

Interview Privacy and Social Conformity Effects on Socially Desirable Reporting Behavior: Importance of Cultural, Individual, Question, Design and Implementation Factors

by

Zeina N. Mneimneh

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey Methodology)
in The University of Michigan
2012

Doctoral Committee:

Professor Michael R. Elliott, Co-Chair
Research Scientist Steven G. Heeringa, Co-Chair
Professor Michael P. Couper
Roger Tourangeau, Westat, Inc.

© Zeina Mneimneh 2012

DEDICATION

I dedicate this work to my family, especially...

to Dad, who has always inspired me at all stages of my life even in his absence

to Mom, who always reminded me to eat because I need my energy

to Rima, my eldest sister, who kept my faith going

to Maha, my second sister, who reminded me to be true to my passions

to Rana, my third sister, who brought out the fighter in me

to all my nieces and nephews, who gave me summers to look forward to

to Omar Katerji, my ex-husband, who listened to my school frustrations

And last, but not least, to John and Doodle Ramirez who give me a reason to look forward and be a better person.

ACKNOWLEDGMENTS

Over the past five years, I have been inspired by many individuals and have received a great deal of support and encouragement from my family, friends, colleagues, and faculty. I would first like to extend a great deal of thanks to my co-chairs, Michael Elliott and Steve Heeringa, who made themselves available for long meetings even when they were away, and who gave me insightful directions and statistical advice as I developed my dissertation chapters. I am also grateful for my committee members, Mick Couper and Roger Tourangeau, who challenged my ideas and gave me extensive feedback on my dissertation chapters, especially during the past year.

I would also like to thank my supervisor at work and my colleague Beth- Ellen Pennell, who was very understanding of my dissertation commitments and very supportive during the process; Frauke Krueter, University of Maryland, who was very encouraging of my dissertation ideas, especially at an early stage of developing them; and all of my friends, but specifically, Tuba Suzer Gurtekin and Aghapi Mordovanakis, who listened to me talk about my research and my frustrations and who regularly met with me for writing “parties.”

Last, but not least, I would like to thank Ronald Kessler and Beth-Ellen who facilitated access to the data I used in this dissertation, and the principal investigators of the World Mental Health Consortium who have given me access to their data. Special thanks to Elie Karam, my previous supervisor and a current colleague and friend in Lebanon, who introduced me to the World Mental Health project and its great members and who was one of the main people who encouraged me to join the Survey Methodology Program at the University of Michigan, whose faculty are the leaders and the best in survey research.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	viii
LIST OF EQUATIONS.....	ix
ABSTRACT	x
CHAPTER	
I. Introduction.....	1
Socially Desirable Response Behavior Framework.....	1
Chapter II.....	6
Chapter III.....	6
Chapter IV	7
References.....	11
II. Cultural and Interviewer Effects on Interview Privacy	13
Introduction	13
Interviewer’s Role	15
Respondent’s and Third Party’s Roles.....	16
Culture’s Role	17
Methods	21
Sample Design for WMH Surveys	21
Questionnaire	22
Questionnaire Administration.....	23
Measures Studied in the Current Chapter.....	23

Analysis	26
Results	30
Presence of Third Party during the Interview in WMH countries	30
Predictors of Third-Person Presence during the Interview	30
Discussion	39
Cultural Variations including Country's Wealth Effects	39
Interviewer Variations	42
Conclusion	46
Appendix	48
References	50
III. Cultural Variations in the Effect of Interview Privacy and the Need for Social Conformity on Reporting Sensitive Information	54
Introduction	54
Methods	61
Sample Designs for the WMH Surveys	61
Questionnaire	62
Questionnaire Administration	62
Measures Studied in the Current Chapter	62
Analysis	66
Results	69
Outcome Rates and Interview Privacy Rates across Countries	69
Effect of Third-Party Presence on Reporting	71
Discussion	75
Conclusion	81
Appendix	83
References	84
IV. Bridging Psychometrics and Survey Methodology: Can Mixed Rasch Models Identify Socially Desirable Reporting Behavior?	88
Introduction	88
Methods	93
Sample Design of Lebanon World Mental Health Study	93

Questionnaire	94
Questionnaire Administration	94
Measures Studied in the Current Chapter.....	95
Analysis	97
Results	103
Respondent Classes Exhibiting Socially Desirable Response	
Behavior	103
Predictors of Class Membership	110
Discussion.....	112
Conclusion	117
Appendix A.....	119
Appendix B.....	122
References	123
V. Conclusion	126
Brief Summary of Findings	126
Recommendations and Implications	127
Research Recommendations.....	127
Design Implications.....	131
Appendix	139
References.....	141

LIST OF FIGURES

FIGURE

I.1	Socially Desirable Response Behavior Framework	5
II.1	Tri-party Influences on Interview Privacy	21
II.2	Predicted Effect of Being a Homemaker vs. Having a Job on Partner Presence during the Interview	35
II. 3	Predicted Effect of Having a Partner with Low-Skill Occupation vs. High-Skill Occupation on Partner Presence	35
II.4	Predicted Effect of Having Very Low Income vs. High Income on Partner Presence	36
III.1	Effect of Third-party Presence on Reporting Suicidal Behavior	72
IV.1	Item Characteristic Curves for Three Items with Increasing Level of Difficulty	99
IV.2a	Distribution of Estimated Probability for Being a Member of Hyperthymic Class I	106
IV.2b	Distribution of Estimated Probability for Being a Member of Depressive Class I	106
IV.3	Two-Class Mixed Rasch Model for Hyperthymic Subscale	108
IV.4	Two-Class Mixed Rasch Model for Depressive Subscale	109
V.1	Computerized Adaptive Testing Logic	133
V.2	Hypothetical Example of an Adaptive Measurement Tool Using One-class Model (i.e. simple IRT) vs. Two-class Model I (Mixed Rasch Model)	135

LIST OF TABLES

TABLE

II.1	<i>WMH Survey Interview Mode, Supervisor-to-Interviewer Ratio, Completed Interviews, and Response Rates</i>	23
II.2	<i>Percent (Standard Error) of World Mental Health Interviews Conducted in the Presence of a Third Party</i>	30
II.3	<i>Odds Ratio from Multilevel Logistic Model Predicting Presence of Any Third Party during the Interview</i>	33
II.4	<i>Odds Ratio from Multilevel Logistic Model Predicting Partner Presence during the Interview</i>	37–38
II.5	<i>Percentage of Interviews Conducted in the Presence of a Third Party during the Interview</i>	48–49
III.1	<i>Weighted Outcome Percent</i>	70
III.2	<i>Percentage of Any Third Party Presence</i>	71
III.3	<i>Percentage of Partner Presence</i>	71
III.4	<i>Odds Ratio and 95% Confidence Interval from Multilevel Logistic Model Predicting Suicide Attempt or Plan</i>	73
III.5	<i>Odds Ratio and 95% Confidence Interval from Multilevel Logistic Model Predicting High Marital Rating Score</i>	74
III.6	<i>Odds Ratio and 95% Confidence Interval from Multilevel Logistic Model Predicting Chronic Conditions</i>	75
IV.1	<i>Fit Indices and Test Statistics for Mixed Rasch Models with Different Number of Classes</i>	105
IV.2	<i>Odds Ratio (OR) and 95% Confidence Interval (CI) from Design-Adjusted Logistic Regression Model Predicting Response Class Membership</i>	111

LIST OF EQUATIONS

Equation IV.1	98
Equation IV.2	100

ABSTRACT

Socially desirable reporting is potentially a major source of measurement error in surveys. It is a response behavior exhibited by respondents when they perceive the information they are requested to provide as sensitive. The literature investigates many of the factors that lead respondents to perceive survey responses as sensitive, but lacks a framework that integrates these factors. We present a framework that displays how these factors relate to one another and lead to socially desirable response behavior. We also demonstrate the importance of integrating these factors by specifically investigating the respondent's cultural background, the respondent's need for social conformity, interview privacy, and their effect on socially desirable reporting.

We present three studies on these factors. The first and second studies focus on predictors of interview privacy and its effect on socially desirable response behavior. Findings from the first study reveal significant variation across cultures and across interviewers in establishing a private interview setting. A country's wealth, level of individualism, and masculinity significantly affect the privacy setting of the interview, either directly or by interacting with respondent and third party characteristics. The second study demonstrates that third-party presence during the interview interacts with the respondent's need for social conformity and his or her cultural background to either decrease or increase the odds of reporting undesirable or desirable outcomes. Ignoring such interaction effects could result in misleading conclusions about third-party presence influences on reporting sensitive information. Study three generates a topic-specific measure of social conformity that integrates both respondent-level and question-level parameters using mixed Rasch models. It demonstrates the importance of using respondent-level topic-specific measures of social conformity

rather than a one-time measure, as social conformity could be affected by external stimuli (such as the questions themselves) that vary across the course of the interview. We conclude by discussing design implications for interview privacy measures, interviewer training, and measures of social conformity, along with recommendations for further research. We also discuss the possible use of mixed models as the backbone for adaptive measurement tools to capture in real time the respondent's activated need for social conformity and ultimately reduce measurement error.

Chapter I

Introduction

Socially Desirable Response Behavior Framework

Socially desirable reporting is potentially a major source of measurement error in surveys (Tourangeau & Yan, 2007). It is a response behavior exhibited by respondents as a mean to project a favorable image of themselves that maximizes social conformity and minimizes negative judgments by others (Johnson & Van De Vijver, 2003). Respondents usually manifest such behaviors when they perceive the information they are requested to provide as sensitive.¹

Several factors lead respondents to perceive survey responses as sensitive. The theoretical and empirical literature addressing these factors comes from several fields. *Respondent-level* factors, such as the respondents' need for social conformity and their cultural backgrounds, are mainly discussed in the fields of psychology (Crowne & Marlowe 1960; Edwards, 1957; Paulhus, 1984), social psychology (Cialdini & Goldstein, 2004), and cross-cultural psychology (Smith, Bond, & Kağıtçıbaşı, 2006). *Question-related, design, and interview implementation* factors are mainly addressed in the survey methodology field (Aquilino, 1997; Fowler, 1995; Singer, Von Thurn, & Miller, 1995; Tourangeau, Rips, & Rasinski, 2000; Tourangeau & Smith, 1996). Tourangeau and Yan (2007) provided a review of the factors that affect reporting on sensitive topics, including mode of administration, interview setting, and question-design strategies. It seems that to better understand response behavior to sensitive topics, an integration of the effects of these four sets of factors

¹ Though the definition of sensitive questions is much broader (Tourangeau & Yan, 2007), throughout this dissertation we use the term "sensitive" to refer specifically to questions asking respondents about desirable or undesirable information.

(respondent-related, question-related, design-related, and implementation-related) is needed. In this chapter, we attempt to represent schematically in Figure I.1 the four sets of factors and how they interact to affect socially desirable reporting behavior.

The first set of factors that influence respondent's perceived sensitivity of the information requested includes the respondent's cultural background, beliefs, values, needs and motivations for social conformity, expectations, demographics, as well as socioeconomic factors. The second set includes question-level characteristics that could either directly or indirectly (by interacting with the respondent's characteristics) affect the respondent's perception of the sensitivity of the information. This set includes the topic of the question itself (e.g., desirable or undesirable, past or current assessment); the type of the question (e.g., whether the question asks about behavior, attitudes, values, traits, or knowledge, etc.); and the wording of the question, including the level of question comprehension and whether the question is phrased in a direct or indirect manner or neutralizing rather than non-neutralizing manner.² The third and fourth sets define the context of the interaction between the respondent and the question. The third set includes design factors such as methods of data collection, confidentiality assurance, study introduction, mode of administration (e.g., interviewer-administered vs. self-administered), and interviewer training. These factors could interact with respondent-level characteristics to increase or decrease the perceived sensitivity of the response. Some design factors, like interview mode and interviewer training, also affect the fourth set of factors, interview implementation. Interview implementation includes the interview location and interview privacy setting (such as whether a third person is present, duration of the third person's stay during the interview, relationship of the third

² Direct phrasing is when the respondent is asked about his or her behavior, attitudes, or traits, etc.; indirect phrasing is when the respondent is asked about the behavior, attitudes, or traits of close friends, family, or neighbors.

Neutralizing questions are those that include a statement that neutralizes the behavior or attitude being asked about, such as an "everybody does it" approach. For example, "Even the calmest parents get angry at their children some of the time. Did your children do anything in the past seven days to make you angry?"

person to the respondent, and characteristics of the third person). This set of factors could also be affected by respondent's characteristics, such as demographics, socioeconomics, and cultural background, and could itself interact with certain respondent characteristics, such as the respondent's need for social conformity affecting his or her perceived sensitivity of the information to be disclosed.

The three main chapters that follow were motivated by this framework. The main interview implementation measure investigated is interview privacy. Chapter II explores the determinants of interview privacy. Chapter III builds on Chapter II and investigates the effect of interview privacy on socially desirable response behavior to different types of survey questions. Both main effects as well as interaction effects with the respondent's need for social conformity and with his or her cultural background are investigated. Chapter IV attempts to develop a quantitative topic-specific measure of respondents' need for social conformity and investigates its association with interview privacy and interview mode.

This research has several implications in relation to socially desirable response behavior, one of the main sources of measurement error. These include research implications that build on the current findings to develop further our understanding of interview privacy effects on social conformity needs and reporting, as well as design implications addressing measures and practices that could reduce such measurement error. These implications are discussed further at the end of this chapter as well as the conclusion chapter.

The rationale for focusing on interview privacy, rather than other design or implementation factors, was driven by the lack of consistent findings in the literature on its effect on responding to different types of sensitive questions. Earlier work has been done by Aquilino on establishing a framework for the effect of interview privacy on reporting sensitive information. In 1997, Aquilino described three features that affect the size and direction of bystanders' effect on reporting sensitive information. The first is whether the survey question asks about factual or subjective information. The second is whether the bystander is

knowledgeable about the factual information requested. The third is the perceived likelihood that the respondent might experience negative consequences by revealing new and unwelcome information to the bystander. If the third party already knows the information to be reported, then their presence might not have an effect or it might even lead to more truthful reporting. If the third party does not know the information requested and the respondent perceives a high likelihood of negative consequences by revealing this information, then the presence of a third party might lead to misreporting. Aquilino's last two conditions relate to the type of relationship that exists between the third party and the respondent. The relationship must be significant to the respondent, and the respondent's answer must bear directly on the relationship, for the respondent to change his/her answers in order to convey a more desirable image in the presence of the third person (Aquilino, 1997; Pollner & Adams, 1997). Other than the type of relationship between the third person and the respondent, the respondent's perceived likelihood of experiencing possible negative consequences by revealing sensitive information could be also affected by respondent-level characteristics, including the respondent's cultural background, motivations, and need for social conformity. Moreover, an important player that could affect interview privacy and who is not included in Aquilino's framework is the interviewer. Interviewers are also the main source of information on interview privacy, as they are responsible for making observations on the interview setting and recording them. Unfortunately, possible variations due to the respondent's characteristics and interviewer effects have not been explored in the literature and could partially explain the mixed findings related to the effect of interview privacy on reporting sensitive information. Before exploring such variations, it is imperative to understand what determines interview privacy as it is not a random occurrence.

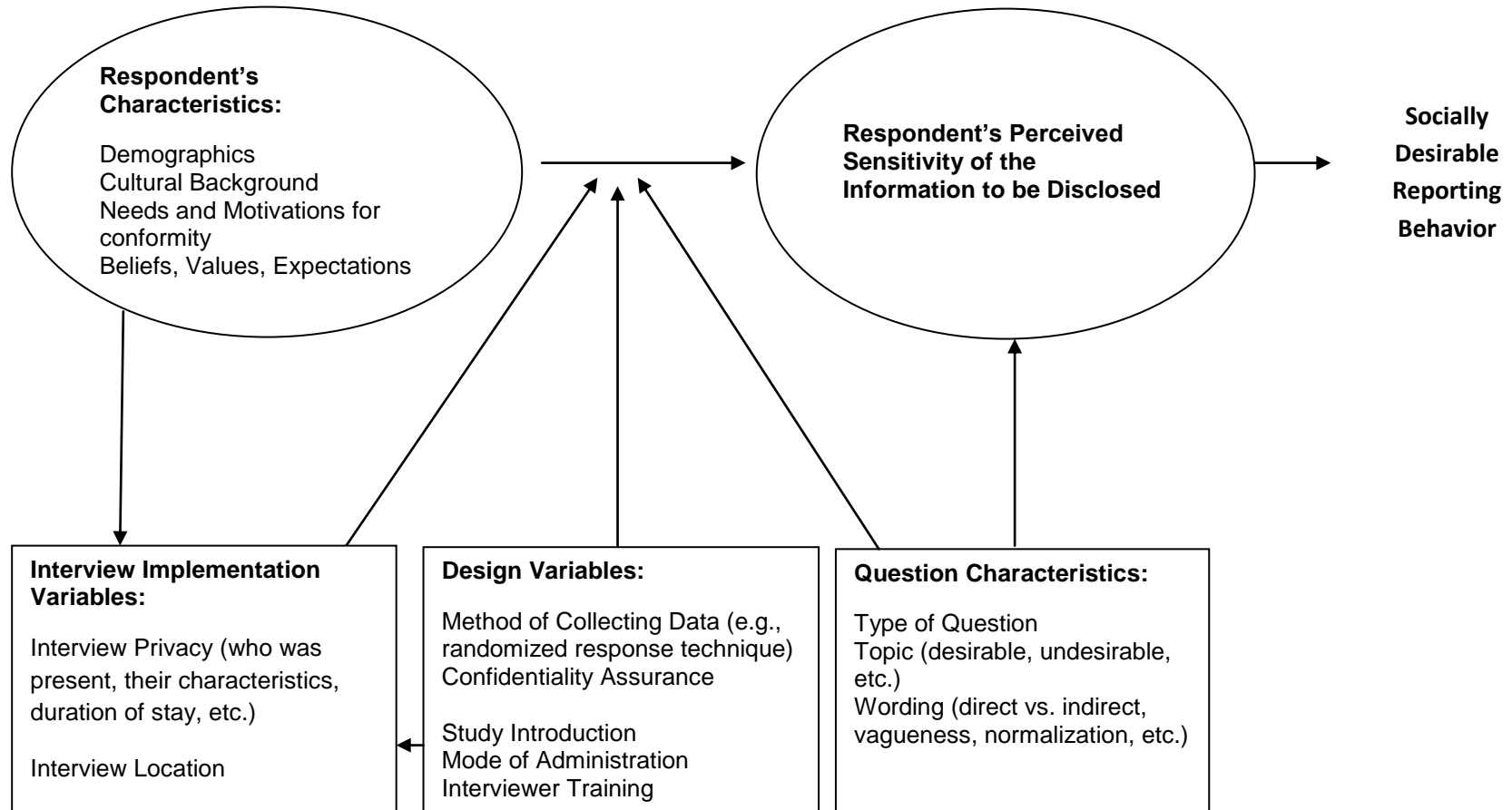


Figure I.1: Socially Desirable Response Behavior Framework

Chapter II

Chapter II explores predictors of interview privacy that are related to the three main players—the interviewer, the respondent, and the third party—in the survey administration. Cultural factors that are common to the three players as well as individual characteristics are explored. Chapter II investigates whether there are any cultural or interviewer variations in achieving interview privacy and whether the effect of respondent and third-party characteristics on interview privacy varies by culture. To investigate such variations, survey data from 14 countries that differ in their level of affluence, family dynamics, and individuals' social power were analyzed. These 14 countries are part of the World Mental Health (WMH) Surveys, a cross-national initiative that collects data from multiple countries using comparable survey design and implementation procedures. The cultural dimensions investigated in Chapter II include the country's level of individualism, masculinity, power distance (a measure of the degree to which less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally), and wealth. Since respondents are nested within interviewers and interviewers are nested within countries, a multilevel logistic regression was used to analyze predictors of interview privacy.

Chapter III

Chapter III builds on Chapter II by using the findings of what determines the interview privacy setting to explore informed hypotheses about the possible interaction effects of the privacy setting and the respondent's characteristics on reporting different types of information. Chapter III investigates whether the effect of third-party presence on reporting sensitive information is moderated by the respondent's need for social conformity and the respondent's cultural background, and whether these effects vary by the type of question: behavioral, attitudinal, or relatively neutral events. This chapter uses the same WMH data as Chapter II as well as multilevel logistic regression to study reporting effects on questions about sensitive behaviors and attitudes such as suicidal behavior and

marital rating, and contrasts these with relatively neutral outcomes, such as having high blood pressure, asthma, or arthritis.

Chapter IV

To measure the respondent's need for social conformity, Chapter III relies on an adapted version of the Marlowe-Crowne Scale that was included in the WMH surveys. The Marlowe-Crowne Scale is one of several social-desirability scales commonly used in the literature. The other two commonly used scales are Edwards's (1957) and Paulhus's Balanced Inventory of Desirable Responding (1984). The Marlowe-Crowne Scale, which is used in Chapter III, was the only measure of social conformity available across all the WMH countries. Using the Marlowe-Crowne Scale requires the assumption that from the time the scale is administered, the respondent's need for social conformity is stable across all the survey questions. However, as discussed by social-psychologists, one's social-conformity needs may be contextual (Cialdini & Goldstein, 2004). Such needs could be activated and strengthened by contextual stimuli, including interview privacy and interview mode. If the stimuli are not consistent across the survey interview, then the respondent's score on the scale might change as the interview progresses.³

One of the important stimuli that could strengthen or dampen such needs affecting the respondent's response behavior are the survey questions themselves. As shown in Figure I.1, the characteristics of survey questions could directly and/or indirectly (through strengthening certain respondent characteristics) affect the respondent's perception of the sensitivity of the information to be disclosed, ultimately leading to socially desirable response behavior. Though, few authors have *theoretically* addressed the simultaneous effect of question characteristics and respondent characteristics on social desirability bias (Cannell, Miller, & Oksenberg, 1981; DeMaio, 1984), most *empirical* literature has studied them separately. In fact, the psychological

³ Other limitations of the Marlowe-Crowne Scale, such as possible measurement distortion and cultural inequivalence, are discussed in chapter IV.

literature has been mainly focused on respondent-level characteristics, and the survey methodology literature has been mainly focused on question-level characteristics. To address this gap, Chapter IV explores the use of mixed Rasch models, a measurement tool that incorporates both levels (respondent and question), to measure the respondent's socially desirable response behavior.

A mixed Rasch model is a mixture of latent class and item response theory (IRT). IRT models are usually applied to scales that are made of multiple items measuring a latent construct (e.g., mathematical ability, reading ability, etc.). They aim to estimate the respondent's measure on the latent construct by modeling the probability of endorsing the scale items, controlling for item characteristics. One of the item characteristics that are modeled is called "item difficulty." Item difficulty refers to the extent to which an item taps into the construct. The term comes from the educational field, where the constructs measured are generally knowledge constructs and the objective is to measure the respondent's level of knowledge by administering items with different levels of difficulty.

IRT models are also heavily used in the field of psychology and, recently, in the health field to measure non-knowledge items such as personality measures, depression, and pain. When applied to non-knowledge items, it is conceivable that question characteristics, such as the level of desirability or undesirability of the item, might affect the probability of endorsing the question, contributing to its estimate of item difficulty. Thus, items that are more desirable would have a higher probability of being endorsed and items that are less desirable would have a lower probability though they both tap into the same level of the construct. In fact, Zickar and Ury (2002) have empirically tested this and found that item desirability was positively associated with item difficulty. However, as discussed earlier, the respondent's perceived sensitivity of the information could be affected by more than the question itself. Respondent-level characteristics, specifically their need for social conformity, might play an important role in how the item is perceived, and in turn, its probability of being endorsed. Respondents whose needs for social conformity get activated and

strengthened by the potentially sensitive question could endorse undesirable items at a lower probability (and thus have higher item difficulty estimates) and desirable ones at a higher probability (and thus have lower estimates of item difficulty) than their counterparts who do not have such needs or whose needs are not necessarily activated by such items.

Thus, if item difficulty partially measures item desirability (or undesirability) when applied to scales with sensitive items, a measurement model that estimates the existence of latent classes of respondents with heterogeneous levels of item difficulty (*while simultaneously controlling for their level of construct*) could capture groups of respondents with different social-conformity needs triggered by the sensitive items. Rost (1990) described such mixtures of latent class and Rasch item response theory models as mixed Rasch models. Rost, however, warned that latent classes of respondents could be measuring certain structural differences in the construct itself. This potential confounding with construct structural differences needs to be considered carefully when interpreting the meaning of the latent classes.

In Chapter IV, we investigate whether applying mixed Rasch models to a range of scales—including desirable, undesirable, and relatively neutral items—captures latent classes of respondents representing those with different socially desirable response behavior. To explore the nature of these classes and confirm that they measure respondents' characteristics linked to socially desirable reporting, such as the need for social conformity, we investigate the association between latent-class membership and independent, objective design and implementation factors that could be activating and enhancing such social-conformity needs, namely interview privacy and interview mode. The rationale is that if the mode of the interview and interview privacy affect social-conformity needs and in turn the response behavior, then respondents who were interviewed in private and by a self-administered mode would be more likely to belong to the latent class(es) exhibiting less socially desirable response behavior.

In sum, the three following chapters emphasize the integration of respondent-level, question-level, design, and interview implementation factors in studying socially desirable response behavior, a potentially major source of measurement error in surveys asking about sensitive information. We conclude with a discussion of research and design implications that aim to develop further our understanding of interview privacy effects on social-conformity needs and reporting, and urge researchers to design measures and practices that could reduce such measurement error. The design implications discussed include implementing interviewer training practices to increase achieving and maintaining interview privacy. By increasing interview privacy, we attempt to better standardize the interview setting across respondents included in the same sample and thus reduce any measurement variations created by the setting effect. Moreover, any possible socially desirable response behavior among certain respondents whose needs for social conformity are activated and strengthened by the settings could be reduced. We also advocate new measures of social conformity, based on our mixed Rasch models. The ultimate aim is to design computer adaptive measurement tools (with mixed Rasch models as a backbone) that estimate real time respondents' response patterns resembling socially desirable response behavior, as well as their corresponding level of construct. Based on these estimates, the adaptive measurement tool would route each respondent to the next scale item that fits his or her estimated level of construct or to intervention messages that aim to reduce their social-conformity concerns, reducing measurement variance (with the former) and measurement bias (with the latter) while also minimizing the number of scale items that need to be asked compared to a static instrument. This will require future work to explore further the nature of the new measures of social conformity suggested in Chapter IV.

References

- Aquilino, W. (1997). Privacy effects on self-reported drug use: Interactions with survey mode and respondent characteristics. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (National Institute on Drug Abuse, Monograph 167). Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, & National Institute on Drug Abuse, Division of Epidemiology and Prevention Research.
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement, 62*, 570–589.
- Caldini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, 389–437.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354.
- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. E. Turner & E. Martin (Eds.), *Surveying subjective phenomena*. New York, NY: Russell Sage.
- Edwards, A. L. (1957). *The social desirability variable in personality research and assessment*. New York, NY: Dryden Press.
- Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks, CA: Sage Publications.
- Johnson, T. P., & Van De Vijver, F. J. R. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods*. Hoboken, NJ: Wiley.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.
- Pollner, M., & Adams, R. E. (1997). The effect of spouse presence on appraisals of emotional support and household strain. *Public Opinion Quarterly, 61*, 615–626.

- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271–282.
- Singer, E., Von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly*, 59, 66–77.
- Smith, P. B., Bond, M. H., & Kağıtçıbaşı, Ç. (2006). *Understanding social psychology across cultures: Living and working in a changing world*. London, U.K. Sage Publications.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, U.K.: Cambridge University Press.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and questions context. *Public Opinion Quarterly*, 60, 275–304.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883.
- Zickar, M. J., & Ury, K. L. (2002). Developing an interpretation of item parameters for personality items: Content correlates of parameter estimates. *Educational and Psychological Measurement*, 62, 19–31.

CHAPTER II

Cultural and Interviewer Effects on Interview Privacy

Introduction

Survey interviewing is a dynamic process that is affected by the setting of the interview. The setting encompasses physical features as well as the events and interactions occurring while the interview is taking place. While the content of the interaction is mainly predetermined by the researcher, the setting is less controllable and is usually jointly determined by the interviewer, the respondent, and other household members. Surveys that are designed to collect sensitive information or to measure the respondent's own knowledge usually call for a private interview setting. However, such a setting is relatively difficult for the interviewer to control and can vary over the course of the interview.

Privacy (or lack of it) is an important feature of the interview mainly due to its possible effect on reporting information, especially sensitive information. The literature on the effects of third-party presence on survey reporting is mixed (see Tourangeau and Yan, 2007, for a meta-analysis). While some studies found that third-party presence is associated with reduced reporting of undesirable outcomes (Aquilino, 1993; Aquilino, 1997; Aquilino, Wright, & Supple, 2000; Moskowitz, 2004; Taietz, 1962), others observed increased reporting of such outcomes (Bulck, 1999; Edwards, Slattery, & Ma, 1998; Hoyt & Chaloupka, 1994). This inconsistency could be attributed to a number of factors related to the type of question, other study design features, respondent characteristics, as well as the relationship of the third party to the respondent, and the information already held by the third party. Regardless of the direction of the effect, variation in interview privacy within a sample changes the interview experience across

respondents, leading to differences in the measurement properties of the collected information. Above all, the presence of others can jeopardize the protection of human subjects and thus should be given more attention (Moskowitz, 2004).

When the design of the survey calls for conducting the interview in a private setting, researchers rely heavily on interviewers to ensure such a setting. In reality, interviewers are essentially guests in respondents' homes and might find it difficult to request, achieve, and maintain privacy. In fact, a substantial proportion of interviews are conducted in the presence of a third party. The appendix at the end of this chapter summarizes the rates of nonprivate interviews that have been reported in the literature across different countries. Most studies conducted in the United States reported that a third party was present in more than 40% of the interviews with rates ranging from 21% to 59% (Anderson & Silver, 1987; Moskowitz, 2004; Pollner & Adams, 1994; Pollner & Adams, 1997; Silver, Abramson, & Anderson, 1986; Smith, 1997). The most common type of third person present was a spouse or other partner followed by children (Hartmann, 1995; Pollner & Adams, 1997; Reuband, 1992; Smith, 1997). Parents, siblings, and others were less often present (Aquilino et al., 2000; Hoyt & Chaloupka, 1994). Similar ranges of rates have also been reported in several countries in Europe, including Belgium (where 48%–52% had a third party present; Bulck, 1999; Welkenhuysen-Gybels & Billiet, 2001), the United Kingdom (40%–50%; Zipp & Toth, 2002), and Germany (the average rate reported by Reuband, 1992, was 37%). In a multinational study conducted in 24 developing countries in Asia, Latin America, the Caribbean, and Africa, the majority of countries (17 out of 24) had a third person present in more than 40% of their interviews. The rates of interviews with third parties present ranged from 17% in Paraguay to 82% in Nepal (Casterline & Chidambaram, 1984). Partners were also the most common type of third party to be present during the interview in these countries. It seems that complete interview privacy is not always possible even when the design of the survey calls for it. Variations within a nation and across nations could be attributed to a number of factors related to the three

main players in the interview interaction: the interviewer, the respondent, and the third party.

Interviewer's Role

Most survey organizations require their interviewers to undergo general interviewing and study-specific training. These cover a wide range of topics, including standardized interviewing techniques, respondent recruitment, respondent consent procedures, study-specific instrument conventions, collection of biomarkers (if needed), and practice interviews. During these training sessions, interviewers are also instructed that the interview needs to be conducted in a private setting (if the study protocol calls for it). Unfortunately, the interview privacy requirement does not seem to get a lot of attention in interviewer trainings (Aquilino, 1993; Smith, 1997; Taletz, 1962). In 1997, Smith described a section on the need for privacy in a typical set of interviewer training material as follows:

A quiet area in which to conduct the interview is crucial. The presence of others may bias the answers. This sometimes requires tact and ingenuity to avoid an audience. The interviewer must inform the respondent that it is necessary to conduct the interview in private, in order to protect his/her confidentiality. Under ideal circumstances, the interview is conducted in a private room equipped with good lighting and a table on which to work. However, these things are often not available, so the interviewer will need to improvise, trying to stay as close as possible to the ideal.⁴

The training material does not provide instructions on how to achieve privacy. Smith (1997) also pointed out that though privacy and confidentiality are emphasized in training materials, the interviewing manuals he reviewed do not explicitly mention avoiding the presence of a third party. It seems likely that interviewers are mainly left to their own judgment and skills to enforce privacy. In

⁴ This quote was taken from Smith, 1997. It comes from the National Opinion Research Center (NORC) training manual on basic interviewing techniques, p. 6. The training manual referenced does not seem to be specific to a study. Thus, it does not necessarily represent other training material in other organizations that could be specific to studies covering sensitive topics.

view of the limited attention given to interview privacy in interviewer training, one would expect interviewers to vary in their attitudes toward privacy and their ability to achieve and to maintain it. A search of the literature has revealed no study that has measured interviewer variation in achieving interview privacy.

Respondent's and Third Party's Roles

The presence of a third party during all or part of the survey interview is not a random occurrence. Respondent characteristics and motivations, household characteristics, and cultural factors may all play a role in the presence of others during the interview. Third parties could include the respondent's spouse or partner, parents, children, siblings, and others. The majority of the literature on predictors of third-party presence focuses on partner presence and any third-party presence.

Hartmann (1995) described a number of conditions that need to be met for a spouse to be present during the interview. Above all, the respondent needs to be married and the spouse needs to be available at the time of the interview. Given that the spouse is available, a number of motivational, demographic, and socioeconomic factors could contribute to his or her presence.

The respondent's age has been found to be associated with spouse presence. Older respondents are more likely to have their spouse present during the interview (Aquilino, 1993; Casterline & Chidambaram, 1984; Welkenhuysen-Gybel & Billiet, 2001). This may reflect the tendency for older couples to spend more time at home or the perceived need for assistance in the interview process. The literature investigating the effect of gender on partner presence has mainly focused on populations in Germany and the United States. According to this literature, women were more often present when their husband was being interviewed than vice versa (Aquilino, 1993; Aquilino, 1997; Aquilino et al., 2000; Hartmann, 1995). Women are more likely to be at home when their husband is being interviewed, and they also may be more interested in the interview process compared to men. Whether husbands are present when their wives are interviewed may be a function of the gender of the interviewer. Hartmann (1995)

found that though women were less likely to have their spouse present than men were, when women were interviewed by males, they were more likely to be interviewed in the presence of their husband.

Respondents who are not working are usually available during the day when their spouses are likely to be at work and thus are more likely to be interviewed in private (Aquilino, 1993; Hartmann, 1995). On the other hand, spouses who are not working may spend more time at home and are more likely to be around when the respondent is being interviewed (Aquilino, 1993; Hartmann, 1995). Two other characteristics related to spouse presence are education and income. Lower education level and lower income both increase the likelihood of having a spouse present during the interview (Aquilino, 1993; Aquilino, 1997). Aquilino (1997) speculated that low socioeconomic status might be a proxy for a dwelling with fewer rooms, making it difficult for the interviewer to insist on privacy. These socioeconomic effects also apply to the presence of any third party and not just spouse presence (Casterline & Chidambaram, 1984).

Culture's Role

Anthropologists have pointed out that there are considerable differences in the way privacy is viewed across cultures (Streib, 1973). In some cultures, privacy is expected and is a welcomed feature in an interaction. In fact, being present while another person is being asked a direct personal question is regarded as intrusive. In other cultures, it might be considered insulting or threatening to ask other household members to leave the room where the interview is conducted (Pollner & Adams, 1994; Streib, 1973).

Hofstede, Hofstede, and Minkov (2010) discussed several cultural dimensions and their impact on different aspects of people's lives. Among the dimensions they discussed that are relevant here are individualism, power distance, and masculinity. The authors assigned scores for each of those dimensions for 76 countries and regions. The scores were based on survey data from Geert Hofstede's International Business Machines (IBM) studies in the 1970s and their replication by Hoppe (1998), Shane (1995), Merritt (2000), de

Mooij and Hofstede (2001), Mouritzen and Savara (2002), and Van Nimwegen (2002). The position of nations relative to one another with respect to these cultural dimensions has been shown to be stable over time (Ingelhart, 1997; Schwartz, Bardi, & Bianchi, 2000). Hofstede et al. (2010) stated that “Dimension scores—at least their relative positions—have remained as valid in the year 2010 as they were around 1970, indicating that they describe relatively enduring aspects of these countries’ societies” (p. 69). Moreover, these cultural dimensions have been validated over the years by measuring their correlation with other measures that are expected to reflect similar cultural differences (Hofstede, Hofstede, & Minkov, 2010; Smith, Bond, & Kağıtçıbaşı, 2006).

The first cultural dimension of interest is individualism:

Individualism pertains to societies in which the ties between individuals are loose; everyone is expected to look after his or her immediate family. Collectivism as its opposite, pertains to societies in which people from birth onwards are integrated into strong, cohesive in-groups, which through people’s lifetime continue to protect them in exchange for unquestioning loyalty (Hofstede et al., 2010, pp. 107–108).

Thus, in collectivist cultures, the self is defined in terms of relationships with others. In order to maintain harmony and interdependence, close attention needs to be given to others in the social context especially if they belong to the in-group circle (Smith et al., 2006).⁵ It is considered normal in a collectivist culture that a member of one’s in-group can invade one’s privacy at any time (Hofstede et al., 2010). Furthermore, direct confrontation, especially with strangers, is considered rude and undesirable in collectivist societies (Hofstede et al., 2010). Interviewers, being guests in the respondent’s home, might find it difficult to request privacy if the respondent insists on the presence of a family member or if family members insist on staying during the interview. This is in contrast to individualist societies where the right to privacy is a central theme.

⁵ The in-group circle is defined as family, friends, and people who are concerned about the individual’s welfare (Triandis, 1995).

The second dimension is power distance. Power distance can be defined as:

The extent to which less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally. Institutions are the basic elements of society such as the family, the school, and the community (Hofstede et al., 2010, p. 77).

It is conceivable that in high power distance countries, respondents who have less authority in the household might find it difficult to request other household members with greater authority (such as their parents, in-laws, and husbands) to leave the interview site.

The third dimension is masculinity:

A society is called masculine when emotional gender roles are clearly distinct: men are supposed to be assertive, tough, and focused on material success, whereas women are supposed to be more modest, tender, and concerned with the quality of life. A society is called feminine when emotional gender roles overlap: both men and women are supposed to be modest, tender, and concerned with the quality of life (Hofstede et al., 2010, pp. 155–156).

Thus, according to this definition, in high-masculinity countries, men might put more emphasis on their work and accomplishments and might not have the time or interest to be present when a member in their household is being interviewed. This might be especially true for men who hold high-level occupations such as professionals, senior managers, and academic professors.

Previous research has shown that there is a negative correlation between individualism and power distance. One of the suggested reasons is that both dimensions are correlated with wealth, though in opposite directions. Power distance was found to be negatively correlated with a nation's wealth while individualism seemed to be positively correlated with wealth. In their book on culture and organizations, the authors reported that Gross National Income (GNI) per capita was found to explain no less than 71% of the nations' individualism scores (Hofstede et al., 2010). According to Hofstede and his colleagues, wealth leads to individualism and not vice versa (Hofstede et al., 2010; Smith et al.,

2006). When a country's wealth increases, its citizens become less concerned with survival problems and more interested in societal values and self-expression. Thus an increase in the wealth of the country could affect interview privacy not only through the possible increase in physical opportunities available for individuals (such as more living space) but also through the shift in their values towards more individualistic views that emphasize self-expression and protection of one's own privacy.

Therefore, to understand the respondent's behavior and the dynamics of the interview interaction, it is useful to attend to such cultural values. Cultural, demographic, and socioeconomic characteristics might make it easier or harder for the interviewer to request and to negotiate the interview setting and, specifically, its privacy. Such effects and their variation across a wide range of cultures have never been studied in the literature.

The three different players, their interaction, and their different characteristics that could affect the privacy setting of the interview as described earlier are summarized in Figure II.1.

Using data from a cross-national survey, the World Mental Health (WMH) Initiative, this chapter aims to investigate between-country and between-interviewer variations in interview privacy. Respondent, third party, interviewer, and cultural effects are examined.

The following five main research questions will be examined:

1. Are there significant cultural variations in interview privacy?
2. What cultural dimensions explain variations across countries in interview privacy?
3. Does the effect of respondent characteristics and/or third-party characteristics on interview privacy vary by culture?
4. Is there a significant variation between interviewers in the level of interview privacy they achieve?
5. Subject to 4, does interviewer variability in achieved privacy rates differ across countries?

From a theoretical point of view, this research will allow survey researchers to better understand the dynamics of the interview and the possible effects of interview privacy (or lack of it) on reporting sensitive information. From a practical point of view, it will help inform interviewer training practices, certification, and materials.

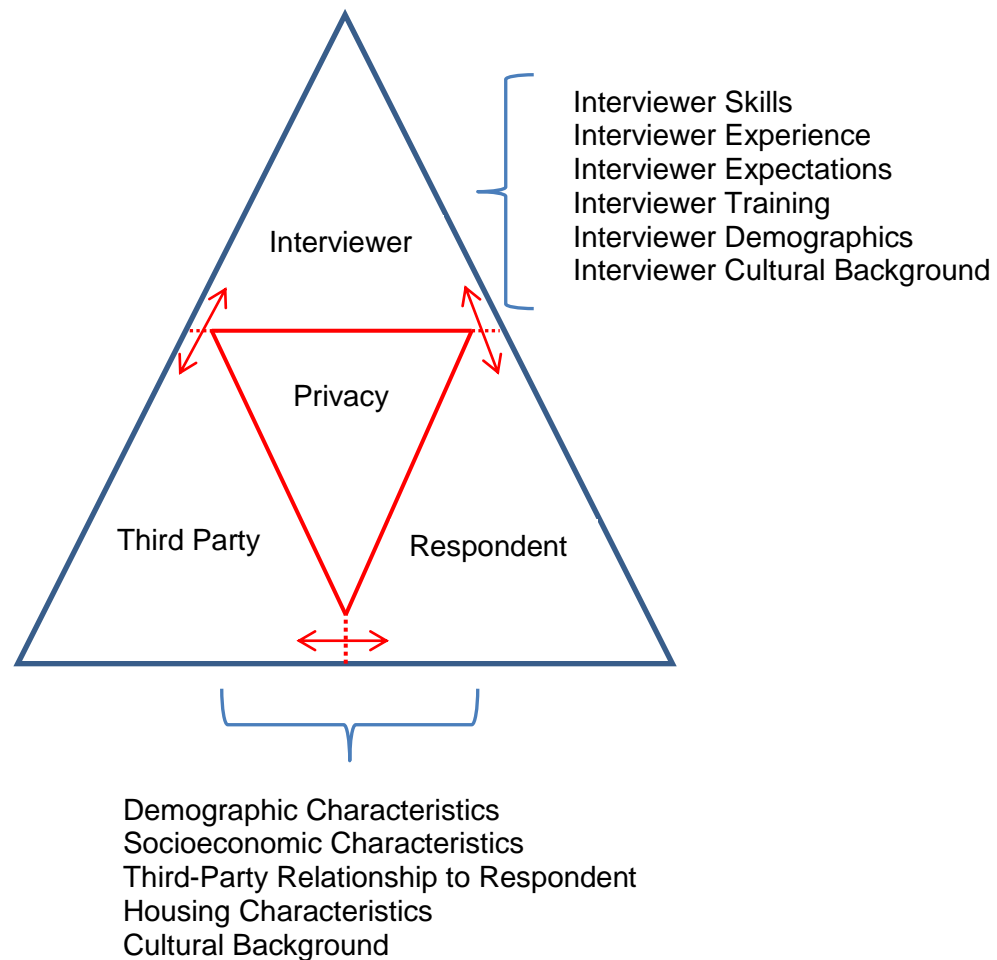


Figure II.1: Tri-party Influences on Interview Privacy

Methods

Sample Design for WMH Surveys

The WMH surveys were administered in 24 countries. Of these countries, 14 collected observations on interview privacy and have published scores on the nation's level of individualism, power distance, and masculinity (Hofstede et al.,

2010). These 14 WMH countries were Belgium, Brazil, Bulgaria, Germany, India, Italy, Japan, Lebanon, Mexico, Nigeria, the People's Republic of China, Romania, Spain, and the United States of America.

All of the surveys targeted the adult population and most of them featured nationally representative household samples (Heeringa et al., 2008). One (Mexico) was representative of urban areas within the country, one of selected states (Nigeria), and four of selected metropolitan areas (Brazil, India, Japan, and the People's Republic of China). Detailed information on the survey design and methods are published elsewhere (Pennell et al., 2008).

To reduce interview length, the WMH interviews were designed to be administered in two parts: Part 1 included core questionnaire sections and Part 2 included noncore sections. All respondents completed Part 1 while Part 2 was administered to a subsample of Part 1 respondents. The current chapter focuses on Part 2 respondents since a number of the independent variables were collected only in Part 2. Country-specific Part 2 completed interviews range from a low of 1,031 (Lebanon) to a high of 5,692 (United States). Table II.1 summarizes the number of completed interviews in each country and the year(s) during which the data were collected.

Questionnaire

The WHO Composite International Diagnostic Interview (CIDI) Version 3.0 was used in all WMH surveys. The CIDI 3.0 is a fully structured interview that generates diagnoses for a wide range of mental-health disorders. It also collects information on treatment, disability, and the burden of mental-health disorders. Detailed questions on social and family life, employment history, finances, and childhood experiences are included. The questionnaire was translated into each country's local language using the WHO translation guidelines (Harkness et al., 2008; Kessler et al., 2004; Kessler & Üstün, 2004).

Questionnaire Administration

All interviews in the 14 WMH countries considered here were conducted face-to-face by trained interviewers using either paper-and-pencil interviewing (PAPI) or computer assisted personal interviewing (CAPI) methods. Interviewer training in each country was modeled after a centralized training session attended by project staff from each country. Table II.1 summarizes the mode of the interview, supervisor-to-interviewer ratio, and response rates in each of the countries where information is available.

Table II.1: WMH Survey Interview Mode, Supervisor-to-Interviewer Ratio, Completed Interviews, and Response Rates

Country	Data Collection Year	Mode	Supervisor-to-Interviewer Ratio	Completed Part 2 Interviews	Response Rate
Belgium	2001–2002	CAPI	1:7	1,015	51
Brazil	2005–2007	PAPI	–	2,942	81
Bulgaria	2003–2007	PAPI	–	2,232	72
China	2001–2002	PAPI	1:5	1,628	75
Germany	2002–2003	CAPI	–	1,315	58
India	2003–2005	PAPI	–	1,368	99
Italy	2000–2003	CAPI	1:20	1,768	71
Japan	2002–2006	CAPI	1:20	1,374	56
Lebanon	2002–2003	PAPI	1:5	1,031	70
Mexico	2001–2002	CAPI	1:6	2,350	77
Nigeria	2002–2003	PAPI	1:4	2,141	79
Romania	2005–2006	CAPI	–	2,356	71
Spain	2001–2002	CAPI	1:20	2,121	79
United States	2001–2003	CAPI	1:10	5,327	71

Note. CAPI= Computer Assisted Personal Interview; PAPI= Paper-and-Pencil Interview.
– Information not available.

Measures Studied in the Current Chapter

The following measures were collected in all WMH surveys and are included in the current analyses.

Main outcome measure.

Interview privacy. Information on interview privacy was collected in an interviewer observation section at the end of the questionnaire. Interviewers

made the following observations: 1) whether a third party was present during any part of the interview and 2) the relationship of the third party to the respondent (parent, partner, child, youth, or “other adults”).

Predictors

Respondent-level measures. Respondent-level measures included gender, age (18–34, 35–49, 50–64, older than 64 years), marital status (never married, currently married or cohabiting, previously married), relative educational level (high, middle, low, very low), current employment status (employed, studying, taking care of home, other), relative household income level (high, middle, low, very low), and household size (one, two, three, more than three). Among those currently married or cohabiting, information on the partner’s level of education (high, middle, low, very low), and type of occupation (not employed, have a low skill job, low-to-average skill job, average-to-high skill job, and high-skill job) was also collected. Respondent and spouse education was categorized at a country-specific level depending on the reported years of education and the local educational system. Spouse employment level categories were defined as follows: Legislators and professionals were considered high skilled; technicians were high-to-average skilled; clerks, service and sales people, skilled workers, craft workers, plant and machine operators, and members of the armed forces were considered average-to-low skilled; and those with elementary occupations were considered low skilled.

Interviewer-level measures. Interviewer identification numbers were available for all countries except India. This enabled the modeling of random effects (intercepts) for the individual members of the interviewing force in each country. Since no other information on interviewers was available, it was not possible to examine the influence of interviewer characteristics, such as gender, on the variation in privacy rates.

Country-level measures. As noted above, three cultural dimensions were investigated in this chapter: individualism, power distance, and masculinity. Country-level measures of these dimensions estimated from data collected in

each WMH country would be ideal for the current analyses. Unfortunately, these measures were not assessed and estimates from other sources were used instead. Hofstede et al. (2010) published dimension scores for each of the 14 countries. The scores were based on survey data about people's values in 76 countries. In these surveys, individualism was measured by several questions, including whether respondents value personal time and freedom. Power distance questions included respondents' attitudes toward disagreeing with authority figures and preference for autocratic versus participative managers. The third dimension, masculinity, was measured using questions on the value of work recognition, advancement, and challenge.

For each dimension and for each country, a standardized score was calculated based on the average score and the standard deviation across the 14 countries. Higher scores indicated higher levels of the underlying dimension. Finally, the countries' economic strength and standard of living was measured by its GNI per capita. GNI measures in nominal dollar values that were calculated according to the Atlas Method for the year the data was collected are used in this chapter.⁶ Because of the wide variation in GNI across the 14 countries and the big gaps in the measure, the countries were divided into three categories: high GNI per capita, middle GNI per capita, and low GNI per capita. This classification matches the World Bank classification of countries by income. Though we test for the effect of the respondent's level of income on interview privacy, we also investigate the effect of country's GNI, as its association with interview privacy could go beyond its impact on the individual's material wealth to include the individual's values as described earlier.

⁶ According to the WHO, the Atlas method "applies a conversion factor that averages the exchange rate for a given year and the two preceding years, adjusted for differences in rates of inflation" ("GNI per capita," n.d.).

Analysis

This chapter focuses on predictors of any third-party presence during the interview (excluding young children) and of partner presence.⁷ Children have not been found to have much impact on survey reports.

Partners were specifically selected among other third parties because they are a well-defined group whose gender can generally be deduced. They share the same household characteristics as the respondent, and information on their educational level and occupation status was collected during the interview. They were also the third party that was most often present in the majority of the WMH countries.

To investigate the research questions listed earlier in relation to cultural and interviewer variation in interview privacy and their predictors, multilevel logistic regression was used for predicting the presence of any third party and the presence of a partner during the interview.

To predict the presence of *any third party*, two models were used. *Model 1* is a three-level random intercept logistic regression with respondents at level 1, interviewers at level 2, and countries at level 3. Predictors include respondent-level characteristics, country's individualism score, and significant interactions between the country's individualism score and respondent characteristics. This model was designed to test the following hypotheses:

Any third-party presence.

Hypotheses related to cultural variations and cultural dimensions.

Hypothesis 1a: Significant cultural variation exists between WMH countries in the presence of any third party during the interview, controlling for the country's level of individualism, respondent-level characteristics, and the interaction between them.

⁷ Partners include spouses and others who live with and are in a marriage-like relationship with the respondent.

Hypothesis 1b: The presence of any third party during the interview is inversely associated with the respondent's country's level of individualism.

Though interaction effects between respondent-level characteristics and the country's level of individualism were tested, we did not hypothesize any specific direction for these effects.

Hypotheses related to interviewer variations.

Hypothesis 2a: Significant variation exists between interviewers within a country in the presence of any third party during the interview, controlling for respondent-level characteristics, the country's level of individualism, and their interaction.

Hypothesis 2b: The level of interviewer variation in the presence of any third party during the interview varies between countries.

Model 2 includes the same predictors as *Model 1* in addition to the country's level of GNI per capita. The purpose of this addition was to investigate changes in the effect of the country's individualism score after GNI was controlled for. *Model 2* allowed for investigating the following two hypotheses:

Hypotheses related to a country's GNI per capita.

Hypothesis 3a: Controlling for all other covariates in the model, respondents interviewed in low- and middle-income countries have increased odds of being interviewed in the presence of a third party compared to respondents interviewed in high income countries.

Hypothesis 3b: Controlling for the country's level of GNI per capita reduces the apparent effect of the country's individualism on third-party presence.

The same set of models was used for predicting the presence of a *partner*. Two additional country-level characteristics were included in these models: power distance and masculinity. The set of hypotheses explored by these models include:

Partner presence.

Hypotheses related to cultural variations and cultural dimensions.

Hypothesis 4a: Significant cultural variation exists between WMH countries in the presence of a partner during the interview, controlling for the country's level of individualism, power distance, masculinity, respondent-level characteristics, and interactions between respondent-level characteristics and each of the country-level dimensions.

Hypothesis 4b: The presence of a partner during the interview is inversely associated with the respondent's country's level of individualism, controlling for all other variables in the model (except GNI per capita).

Hypothesis 4c: The presence of a partner during the interview is positively associated with the respondent's country's level of power distance, controlling for all other variables in the model (except GNI per capita).

Hypothesis 4d: The presence of a partner during the interview is inversely associated with the respondent's country's level of masculinity, controlling for all other variables in the model (except GNI per capita).

Hypothesis 4e: There is an interaction between the partner's occupation's prestige and the country's level of masculinity, with higher levels of masculinity associated with greater partner presence differences due to the partner's occupation prestige, controlling for all other variables in the model (except GNI per capita).

Though interactions between respondent-level characteristics and the other two cultural dimensions (individualism and power distance) were tested, no specific hypotheses regarding the directions of the interactions were made.

Hypothesis related to interviewer variations.

Hypothesis 5: Significant variation exists between interviewers within a country in the presence of a partner during the interview, controlling for all other variables in the model (except GNI per capita).

Hypotheses related to a country's level of GNI per capita.

Hypothesis 6a: Controlling for all other covariates in the model, respondents interviewed in low- and middle-income countries have an increased odds of partner presence during the survey interview compared to respondents interviewed in high-income countries.

Hypothesis 6b: After controlling for the country's GNI per capita, the effect of the country's individualism on partner presence is attenuated.

All multilevel models presented in this chapter are unweighted. To explore the effect of weights on the findings, the analyses were replicated using weighted and unweighted single-level logistic regression models with dummy variables representing countries. The weights accounted for within-country differential probability of selection, post-adjustment to the country's sociodemographic distributions and subsampling to specific questionnaire sections. Adjusted and unadjusted point estimates were consistent suggesting that the use of weights to adjust for informative sample design was not necessary for the relationships investigated. All analyses were conducted using PROC GLIMMIX SAS version 9.2 (SAS Institute, NC).

Results

Presence of Third Party during the Interview in WMH Countries (Table II.2)

Third-party presence during the WMH survey interviews ranged from a low of 12.6% (Japan) to a high of 70.0% (India). Most countries (eight out of 14) had a third party present in 34%–43% of the interviews.

Partner presence constituted the most common type of third-party presence in the majority of the countries, followed by “other adults.” The least common type of third-party presence was a parent.

Table II.2: Percent (Standard Error) of World Mental Health Interviews Conducted in the Presence of a Third Party

Country	N	Any Third Person	Parent	Partner	Youth	Adult
Belgium	1,015	25.4 (1.4)	2.1 (0.5)	18.7 (1.2)	2.7 (0.5)	3.2 (0.6)
Brazil	2,942	59.0 (0.9)	8.2 (0.5)	26.7 (0.8)	18.8 (0.7)	23.9 (0.8)
Bulgaria	2,232	39.3 (1.0)	3.9 (0.4)	26.1 (0.9)	4.4 (0.4)	9.0 (0.6)
China	1,628	36.1 (1.2)	8.7 (0.7)	19.9 (1.0)	4.1 (0.5)	8.8 (0.7)
Germany	1,315	25.9 (1.2)	2.4 (0.4)	16.4 (1.0)	2.7 (0.5)	6.0 (0.7)
India	1,368	70.0 (1.2)	12.4 (0.9)	31.0 (1.3)	7.3 (0.7)	31.4 (1.3)
Italy	1,768	36.0 (1.1)	6.4 (0.6)	21.2 (1.0)	2.1 (0.4)	9.4 (0.7)
Japan	1,374	12.6 (0.9)	1.4 (0.3)	6.2 (0.7)	1.0 (0.3)	4.5 (0.6)
Lebanon	1,031	66.4 (1.5)	11.8 (1.0)	31.7 (1.5)	15.8 (1.1)	32.1 (1.5)
Mexico	2,350	34.6 (1.0)	7.0 (0.5)	11.4 (0.7)	10.7 (0.6)	11.1 (0.7)
Nigeria	2,141	32.0 (1.0)	3.1 (0.4)	8.4 (0.6)	9.1 (0.6)	19.4 (0.9)
Romania	2,356	35.4 (1.0)	3.7 (0.4)	20.7 (0.8)	5.0 (0.5)	10.6 (0.6)
Spain	2,121	41.2 (1.1)	6.7 (0.5)	23.4 (0.9)	4.6 (0.5)	10.9 (0.7)
United States	5,327	30.2 (0.6)	2.8 (0.2)	17.3 (0.5)	7.6 (0.4)	9.1 (0.4)

Note. Results based on Part II sample with valid observations on interview privacy.

Predictors of Third-Person Presence during the Interview—Multilevel Model Results

Any third person (Table II.3).

Cultural variations. As hypothesized, a significant variation in third-party presence during the interview was found across countries, controlling for respondent demographic and socioeconomic characteristics and the country’s level of individualism. Before controlling for the country’s level of GNI per capita

(Model 1, Table II.3), the country's level of individualism main effect on third-party presence was significant. The country's level of individualism also significantly interacted with several respondent-level characteristics in predicting third-party presence. In general, being a female, under 64 years (as compared to 65 or older), or a homemaker (as compared to employed) *reduced* the odds of having a third person present during the interview (ORs= 0.87, 0.70, and 0.89 respectively at the mean level of individualism across all countries). These differences between gender, age, and employment status mattered more as the country's level of individualism increased. For example, in a country with a level of individualism of one standard deviation above the mean, females had 0.83 times the odds to have any other person present during an interview compared to males, but females living in a country with a mean level of individualism had only 0.87 times the odds compared to males (Model 1, Table II.3).

Interactions were also found between the country's level of individualism and the respondent's marital status, educational level, and household size.⁸ Generally speaking, being married (as compared to single), unemployed (as compared to being employed), having very low education (as compared to high education), and very low income (as compared to high income), and living in larger families (three or more as compared to living alone), *increased* the odds of having a third person present during the interview (ORs= 1.17, 1.11, 2.27, 1.14, and 3.27 respectively at the mean level of individualism across all countries). Such marital, educational, and family size differences in third-party presence during the interview were more pronounced among respondents living in more individualized countries. All interaction effects are presented in Model 1, Table II.3.

GNI per capita effect. Though the country's level of individualism loses its main effect once the country's GNI per capita was controlled for, all interactions between respondent-level characteristics and the country's level of individualism remained significant (Model 2, Table II.3). As hypothesized, the country's level of

⁸ Though these interactions were tested for, no specific hypotheses were made regarding the direction of the interaction effects.

GNI per capita significantly predicted third-party presence during the interview. Respondents interviewed in a country with middle-level GNI per capita had five times the odds (OR=4.96) of having any third person present during the interview compared to respondents interviewed in high-income countries. Though third-party presence was also more common in countries with low-level GNI per capita (as compared to high-income countries), the difference was not significant. The country's GNI per capita explained a big proportion of the between-country variation in third-party presence. The between-country variance dropped from 0.41 to 0.25 after the country's GNI per capita was added to the model.

Interviewer variations. In addition to between-country variation in third-party presence, significant between-interviewer variation was found (as hypothesized). Since the dependent variable was binary, it was not possible to directly examine the total variance decomposition. However, comparing the magnitude of between-country variance to the between-interviewer variance before controlling for any cultural dimension, we found that the latter is larger than the former 0.50 vs. 0.59. The average between-interviewer variance across all countries was 0.59. In fact, the majority of the countries had an estimated between-interviewer variance between 0.46 and 0.67. The lowest country-specific between-interviewer variance was 0.22.

The interviewer-level rate of third-party presence in this latter country ranged from 11% to 64%. The country with the highest between-interviewer variance (1.22) had an interviewer-level third-party presence rate that ranged from 0% to 69%.

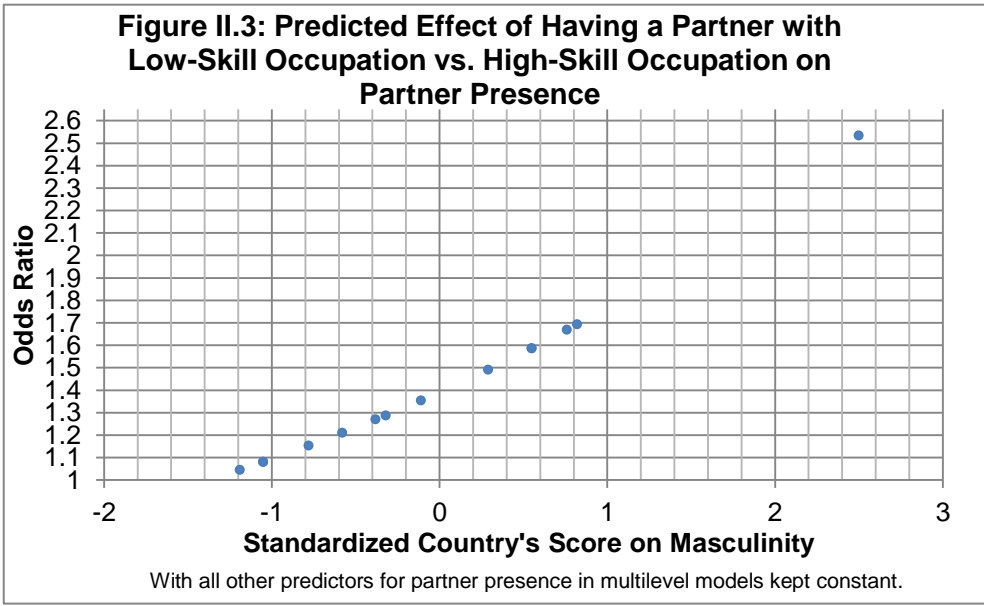
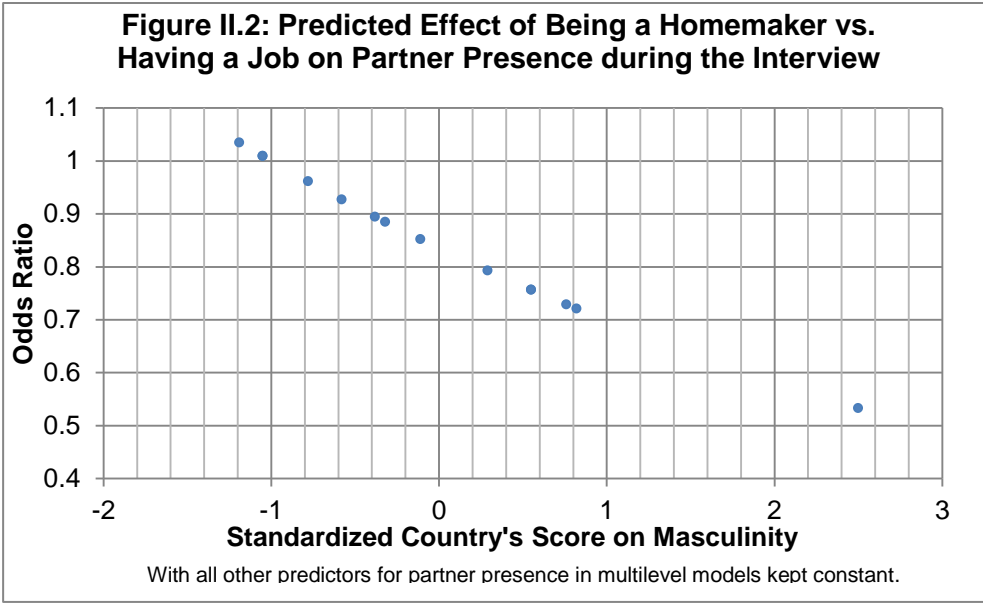
Table II.3: Odds Ratio from Multilevel Logistic Model Predicting Presence of Any Third Party during the Interview (Total N= 27573)^a

	Model 1	Model 2
Sex		
Females	0.87 (0.82–0.92)	0.87 (0.82–0.92)
Males	1.00	1.00
Age in years		
18–34	0.70 (0.62–0.79)	0.70 (0.62–0.79)
35–49	0.70 (0.62–0.78)	0.69 (0.62–0.78)
50–64	0.69 (0.63–0.77)	0.69 (0.63–0.77)
65 and over	1.00	1.00
Marital Status		
Married/Cohabiting	1.17 (1.06–1.28)	1.17 (1.06–1.28)
Divorced/Widowed/Separated	0.73 (0.64–0.82)	0.73 (0.64–0.82)
Single	1.00	1.00
Education Level		
Very low	2.27 (2.05–2.52)	2.27 (2.05–2.52)
Low	1.78 (1.62–1.94)	1.78 (1.63–1.94)
Mid	1.39 (1.28–1.51)	1.39 (1.28–1.51)
High	1.00	1.00
Employment Status		
Homemaker	0.89 (0.80–0.98)	0.89 (0.80–0.98)
Student	1.05 (0.88–1.25)	1.05 (0.88–1.25)
Unemployed/Retired/Other	1.11 (1.03–1.21)	1.11 (1.03–1.20)
Employed	1.00	1.00
Income Level		
Very Low	1.14 (1.03–1.26)	1.14 (1.03–1.26)
Low	1.07 (0.98–1.17)	1.07 (0.98–1.17)
Mid	1.01 (0.92–1.10)	1.01 (0.93–1.10)
High	1.00	1.00
Household Size		
HH size two	2.80 (2.49–3.14)	2.80 (2.49–3.15)
HH size three	3.10 (2.74–3.50)	3.10 (2.74–3.50)
HH size more than three	3.27 (2.89–3.71)	3.27 (2.89–3.71)
HH size one	1.00	1.00
Any 12 month Disorder		
Have a disorder vs. no disorder	1.18 (1.10–1.26)	1.18 (1.10–1.26)
Standardized Individualism Score (IND)	0.63 (0.41–0.96)	1.20 (0.61–2.35)
Level of GNI per capita		
Low	---	4.13 (0.76–22.49)
Middle	---	4.96 (1.44–17.06)
High	---	1.00
Female*IND	0.95 (0.90–1.00)	0.95 (0.90–1.00)
Age 18–34* IND	0.95 (0.86–1.04)	0.95 (0.86–1.04)
Age 35–49* IND	0.94 (0.85–1.02)	0.93 (0.85–1.02)
Age 50–64* IND	0.91 (0.84–1.00)	0.91 (0.84–1.00)
Married*IND	1.09 (1.01–1.17)	1.09 (1.01–1.17)
Divorced/Widowed*IND	1.12 (1.01–1.24)	1.12 (1.01–1.24)
Very low education*IND	1.10 (1.01–1.20)	1.10 (1.01–1.20)
Low education* IND	1.06 (0.98–1.14)	1.06 (0.98–1.14)
Mid education * IND	1.05 (0.98–1.13)	1.05 (0.98–1.13)
HH size two* IND	1.33 (1.20–1.47)	1.33 (1.20–1.47)
HH size three* IND	1.34 (1.20–1.48)	1.34 (1.20–1.48)
HH size more than three* IND	1.35 (1.22–1.50)	1.35 (1.21–1.50)
Between-Country Variance (s.e.)	0.41 (0.18)	0.25 (0.12)
Between-Interviewer Variance (s.e.)	0.59 (0.04)	0.59 (0.04)

Note. a Significant variables with $p < 0.05$ are presented in bold. Dashes “---” indicate variables not included in the model. Models exclude India since no interviewer identification was available. s.e.= standard error; IND= individualism.

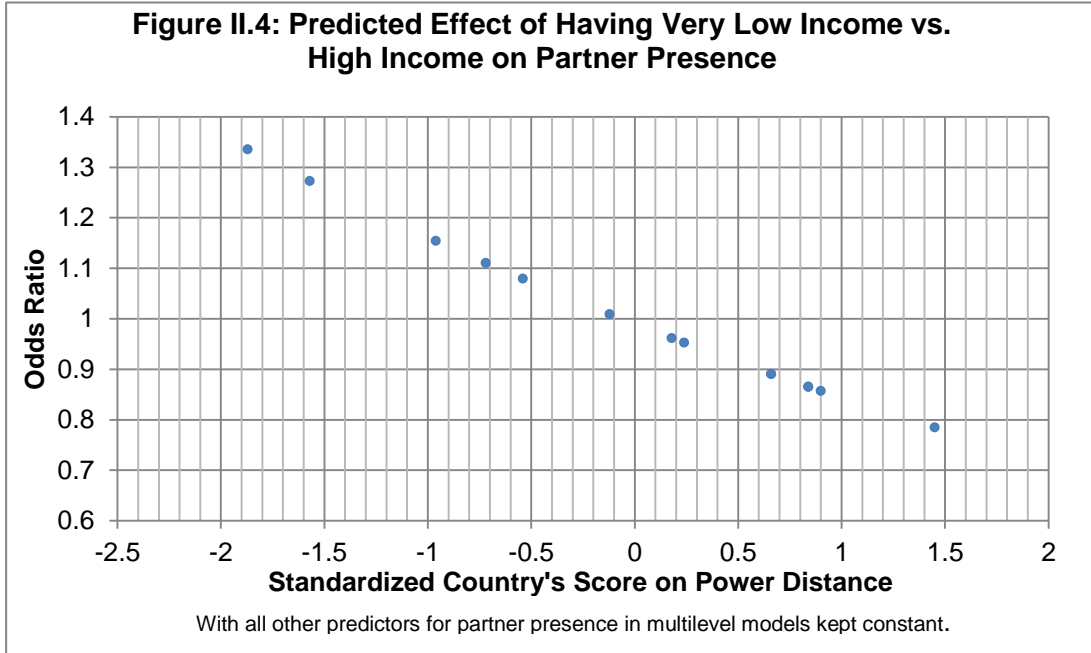
Partner presence (Table II.4).

Cultural variations. The rate of partner presence during the interview also varied significantly by country. Though we hypothesized that the three cultural dimensions—masculinity, power distance, and individualism—are associated with partner presence, only the country’s level of masculinity had a significant main effect on partner presence. As the country’s level of masculinity increased by one standard deviation above the mean, the odds of having a partner present during the interview decreased by 0.68 (at the mean level of masculinity across all countries, Model 1, Table II.4). The country’s level of masculinity also moderated the relationship between partner presence and being a homemaker (as compared to being employed) and between partner presence and the partner’s occupation skill level as hypothesized. While being a homemaker generally *reduced* the odds of having a partner present, having a partner with a low-skill occupation *increased* the odds of his or her presence (compared to having a partner with a high-skill occupation). These differences mattered even more when the country’s level of masculinity increased as is shown in Figures II.2 and II.3. For example, in a country with the mean level of masculinity, respondents who had a partner with a low-skill occupation had 1.4 greater odds to have their partner present during the interview as compared to those who had a partner with a very high-skill occupation. The odds ratio increased to 1.79 in a country with a level of masculinity of one standard deviation above the mean (under a linear interaction assumption).



The country's level of power distance seems to modify the relationship between the household's income level and partner presence as is shown in Figure II.4. In countries with a low power distance score, respondents living in households with very low income were more likely to have a partner present during the interview compared to respondents with a high household income level. Assuming a linear interaction, the odds ratio reached a value of 1.2 in the country with the lowest power distance score (-1.9 standard deviation below the

mean), declining to 0.7 in the country with the highest level of power distance as is shown in Figure II.4.



The third country-level dimension that was investigated is individualism. Though the country's level of individualism did not have a significant main effect on the presence of a partner during the interview, it significantly interacted with age, education, and family size. For example, respondents with a very low education level (as compared to respondents with a high education level) and those living in larger families (more than three members as compared to single families) had 1.59 times and 2.93 times the odds of having a partner present during the interview (at the mean level of individualism across all countries). These differences got larger as the country's level of individualism increased. The interactions effects are summarized in Model 3, Table II.4.

Table II.4: Odds Ratio from Multilevel Logistic Model Predicting Partner Presence during the Interview (Total N=26,316)^a

	Model 1	Model 2 ^b	Model 3
Sex			
Females	0.62 (0.57–0.67)	0.61 (0.56–0.66)	0.61 (0.56–0.66)
Males	1.00	1.00	1.00
Age in years			
18–34	0.73 (0.63–0.85)	0.72 (0.61–0.85)	0.72 (0.61–0.85)
35–49	0.69 (0.60–0.79)	0.67 (0.57–0.78)	0.67 (0.57–0.78)
50–64	0.74 (0.65–0.84)	0.72 (0.63–0.82)	0.71 (0.62–0.82)
65 and over	1.00	1.00	1.00
Education Level			
Very low	1.52 (1.32–1.76)	1.59 (1.36–1.86)	1.59 (1.36–1.86)
Low	1.33 (1.18–1.51)	1.35 (1.18–1.54)	1.35 (1.18–1.54)
Mid	1.22 (1.09–1.36)	1.20 (1.06–1.36)	1.20 (1.06–1.36)
High	1.00	1.00	1.00
Employment Status			
Homemaker	0.84 (0.74–0.95)	0.82 (0.72–0.95)	0.82 (0.72–0.95)
Student	0.64 (0.37–1.12)	0.57 (0.32–1.03)	0.57 (0.32–1.03)
Unemployed/Retired/Other	0.97 (0.87–1.08)	0.96 (0.86–1.08)	0.97 (0.86–1.08)
Employed	1.00	1.00	1.00
Income Level			
Very Low	0.92 (0.80–1.04)	0.93 (0.80–1.08)	0.93 (0.81–1.08)
Low	0.99 (0.88–1.11)	1.01 (0.89–1.15)	1.01 (0.89–1.15)
Mid	0.97 (0.87–1.10)	0.98 (0.87–1.10)	0.98 (0.87–1.10)
High	1.00	1.00	1.00
Household Size			
HH size two	3.20 (2.31–4.45)	3.39 (2.41–4.77)	3.39 (2.41–4.77)
HH size three	3.11 (2.23–4.34)	3.34 (2.36–4.73)	3.34 (2.36–4.73)
HH size more than three	2.72 (1.95–3.80)	2.93 (2.07–4.16)	2.93 (2.06–4.15)
HH one	1.00	1.00	1.00
Spouse Education Level			
Married*Very low	1.11 (0.95–1.29)	1.10 (0.94–1.30)	1.10 (0.94–1.30)
Married*Low	1.11 (0.97–1.27)	1.08 (0.94–1.24)	1.08 (0.94–1.24)
Married*Mid	1.07 (0.94–1.21)	1.06 (0.93–1.21)	1.06 (0.93–1.21)
Spouse Occupation Status			
Married*Spouse does not work	2.00 (1.72–2.33)	2.01 (1.71–2.36)	2.01 (1.71–2.36)
Married*Low skill	1.40 (1.16–1.68)	1.20 (1.00–1.46)	1.20 (1.00–1.45)
Married*Low–average skill	1.19 (1.02–1.38)	1.15 (0.98–1.35)	1.15 (0.98–1.35)
Married*Average–high skill	1.15 (0.93–1.43)	1.06 (0.85–1.33)	1.06 (0.85–1.33)
High skill	1.00	1.00	1.00
Any 12 month Disorder			
Have a disorder vs. no disorder	1.11 (1.02–1.21)	1.13 (1.02–1.24)	1.12 (1.02–1.24)
Individualism Index (IND)	1.03 (0.55–1.90)	0.97 (0.50–1.87)	1.17 (0.55–2.52)
Masculinity Index (MAS)	0.68 (0.46–0.99)	0.70 (0.49–0.99)	0.76 (0.54–1.06)
Power Distance Index (PD)	1.32 (0.76–2.29)	1.17 (0.64–2.12)	0.90 (0.48–1.71)
Level of GNI per capita			
Low	---	---	2.25 (0.40–12.61)
Middle	---	---	3.06 (0.75–12.52)
High	---	---	1.00

(continued)

Table II.4: Odds Ratio from Multilevel Logistic Model Predicting Partner Presence (continued)

	Model 1	Model 2 ^b	Model 3
Age 18–34* IND	0.99 (0.88–1.11)	0.99 (0.88–1.12)	0.99 (0.88–1.12)
Age 35–49* IND	0.90 (0.81–1.01)	0.92 (0.82–1.03)	0.92 (0.82–1.03)
Age 50–64* IND	0.88 (0.79–0.98)	0.88 (0.79–0.99)	0.88 (0.79–0.99)
Very low education*IND	1.17 (1.05–1.31)	1.21 (1.08–1.35)	1.21 (1.08–1.35)
Low education* IND	1.06 (0.97–1.16)	1.10 (1.00–1.21)	1.10 (1.00–1.21)
Mid education * IND	1.07 (0.98–1.17)	1.09 (0.99–1.19)	1.09 (0.99–1.19)
HH size two* IND	1.44 (1.07–1.93)	1.42 (1.05–1.91)	1.42 (1.05–1.92)
HH size three* IND	1.37 (1.02–1.85)	1.39 (1.02–1.88)	1.39 (1.02–1.88)
HH size more than three* IND	1.42 (1.06–1.92)	1.45 (1.07–1.97)	1.46 (1.07–1.98)
Very low income*PD	0.85 (0.75–0.97)	---	---
Low income*PD	0.92 (0.83–1.03)	---	---
Middle income*PD	0.93 (0.84–1.03)	---	---
Homemaker*MAS	0.83 (0.71–0.78)	---	---
Student*MAS	0.77 (0.34–1.75)	---	---
Unemployed/Retired/Other*MAS	0.94 (0.84–1.06)	---	---
Spouse does not work *MAS	1.07 (0.89–1.28)	---	---
Low-skill spouse occupation*MAS.	1.28 (1.02–1.60)	---	---
Low-average skill spouse occupation*MAS	1.06 (0.88–1.27)	---	---
Average-high skill spouse occupation*MAS	1.02 (0.76–1.37)	---	---
Between-Country Variance (s.e.)	0.22 (0.11)	0.23 (0.12)	0.19 (0.11)
Between-Interviewer Variance (s.e.)	---	0.44 (0.04)	0.44 (0.04)

Note. a Significant variables with $p < 0.05$ are presented in bold. Dashes “---” indicate variables not included in the model. b Excludes India since no interviewer information was available. Sample size drops 24,987. Interactions between PDI and MAS and respondent-level characteristics were dropped since once India was excluded they became not significant. Model 1 was replicated without India and these interactions were not significant. s.e.= standard error; IND= individualism; PD=power distance; MAS=masculinity.

Country’s GNI per capita effects. Adding the country’s level of GNI per capita did not affect the interaction effects with the country’s level of individualism, but it did attenuate the effect of masculinity on partner presence and render it insignificant. Contrary to what was hypothesized, the country’s level of GNI per capita did not significantly predict partner presence during the interview.

Interviewer variations. Again, similar to any third-party presence, significant between-interviewer variation existed in the rates of partner presence across all countries. The magnitude of between-interviewer variance was 0.44, larger than the between-country variance of 0.35 (before accounting for any country-level characteristics).

Discussion

Cultural Variations including Country's Wealth Effects

Consistent with the literature, a substantial percentage of interviews were conducted in the presence of a third party across all countries. The lowest rate was in Japan where one in every seven interviews had a third party present. This rate increased to five in seven interviews in India. The variation in non-private interviews between countries was found to be significant, controlling for respondent-level characteristics. Both the country's level of individualism and wealth, as measured by GNI per capita, were significantly associated with the presence of a third party during the interview in the expected direction. The odds of having a third person present during the interview decreased as the country's level of individualism increased. However, once the country's GNI per capita is controlled for, the main effect of individualism was no longer significant. This is not surprising since a country's level of individualism is related to its wealth. In fact, as hypothesized elsewhere, wealth leads to individualism and not vice versa (Hofstede et al., 2010; Smith et al., 2006). When a country's wealth increases, its citizens might become less concerned with survival problems and might give more attention to societal values associated with self-expression and personal choice. Moreover, as wealth increases, resources and commodities become more available, this could allow citizens to become more independent rather than interdependent. Thus, it seems that wealth drives the relationship between individualism and interview privacy.

Variation in partner presence across countries shows a slightly different picture. Though there is significant between-country variation in the rate of partner presence, the relationship between the different cultural dimensions and partner presence was somewhat different than hypothesized. Controlling for the three different cultural dimensions—individualism, power distance, and masculinity—only masculinity was found to have a significant main effect on partner presence. The higher the masculinity of the country, the less likely it is to have a partner present during the interview. The direction of the relationship

could be explained by what defines masculine countries. Masculine countries are characterized by distinct gender roles where men put more emphasis on work and accomplishment and thus are more focused on providing material support while women are more involved with the quality of life and providing emotional support to the family. Such differences in gender roles could affect the availability of partners during the interview, as each is occupied with their distinct roles. This is clearly reflected in the larger partner presence difference between homemakers and employed respondents in more masculine countries compared to less masculine ones, as discussed later.

However, once the country's wealth was controlled for, the main effect of masculinity lost its significance. This is similar to the GNI per capita effect on the relationship between the country's level of individualism and any third-party presence. The country's wealth seems to mediate the effect of both individualism and masculinity with a bigger effect on the former.

In addition to the cultural dimensions' main effects on interview privacy measures, these country-level dimensions also moderate the effect of respondent-level characteristics. Such interaction effects could be explained by the common set of values that define the social role of the different players involved in the interview interaction: the respondent, the third party, and the interviewer. The decrease in interview privacy between females and males in less individualistic countries could be attributed to the general lack of privacy given to females by their family members, including their parents or even their in-laws (Casterline & Chidambaram, 1984). Individualism also moderates the effect of the respondent's educational level on third-person presence. Again, the effect of the respondent's educational level on interview privacy gets larger as the country's level of individualism increases. Respondents with a very low education in a country with China's level of individualism were predicted to have twice the odds of having a third party present during the interview compared to respondents with a high education. The odds ratio is predicted to increase to 2.8 in countries with an individualism level equal to that of the United States.

The country's level of masculinity interacts with the effect of the respondent and partner's socioeconomic characteristics on partner presence.⁹ While the literature mainly uses occupation status as a measure of individuals' availability during the interview (i.e., the proportion of time they are around during the interview), the nature of the occupation also seems to play a significant role in shaping the social interaction between the respondent and his or her partner. Homemakers are usually present at home when other household members are not available and thus are more likely to be interviewed in private compared to employed respondents. The effect gets larger as the country's level of masculinity increases. In less masculine countries like Nigeria, the predicted odds ratio for partner presence when a homemaker was being interviewed, as compared to employed respondents, was 0.96. The corresponding odds ratio in highly masculine countries like Japan was predicted to decrease to 0.53. A similar moderating effect is observed for the partner's type of occupation. Masculinity increases the difference between spouses holding a low-skill occupation and those holding a high-skill occupation. As described earlier, countries that have a high level of masculinity are characterized by a high emphasis on material success and achievement. In such countries, respondents who have an occupation (compared to homemakers) or that have partners that hold high-skill jobs are more focused on financial and work achievements limiting their time availability, their interest, and their attention to nonwork-related matters.

Thus, it seems that the behavior of respondents and other household members and the social interaction between them is greatly influenced by their shared cultural values. To understand the motivations behind the presence of a third party during the interview, researchers need to pay closer attention to the dynamics of the relationship between the respondent and other household members and how it is shaped by prevailing cultural values.

⁹ When all the countries were included in the model including India (Model 2 in Table II.4)

Interviewer Variations

The third major player that influences the interview setting is the interviewer. Researchers greatly rely on interviewers to ensure privacy. Achieving and maintaining privacy usually requires some negotiation between the interviewer and household members. In reality, interviewers are essentially guests, and household members do not normally expect them to dictate the interview setting. Interviewers who have a certain set of social skills, who are friendly negotiators, and who seem trustworthy might be better at achieving privacy. Interviewers' traits and skills, however, could vary. The current analyses constitute the first attempt to measure interviewer variation in achieving a private interview setting. However, since an interpenetrated design was not implemented, it is important to note that it was not possible to isolate the effect of possible geographic clustering from interviewer clustering.¹⁰ Keeping this in mind, between-interviewer variance was estimated across countries and within each country. Findings reveal considerable variation. The estimated between-interviewer variance was even larger in magnitude than the estimated between-country variation, 0.59 compared to 0.50 (before controlling for any country-level characteristics).

Contrary to what was hypothesized, the magnitude of the country-specific between-interviewer variance did not greatly differ from country to country. This could be explained by the standard training protocols that were administered in *all* WMH surveys that did not put a great emphasis on interview privacy, like most commonly reported interviewer trainings (Smith, 1997). Still, some level of variation in *the range of interviewer-privacy rates* exists between countries. Variation in the proportion of private interviews by interviewer between countries ranged from 0% to 69% in the country with the highest between-interviewer variance to 11% to 64% in the country with the lowest interviewer variance.

¹⁰ Information on interviewer assignment in each country was not available. It is possible that part of the between-interviewer variations could be attributed to some geographic clustering related to interview privacy; however, this has never been empirically tested in the literature.

Variation in interview privacy between interviewers could also be attributed to differences in the measurement property of the privacy measures. These measures are based on observations made by interviewers at the end of the interview. Moreover, interviewers are informed during their training that they will be monitored during the field production and evaluated on their performance. It is possible that some interviewers could be underreporting such measures to avoid negative evaluations. Systematic variations in underreporting such measures between interviewers could also contribute to the between-interviewer variance. Unfortunately such differences can't be isolated unless the interview setting is video recorded and interviewer observations are compared against such recordings.

Variation in interview privacy within a sample whether caused by interviewers or by respondent characteristics leads to variation in the interview stimuli that could affect reporting. In fact the effect of interview privacy on reporting sensitive information has been mixed. While some studies found that third-party presence is associated with reduced reporting of undesirable outcomes (Aquilino, 1993; Aquilino, 1997; Aquilino, Wright, & Supple, 2000; Moskowitz, 2004; Taietz, 1962), others observed increased reporting of such outcomes (Bulck, 1999; Edwards, Slattery, & Ma, 1998; Hoyt & Chaloupka, 1994). To reduce such effects, irrespective of their direction, researchers sometimes rely on more private interview modes such as self-administered questionnaires (SAQ) within an interviewer-administered interview, computer assisted self-interviewing (CASI), or audio-computer assisted self-interviewing (A-CASI). While such modes have been effective in reducing interviewer effects on reporting sensitive information (Tourangeau and Smith, 1996), their success in mitigating the effect of third party presence on reporting has not been established. In fact, there are only four published studies that looked at the benefit of using such modes to counter the effect of third-party presence on reporting sensitive information (Aquilino, 1997; Aquilino et al., 2000; Couper,

Singer, & Tourangeau, 2003; Moskowitz, 2004).¹¹ These studies found mixed results. Aquilino attributes this to the possible psychological presence created by the physical presence of the third person even if a more private mode is being used. The interviewing experience still seems to vary between respondents who completed their interviews in private and those who did not.

Thus, it seems that a more certain approach to minimize variations in interview privacy within a sample is to reduce any presence of third party during the interview. One possible way to achieve this and that could also reduce between-interviewer variation in interview privacy is to greatly underscore the importance of interview privacy during interviewer trainings and in all training material (if the study protocol calls for it). Interviewers need to be better trained on how to request, negotiate, achieve, and maintain privacy. To our knowledge, very few studies have emphasized interview privacy in their actual interviewer training and their training material (Smith, 1997). One such example is the National Study of Family Growth (NSFG). The NSFG training manual instructs interviewers that if they believe the setting is not private, they should make another appointment or suggest an alternative location, such as a conference room at the public library, a community center, a porch or patio outside the home, or a restaurant. In fact, in the NSFG Cycle 7, 17% of the interviews were conducted in such settings. The training manual also suggests reminding the third person about the parental consent form that states the importance of conducting the interview in private. Yet, still in cycle 7, about one in eleven interviews were conducted in the presence of a third person (excluding children under the age of 4 years). It seems that more work needs to be done on all the different steps that are needed to guarantee interview privacy: requesting, negotiating, achieving, and maintaining it. Interviewers need to be drilled during

¹¹ In one of these studies (Couper et al., 2003), interviews were conducted in a research lab with volunteers, the privacy setting was randomized, and the third person was a stranger. In another study, interviews were conducted over the phone and respondents were asked whether a third person was present (Moskowitz, 2004). In the remaining two studies, interviews were conducted in respondents' homes and interviewers took observations on whether a third person was present during the interview.

training sessions on different scenarios that they could face and how to address them. Practice sessions on possible techniques tailored to the cultural norms need to be incorporated in interviewer training sessions and material. For example, having interviewers remind the third person about the study confidentiality requirements or having them induce the respondent to request more privacy might be appropriate in certain cultures. In other cultures, however, interviewers might have to repeatedly satisfy the third person's curiosity by answering questions they have, sidetrack the third person, or create an environment that is less encouraging for certain members to be present, such as gender matching interviewers to respondents, or occupying the third party with a self-administered questionnaire designed to collect additional information (possibly on the household) that benefit the research being undertaken.

The findings presented in this chapter should be interpreted with the following considerations in mind. The cultural dimension indices that are used in this chapter come from data collected several years ago. One concern is the applicability of those indices to the WMH data collected between 2000 and 2004 (Pennell et al., 2008). Yet, a number of researchers have shown that while the values of many nations have been changing, the relative positioning of nations has been retained (Hofstede et al., 2010; Ingelhart & Baker, 2000; Schwartz et al., 2000). Moreover, for two of the countries, Lebanon and Nigeria, country-specific values were not available. These two countries were assigned their corresponding regional value (Arab region and West African region, respectively). Some variations, however, may exist within a region. Finally, with respect to between-interviewer variance, it is important to keep in mind the following issues: First, since an interpenetrated design was not implemented and since information on the assignment of interviewers to geographically clustered areas or segments was not available, isolating the possible geographic clustering component from the estimated interviewer variance was not possible. Though, no empirical evidence exists on geographical clustering of interview privacy within a sample, it is conceivable that households clustered in certain geographical sampling units accept and respect interview privacy more than other households

clustered in other units, contributing to the large between-interviewer variance. Second, interview privacy measures were based on interviewer observations. Interviewers are regularly monitored and evaluated in the field. If the study protocol explicitly calls for interview privacy, it is possible that certain interviewers might underreport the presence of a third person to avoid possible negative evaluation. Such misreporting in addition to the possibility of unintentional measurement error in recording the type of third person present might contribute to some of the observed variation. Third, interviewer-level characteristics were unfortunately not available, which hindered explaining possible sources of between-interviewer variation.

Conclusion

This chapter highlights the role of cultural dimensions and country's wealth in explaining interview privacy variations across a wide range of countries. Such country-level characteristics play a direct effect in explaining cultural differences in rates of interview privacy and also shape the effect of the respondent's and other household members' characteristics on privacy. The interviewers' role is also brought back to light and their effect is accentuated. These findings call for practical measures related to training interviewers on requesting, negotiating, achieving, and maintaining interview privacy. However, before such measures are implemented, more quantitative and qualitative data about the interviewer's behavior and attitude toward privacy; the dynamics of the interaction between the respondent, other household members, and the interviewer; and other characteristics of the interview setting need to be collected and studied. Such data could be directly collected by the interviewer or, ideally, through recording the interaction. Investigating those factors will help researchers better understand interviewer variations and cultural influences and help them to develop tailored methods to shape interviewers' attitudes and behaviors toward achieving interview privacy. Most important, collecting more detailed information on the interview setting and the process through which the interpersonal context of the interview is assembled will allow survey researchers to better understand the

effect of third-party presence on the interviewer's behavior and the respondent's cognitive processes when answering survey questions.

Appendix

Table II.5: Percentage of Interviews Conducted in the Presence of a Third Party during the Interview^a

Study	Country	Sample Type	Presence of Third Party during Interview
United States			
Silver, Abramson, and Anderson (1986)	United States	Adult population in 11 National Election Surveys	Range between 41% to 57% (any third party)
Anderson and Silver (1987)	United States	Couple immigrants from Soviet Union to United States	52% (any third party including spouse and children)
Aquilino (1993)	United States	Married couples from the 1987–1988 National Survey of Families and Households	36% (spouse)
Hoyt and Chaloupka (1994)	United States	Youths (14–21) from National Longitudinal Survey of Youth in 1984 and 1988	1984: 5% (parents); 2% (friends); 9% (other household members); 3% (others) 1988: 3% (parents); 2% (friends); 12% (other household members); 3% (others)
Pollner and Adams (1994)	United States	Adult population in Los Angeles site of the Epidemiological Catchment Area Program	48% (any third party)
Aquilino (1997)	United States	Couples 18–45 years old	25% (spouse)
Pollner and Adams (1997)	United States	Couples living together in Los Angeles site of the Epidemiological Catchment Area Program	48% (any third party)
Smith (1997)	United States	Leisure Study, 1975	26% (any third party)
		National Health and Social Life Survey, 1992	21% (any third party)
		General Social Survey, 1994	37% (any third party)
Aquilino, Wright, and Supple (2000)	United States	Adolescents and Adults (12–34) living with parents	Parent presence varied by Interview Section from 6.1% to 10% Sibling presence varied by Interview section from 2% to 3%

(continued)

**Table II.5: Percentage of Interviews Conducted in the Presence of a Third Party
(continued)**

Study	Country	Sample Type	Presence of Third Party during Interview
United States			
Moskowitz (2004)	United States	Adolescents sample ages 12–17 years old	CATI mode: 59% (any third party) T-ACASI: 42% (any third party) ^b
Europe			
Taietz (1962)	The Netherlands	Elderly in two municipalities	78% (child and/or partner)
Reuband (1992)	Germany	Adult population	Range between 20% to 66% (any third party); Average: 37% (any third party)
Reuband (1992)	West Germany	Adult population	28%
Hartmann (1995)	West Germany	Adult population of German General Social Survey from 1980 to 1990	Range between 24% to 35% (any third party)
Bulck (1999)	Belgium	Adult population in Flanders	52% (any third party)
Welkenhuysen-Gybels and Billiet (2001)	Belgium	Flemish adult population ages 18–65 years old	48% (any third party)
Zipp and Toth (2002)	United Kingdom	Couples in the British Household Panel Study	40% among Females; 50% among Males; (spouse)
Developing Countries			
Casterline and Chidambaram (1984)	Developing countries in Asia, Latin America, the Caribbean, and Africa	Females less than 45 years old	Range between 17% (Paraguay) to 82% (Nepal); Majority of Countries (17 out of 24): More than 40%
Podmore, Chaney and Golder (1975)	Hong Kong	Household sample of young population ages 15–29 years old	51% had one or both parents present

Note.^a The majority of the studies summarized above explicitly mention that the design required interviewers to conduct their interviews in private. For the rest, it was not explicitly mentioned in the article, but because they deal with sensitive topics (such as drug use), include a more private mode experiment, or aim to measure the respondent's knowledge, it was assumed that interview privacy was required in their design.

All studies summarized above, except two (Anderson & Silver, 1987, and Welkenhuysen-Gybels & Billiet, 2001), included sensitive questions on a variety of desirable and undesirable outcomes such as reading books, voting behavior, attitudes toward marriage, attitudes toward extended family, TV viewing, mental health, smoking behavior, alcohol and drug use, and intercourse frequency. Since the biggest majority included such sensitive questions and since the range of topics greatly varied within a study, any conclusions about patterns of nonprivate interviews across studies was avoided.

^b Interview privacy measures were collected by asking the respondent the following question over the phone: "For our final question, we would like to know if a parent could hear you being interviewed at any time?"

References

- Anderson, B. A., & Silver, B. D. (1987). The validity of survey response: Insights from interviews of married couples in a survey of Soviet emigrants. *Social Forces*, 66, 537–554.
- Aquilino, W. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly*, 57, 358–376.
- Aquilino, W. (1997). Privacy on self-reported drug use: Interactions with survey mode and respondent characteristics. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (National Institute on Drug Abuse, Monograph 167). Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, & National Institute on Drug Abuse, Division of Epidemiology and Prevention Research.
- Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse*, 35, 845–867.
- Bulck, J. V. (1999). Does the presence of a third person affect estimates of TV viewing and other media use? *Communications*, 24, 105–115.
- Casterline, J., & Chidambaram, V. C. (1984). The presence of others during the interview and the reporting of contraceptive knowledge and use. In J. A. Ross & R. McNamara (Eds.), *Survey analysis for the guidance of family planning programs* (pp. 267–298). Liege, Belgium: Ordina Editions.
- Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly*, 67, 385–395.
- Edwards, S. L., Slattery, M. L., & Ma, K. (1998). Measurement errors stemming from nonrespondents present at in-person interviews. *Annals of Epidemiology*, 8, 272–277.
- GNI per capita, Atlas method (current US\$). (n.d.). Retrieved July 26, 2012, from <http://data.worldbank.org/indicator/NY.GNP.PCAP.CD>
- Harkness, J., Pennell, B., Villar, A., Gebler, N., Auilar-Gaxiola, S., & Bilgen I. (2008). Translation procedures and translation assessment in the World Mental Health Survey Initiative. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.

- Hartmann, P. (1995). Response behavior in interview settings of limited privacy. *International Journal of Public Opinion Research*, 7, 383–390.
- Heeringa, S. G., Elisabeth, J. W., Hubbard, F., Mneimneh, Z. N., Chiu, W., Sampson, N. A., & Berglund, P. A. (2008). Sample designs and sampling procedures. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Culture and organizations: Software of the mind* (3rd ed.). New York, NY: McGraw-Hill.
- Hoppe, M. H. (1998). Validating the masculinity/femininity dimensions on elites from nineteen countries. In G. Hofstede et al. (Eds.), *Masculinity and femininity: The taboo dimension of national cultures* (pp. 29–43). Thousand Oaks, CA: Sage.
- Hoyt, G. M., & Chaloupka, F. J. (1994). Effect of survey conditions on self-reported substance use. *Contemporary Economic Policy*, 7, 109–121.
- Ingelhart, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.
- Ingelhart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65, 19–51.
- Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M. J., Jin, R., Vega, W. A., Walters, E. E., Wang, P., Zaslavsky, A., & Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, 13, 122–139.
- Kessler, R. C., & Üstün, T. B. (2004). The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, 13, 93–121.
- Merritt, A. (2000). Culture in the cockpit: Do Hofstede's dimensions replicate? *Journal of Cross-Cultural Psychology*, 31, 283–301.
- Mooij, M. de, & Hofstede, G. (2001). Convergence and divergence of customer behavior: Implications for international retailing. *Journal of Retailing*, 78, 61–69.

- Moskowitz, J. M. (2004). Assessment of cigarette smoking and smoking susceptibility among youth: Telephone computer-assisted self-interviews versus computer-assisted telephone interviews. *Public Opinion Quarterly*, 68, 565–587.
- Mouritzen, P. E., & Savara J. H. (2002). *Leadership at the apex: Politicians and administrators in western local governments*. Pittsburgh, PA: University of Pittsburgh Press.
- Pennell, B., Mneimneh, Z. N., Bowers, A., Chardoul, S., Wells, J. E., Viana, M. C., Dinkelmann, K., Gebler N., Florescu, S., He, Y., Huang, Y., Toma, T., & Saiz, G. V. (2008). Implementation of the World Mental Health surveys. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.
- Podmore, D., Chaney, D., & Golder, P. (1975). Third parties in the interview situation: Evidence from Hong Kong. *The Journal of Social Psychology*, 95, 227–231.
- Pollner, M., & Adams, R. E. (1994). The interpersonal context of mental health interviews. *Journal of Health and Social Behavior*, 35, 283–290.
- Pollner, M., & Adams, R. E. (1997). The effect of spouse presence on appraisals of emotional support and household strain. *Public Opinion Quarterly*, 61, 615–626.
- Reuband, K. (1992). On third persons in the interview situation and their impact on responses. *International Journal of Public Opinion Research*, 4, 269–274.
- Schwartz, S. H., Bardi, A., & Bianchi, G. (2000). Value adaptation to the imposition and collapse of communist regimes in East-central Europe. In Stanley A. Renshon, & John Duckitt (Eds.), *Political psychology: Cultural and cross-cultural foundations*. London, U.K.: Macmillan.
- Shane, S. A. (1995). Uncertainty avoidance and the preference for innovation championing roles. *Journal of International Business Studies*, 26, 47–68.
- Silver, B. D., Abramson P. R., & Anderson, B. A. (1986). The presence of others and overreporting of voting in American national elections. *Public Opinion Quarterly*, 50, 228–239.
- Smith P. B., Bond, M. H., & Kağıtçıbaşı, Ç. (2006). *Understanding social psychology across cultures: Living and working in a changing world*. London, U.K.: Sage Publications.

- Smith, T. W. (1997). The impact of the presence of others on a respondent's answers to questions. *International Journal of Public Opinion Research*, 9, 33–47.
- Streib, G. F. (1973). Privacy in the research interview: Issues related to design, field tactics and analysis. *Journal of Comparative Family Studies*, 4, 276–285.
- Taietz, P. (1962). Conflicting group norms and the “third” person in the interview. *American Journal of Sociology*, 68, 97–104.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and questions context. *Public Opinion Quarterly*, 60, 275–304.
- Van Nimwegen, T. (2002). *Global banking global values: The in-house reception of the corporate values of ABN AMRO*. Delft, The Netherlands: Ebuorn.
- Welkenhuysen-Gybels, J., & Billiet, J. (2001). The impact of third party presence in survey interviews on the measurement of political knowledge. *Acta Politica*, 36, 287–306.
- Zipp, J. F., & Toth, J. (2002). She said, he said, they said: The impact of spousal presence in survey research. *Public Opinion Quarterly*, 66, 177–208.

CHAPTER III

Cultural Variations in the Effect of Interview Privacy and the Need for Social Conformity on Reporting Sensitive Information

In the second chapter, “Cultural and Interviewer Effects on Interview Privacy,” we presented how the three major players—the respondent, the interviewer, and the third person—affect the privacy setting of the interview. In this chapter, we look at the effect of third-party presence on reporting sensitive and relatively neutral outcomes. The findings from the earlier chapter guides our investigation in this chapter by: 1) highlighting the importance of interviewer variations in determining interview privacy and the need to account for such variations in our models; 2) underscoring the importance of cultural as well as individual-level factors that are related to the respondent and the third party in determining interview privacy; and 3) recognizing possible moderating effects of certain cultural dimensions on third-party presence and reporting survey data. This chapter uses the same cross-national data from the World Mental Health Initiative as in Chapter II with a few countries excluded due to uncollected measures.

Introduction

Many studies require interviewers to conduct their interviews in a private setting. The rationale is that the presence of a third party during the interview might interfere with the response process, possibly causing respondents to misreport information (especially that of a sensitive nature), or to rely on others present during the interview for answers to knowledge questions.

However, establishing interview privacy might not always be feasible, even when the study protocol calls for it. Individual and cultural characteristics of

the respondent, the third party, and the interviewer can affect the process of requesting and negotiating privacy, making it harder for the interviewer to achieve and maintain it during the course of the interview. Using data from the World Mental Health (WMH) Initiative, we have found relatively high rates of non-private interviews across all countries in the range of 13% (Japan) to 70% (India). Other multinational survey programs experienced even a wider range across countries, from a low of 17% (Paraguay) to a high of 82% (Nepal) (Casterline & Chidambaram, 1984).

The relatively common presence of a third party during the interview has led researchers to propose several frameworks describing the conditions required for response effects to occur. Hofhuis (1995) proposed a framework on the effect of third-party presence when responding to knowledge questions. Hofhuis (1995), as reported in Welkenhuysen-Gybel and Billiet (2001), described three conditions that are needed for a response effect to occur when asking knowledge questions. First, a third person needs to be present during the interview and has to be able to hear the survey questions. Second, the third person has to interfere during the interview, and lastly, the third person must be knowledgeable about the subject of the question. Welkenhuysen-Gybel and Billiet (2001) investigated this framework, and found that female respondents interviewed in the presence of an interfering and informed partner gave more correct answers and less “don’t know” answers to 12 political questions compared to female respondents who were interviewed in private. The responses given by these female respondents represent not only their own knowledge but also the knowledge of their partner. Such interference could lead to measurement error if the researcher is aiming to measure only the respondent’s knowledge.

Aquilino (1997) proposed a similar framework that deals with the reporting of sensitive information. Aquilino (1997) described three factors that affect the size and direction of bystanders’ effects on reporting sensitive information. The first is whether the survey question asks about factual or subjective information. The second is whether the bystander is knowledgeable about the factual

information requested. The third is the perceived likelihood that the respondent might experience negative consequences by revealing new and unwelcome information to the bystander. If the third party already knows the information to be reported, then their presence might not have an effect, or it might even lead to more truthful reporting. If the third party does not know the information requested, and the respondent perceives a high likelihood of negative consequences by revealing this information, then the presence of a third party might lead to misreporting. The last two factors are related to the type of relationship that exists between the third party and the respondent. The relationship must be significant to the respondent, and the respondent's answer must bear directly on the relationship for the respondents to change their answers in order to convey a more desirable image in the presence of the third person (Aquilino, 1997; Pollner & Adams, 1997). This has led a number of researchers to investigate the effect of specific types of relationships on reporting sensitive factual and attitude questions.

The most commonly studied types of relationship between a third party and a respondent are parents and spouses. Several studies have investigated the effect of parent or spouse presence during the interview on reporting sensitive *factual* information. The effect of parent presence on reporting substance use among youth and young adults is pretty consistent. Youth and young adults interviewed in the presence of their parents were less likely to report substance use (Aquilino, 1997; Aquilino, Wright, & Supple, 2000; Gfroerer, 1985; Hoyt & Chaloupka, 1994; Moskowitz, 2004). In a meta-analysis conducted by Tourangeau and Yan (2007), the authors concluded that parental presence significantly reduced reporting of sensitive information.

The effect of spouse presence has been less consistent. In one of the studies conducted by Aquilino (1997) among married couples less than 34 years old, spouse presence had no effect on reporting substance use. However, in another sample, Aquilino, Wright, and Supple (2000) found higher rates of substance use among respondents (less than 45 years old) who were interviewed in the presence of their spouse. Casterline and Chidambaram (1984)

studied the effect of third-party presence on reporting contraceptive use in several countries in Latin America, the Caribbean, Asia, and Africa. The authors found that husband presence during the interview reduced the odds of reporting contraceptive use.¹² Pollner and Adams (1994) found that spouse presence reduced the reporting of depression symptoms among adult respondents residing in Los Angeles, but it did not have an effect on reporting other mental-health symptoms.

The effect of third-party presence on reporting *subjective* information has also been mixed. In India, youth and young adults (15–29 years old) interviewed in the presence of their parents reported more positive attitudes toward family (Podmore, Chaney, & Golder, 1975). In the United States, respondents interviewed in the presence of their spouse reported a better quality of marital life (Aquilino, 1997). On the other hand, Anderson and Silver (1987) found that partner presence had no effect on agreement between emigrant Soviet couples when asked about their satisfaction with housing and standard of living.¹³ Pollner and Adams (1997) and Smith (1997) also reported that spouse presence had no effect on respondents' attitudes toward spouse support and satisfaction with household arrangements, and subjective questions on marriage and gender differences, respectively. Pollner and Adams (1997) concluded that the inconsistency of findings and the ambiguity surrounding the interview conditions of some studies indicated that a conclusive judgment about third-party effects on reporting was not quite at hand and needed further investigation.

Several factors might have contributed to the inconsistency of findings on the effect of third-party presence on reporting sensitive information. These

¹² The authors also investigated whether the effect differed by the type of contraceptive method. They specifically compared all contraceptive methods to condom and withdrawal use. Overall, they reported that husband-presence effects for the specific methods resembled the patterns for all methods. In a few countries, however, the reporting of condom use when the husband was present was relatively higher than the reporting of all methods together.

¹³ The design of the survey included interviewing couples separately and simultaneously and interviewing couples separately but not simultaneously. Agreement measures did not significantly differ whether the interviews were simultaneous or not.

factors include inconsistent definitions of third-party presence and duration of presence across studies, interviewer differences in reporting privacy measures, differences in controlling for possible confounding factors, differences in investigating respondent characteristics that could moderate the effect of third-party presence on reporting, and cultural differences across studied populations.

Interviewers usually collect interview privacy measures at the end of the interview by observing the interview setting. As we have shown in our previous chapter, between-interviewer variation in reporting privacy measures has been shown to be high and could possibly vary from one study to another depending on interviewer training protocols and the population studied. Yet, none of the studies investigating the effect of third-party presence on reporting controlled for such possible interviewer variations.

Respondent and third-person characteristics also play a significant role in determining the privacy of the interview. Age, gender, education level, occupation status, and household size are all significantly associated with third-party presence during the interview (Aquilino, 1993, 1997). These same factors are also commonly associated with many of the key outcomes investigated. Differences in controlling for such characteristics could lead to different conclusions.

One of the respondent characteristics that might also affect reporting of sensitive information is the respondent's need for social conformity. Such a need is driven by respondent's motivations and desire to obtain social approval from others (Cialdini & Goldstein, 2004) and minimize possible negative evaluation from them (Johnson & Van De Vijver, 2003). Such conformity desires and motivations could be activated and strengthened depending on contextual stimuli, such as threats to fit in and anonymity of the interaction (Cialdini & Goldstein, 2004). Therefore, it is conceivable that the presence of others could intensify such desires and motivations already held by certain respondents, making them more likely to misreport sensitive information compared to respondents with low conformity motivations. To our knowledge, only one study controlled for a similar construct when investigating the effect of third-party

presence on reporting (Moskowitz, 2004). The author, however, did not investigate the possible moderating effect of the respondent's need for social conformity on interview privacy and reporting.

The need for social conformity in addition to many values and rituals differ by the respondent's cultural background (Cialdini & Goldstein, 2004; Hofstede, Hofstede, & Minkov, 2010). The perception of one's role and status vis-à-vis the role and status of others in any social interaction is unconsciously guided by their cultural background (Hofstede et al., 2010). Thus, to understand people's behavior and the social rules surrounding their behavior, closer attention to the role of culture is needed. This is especially important in cross-cultural studies that aim at comparing findings across cultures.

The relationship between socially desirable reporting and culture has mainly focused on two country-level characteristics: a country's level of collectivism and its wealth.

In a meta-analysis conducted by Bond and Smith (1996), a significant positive relationship was found between the measure of a country's conformity and its level of collectivism. Triandis (1995), as cited in Johnson and Van de Vijver (2003), discussed how honesty in interactions with strangers is valued more in individualist societies while saving face is more salient in collectivist societies. In collectivist societies, the self is defined in terms of relationships with others. To maintain harmony with others, collectivist societies usually pay closer attention to the context of the interaction and to others present in the same social context (Puente-Díaz, 2011; Smith, Bond, & Kağıtçıbaşı, 2006). In such societies, the loss of face of an individual can also cause the loss of face of the group to which they belong. Thus, members that belong to the same in-group have a shared interest in avoiding any loss of face to maintain in-group harmony (Ting-Toomey, 1999). Therefore, respondents' cultural backgrounds might make respondents more concerned about the presence of others during the interview, leading them to misreport information.

A country's level of individualism is also significantly correlated with wealth. It is now established that wealth leads to individualism and not vice versa

(Hofstede et al., 2010; Smith et al., 2006). A country's wealth also affects interview privacy. As we have shown in Chapter II, wealthier countries have been shown to have higher rates of interviews conducted in private. Moreover, a country's level of individualism and masculinity interact with several respondent- and spouse-level characteristics to predict interview privacy. Definitions of these dimensions are provided in the appendix at the end of this chapter.

Thus, to better understand the effect of third-party presence on reporting sensitive information, we need to simultaneously investigate respondent-level characteristics, including the need for social conformity; interviewer variations in reporting third-party presence; and cultural factors that could facilitate or hinder the privacy of the interview. Understanding such effects is important in light of possible difficulties faced by interviewers in attaining interview privacy, and the conceivable psychological effect of third-party presence on the response process even when more private self-administered modes are used (Aquilino, 1997; Moskowitz, 2004).

This chapter examines the effect of third-party presence during the interview on reporting sensitive information by taking into consideration respondent-level characteristics, cultural factors, and interviewer variations. The chapter aims to investigate the following research questions:

1. Does third-party presence affect reporting of sensitive information across cultures?
2. Does the effect of third-party presence on reporting sensitive information vary by a country's level of wealth and a country's level of individualism?
3. Does the effect of third-party presence on reporting sensitive information vary by respondent's motivations for social conformity?
4. Do the effects of third-party presence, cultural factors, and respondent's need for social conformity vary by the type of sensitive question, that is behavioral vs. attitudinal?

From a theoretical point of view, this research will help survey researchers to better understand the effects of interview privacy on reporting sensitive information and resolve some of the inconsistencies in the literature. It emphasizes the importance of respondents' and third party cultural and individual-level characteristics in moderating such effects and the need to account for them in future investigations. From a practical point of view, this research highlights the need for designing well-defined interviewer privacy observations to capture the dynamics of the interview interaction and further investigate its effects on different types of questions.

Methods

Data from the WMH Initiative were used to address these research questions and test the corresponding hypotheses described later in the analysis section.

Sample Designs for the WMH Surveys

The WMH surveys were administered in 24 countries. Of these countries, 9—Brazil, Bulgaria, Japan, Lebanon, Mexico, Nigeria, the People's Republic of China, Romania, and the United States—collected data on the respondent's social-conformity motivations and interview privacy. The current analyses focused on these nine countries.

All WMH surveys targeted the adult population and most of them featured nationally representative probability samples of individuals in households (Heeringa et al., 2008). One (Mexico) was representative of urban areas, one (Nigeria) of selected states, and four (Brazil, India, Japan, and the People's Republic of China) of selected metropolitan areas. Detailed information on the survey sample design is published elsewhere (Heeringa et al., 2008).

To reduce interview length, the WMH interviews were designed to be administered in two parts: Part 1 included core questionnaire sections and Part 2 included non-core sections. All respondents completed Part 1; Part 2 was administered to a subsample of Part 1 respondents. The current analyses

focused on Part 2 respondents since the scale measuring the respondent's need for social conformity and the majority of key outcomes were collected only in Part 2. Country-specific Part 2 samples ranged from a low of 1,031 (Lebanon) to a high of 5,692 (United States).

Questionnaire

The WHO Composite International Diagnostic Interview (CIDI) Version 3.0 was used in all WMH surveys. The CIDI 3.0 is a fully structured interview that generates diagnoses for a wide range of mental-health disorders. It also collects information on treatment behavior, disability, and physical chronic conditions. Detailed questions on social and family life, employment history, finances, and childhood experiences are also included. The questionnaire was translated into each country's local language using the WHO translation guidelines (Harkness et al., 2008; Kessler et al., 2004; Kessler & Üstün, 2004).

Questionnaire Administration

In all countries, trained interviewers conducted face-to-face interviews by using either paper-and-pencil interviewing (PAPI) or computer-assisted personal interviewing (CAPI) methods. Interviewer training in each country was modeled after a centralized 5-day training that project staff from each country had attended. All data was collected before 2008 with the majority of the field work taking place between 2001 and 2003. Detailed information on the specific years of data collection, mode used, response rates, and supervisor-to-interviewer ratio is presented in Chapter II and published elsewhere (Pennell et al., 2008).

Measures Studied in the Current Chapter

As detailed later in the analysis section, a multi-level model for discrete outcomes was used to explore our research questions. Below we list the key binary outcomes, followed by predictors.

Key outcomes. Three categories of key outcome measures were investigated to represent a variety of sensitive behaviors, attitudes, and relatively neutral measures.

Behavioral outcome: CIDI 3.0 included a section on suicidality. In this section, respondents were asked to report if they had ever made a suicide plan or attempted suicide.

Attitudinal outcome: Married respondents were asked to rate their relationship with their current partner on a scale from zero to 10.¹⁴ Since we were interested in investigating the effect of partner presence on reporting high marital ratings (rather than an average increase of one point on the scale), the score was categorized into two groups: a high rating defined as a score above the midpoint of the scale (which is five) and a low rating (five or below).¹⁵

Physical Chronic Condition Outcomes: Respondents were asked to report whether they were ever told by any health professional that they had the following conditions: high blood pressure, asthma, or arthritis.

A number of substantive and sociodemographic predictors were used to investigate their effects on reporting these outcomes.

Respondent-level predictors.

Interview privacy: Interviewer observations on the privacy of the interview setting were collected at the end of the interview. Interviewers were instructed to specify: 1) whether a third person was present at any time during the interview; 2) the relationship of the third party to the respondent (parent, partner, child, youth, or “other adults”); and 3) the duration of the third-party stay during the interview (all the time, most of the time, about half of the time, about

¹⁴ Respondents were asked the following question: Using a scale from 0 to 10 where 0 means “the worst possible marriage and 10 means “the best”, how would you rate your marriage?

¹⁵ The analysis has been replicated with higher cutoffs of six and seven. Results were consistent, with the seven cutoff showing a marginal significance.

one quarter of the time, or less than one quarter of the time). The current analyses focus on any third-party presence (excluding children under the age of six) and partner presence. The duration of the third-party stay was divided into two categories: 1) all of the interview time and 2) some of the interview time.

Respondent's need for social conformity: CIDI 3.0 included an adapted version of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960). Crowne and Marlowe designed the scale to measure the construct they labeled as “the need for social approval.” The CIDI adapted scale consisted of 10 true or false statements such as “I never met a person that I didn’t like” and “I have always told the truth.” Respondents were instructed to choose the answer that first came to their mind and not take too much time thinking before they answered. A value of one was assigned to each endorsed item. The sum of these values formed a score ranging from zero to 10. Within each country, the respondent’s total score was standardized by the country’s average and standard deviation. A score was considered high if it was at least one standard deviation above the national-level mean. This measure will be referred to as the CIDI social conformity scale, and was used to investigate whether there was an interaction between third-party presence and a high score on this scale on reporting sensitive information.¹⁶ It is important to note that the Marlowe-Crowne Scale has its own limitations, including being susceptible to measurement distortion. Unfortunately, this was the only scale available in the WMH surveys for measuring variations in the respondent’s motivations for social conformity. A discussion on the limitations of the scale is included later in this chapter.

Respondent sociodemographic predictors: These variables were treated as control variables and included the respondent’s gender, age (18–34, 35–49, 50–64, older than 64 years), marital status (never married, currently married or cohabiting, previously married), education level (high, middle, low,

¹⁶ Interaction between the respondent’s need for social conformity and the cultural dimensions investigated could have also been researched. However, this was not one of the main hypotheses of the chapter, and most important, the scale was not designed to be compared across cultures and has been criticized for possible measurement inequivalence across cultures.

very low), income level (high, middle, low, very low), current employment status (employed, studying, taking care of home, other), and household size (fewer than two, two, three, more than three). Among those currently married or cohabiting, the partner's level of education (high, middle, low, very low) and type of occupation (not employed, have a low-skill job, low-to-average skill job, average-to-high skill job, and high-skill job) were also accounted for.

Interviewer-level predictors: Interviewer identification numbers were available for all nine countries. This information enabled the modeling of random effects (intercepts) for the individual members of the interviewing force in each country (details provided in the analysis section). However, since no other information on interviewers was available, it was not possible to examine the influence of interviewer characteristics, such as gender, on the variation in outcome rates.

Country-level predictors. Two cultural dimensions were included in the current analyses: individualism and masculinity. Definitions for these dimensions are provided in the appendix at the end of this chapter. Country-specific scores for each of the two dimensions and more details on their assessment are also available in Hofstede, Hofstede, and Minkov (2010). For each dimension and for each country, a standardized score was calculated based on the average score and the standard deviation across all the countries included in the analyses. Higher scores indicated higher levels of the underlying dimension.

Finally, the countries' economic strength and standard of living was measured by its Gross National Income (GNI) per capita. GNI measures in nominal dollar values that were calculated according to the Atlas Method for the year the data was collected were used in this chapter.¹⁷ Because of the limited

¹⁷ According to the WHO, the Atlas method "applies a conversion factor that averages the exchange rate for a given year and the two preceding years, adjusted for differences in rates of inflation" ("GNI per capita," n.d.).

number of countries and the big gaps in the GNI measure across the nine countries, the countries were divided into three categories: high GNI per capita, middle GNI per capita, and low GNI per capita. This classification matches the World Bank classification of countries by income.

Analysis

Because of the limited number of countries, a two-level logistic regression with respondents nested within interviewers was used for all outcomes. Interviewers were treated as random effects. All predictors were treated as fixed effects including cultural dimensions. Countries were not included as fixed effects due to running into a low rank design matrix. Country effects were adjusted for through the three cultural factors individualism, masculinity, and GNI per capita.¹⁸ The analysis was repeated using a three-level model with respondents at level 1, interviewers at level 2, and countries at level 3 and the results with respect to our main hypotheses were *consistent* but were less stable because of the limited number of countries in the analyses. Thus, the analyses and the findings reported below were based on the two-level model.

For suicidal behaviors and physical chronic condition outcomes (high blood pressure, asthma, or arthritis), main predictors included any third party present during all of the interview time (vs. none of the time), any third party present during some of the interview time (vs. none of the time), the respondent's level on the CIDI social conformity scale (high vs. low), the country's level of individualism, and the country's level of GNI per capita (middle vs. high and low vs. high). The country's standardized masculinity score as well as respondent demographics and socioeconomic characteristics were also included in all the models. Interactions between third-party presence measures and a high score on the CIDI social conformity scale, and between third-party presence measures

¹⁸ These dimensions significantly predict interview privacy and render between-country variation non-significant as shown in Chapter II.

and the country's level of individualism and wealth were tested. Only significant interactions (with a p-value less than 0.05) were kept in the final model.

These models were designed to test the following hypotheses:

Suicide behavior outcome.

Hypothesis 1: Third-party presence is inversely related to the likelihood of reporting suicidal behavior, controlling for respondent demographics and socioeconomic characteristics and the country's level of individualism, masculinity, and wealth.

Hypothesis 2: There is an interaction between the respondent's need for social conformity and third-party presence. Reporting differences due to third-party presence are larger among respondents with a high need for social conformity, controlling for all other variables in the model.

Hypothesis 3: There is an interaction between the country's level of individualism and third-party presence. The effect of third-party presence on reporting less suicidal behavior is reduced as the country's level of individualism increases, controlling for all other variables in the model.

Hypothesis 4: There is an interaction between the country's level of wealth and third-party presence. The effect of third-party presence on reporting less suicidal behavior is accentuated in countries with a middle and low GNI per capita (compared to countries with a high GNI per capita), controlling for all other variables in the model.

Physical chronic conditions outcome.

Hypothesis 5: Third-party presence is not significantly related to the likelihood of reporting any of the physical chronic conditions, controlling for respondent demographics and socioeconomic characteristics and the country's level of individualism, masculinity, and wealth.

Hypothesis 6: There is no significant interaction between the respondent's need for social conformity and third-party presence on reporting any of the physical chronic conditions, controlling for all other variables in the model.

Hypothesis 7: Neither the country's level of individualism nor the country's level of wealth significantly moderates the effect of third-party presence on reporting any of the physical chronic conditions, controlling for the rest of the variables.

Marital rating outcome. Only five out of the 9 countries collected marital ratings. Substantive predictors for the marital outcome rating included partner presence, the respondent's level on the CIDI social conformity scale (high vs. low), and the country's level of GNI per capita (middle vs. high and low vs. high). The country's level of individualism was not included in the model since two of the five countries had the same level of individualism, and there was not enough variation to explore. Respondent sociodemographic characteristics as well as the presence of any other third party (other than a partner) were also included in all the models. Interactions between partner presence measures and a high score on the CIDI social conformity scale, and between partner presence measures and the country's level of wealth were tested separately. Only significant interactions (with a p-value less than 0.05) were kept in the final model. These models were designed to test the following hypotheses:

Hypothesis 8: Partner presence is positively related to the likelihood of reporting a high marital rating, controlling for respondent demographics and socioeconomic characteristics, other bystanders' presence, and the country's level of wealth.

Hypothesis 9: There is an interaction between the respondent's need for social conformity and partner presence. Reporting differences due to partner presence are larger among respondents with a high need for social conformity, controlling for all other variables in the model.

Hypothesis 10: There is an interaction between the country's level of wealth and partner presence. The effect of partner presence on reporting a high marital rating is accentuated in countries with a middle and low GNI per capita (as compared to countries with a high GNI per capita), controlling for all other variables in the model.

All multilevel models presented in this chapter are unweighted. To explore the effect of weights on the findings, the analyses were replicated using weighted and unweighted single-level logistic regression models. The weights accounted for within-country differential probability of selection, post-adjustment to the country's sociodemographic distributions and subsampling to specific questionnaire sections. Adjusted and unadjusted point estimates were consistent (with single-level models being more stable) suggesting that the use of weights to adjust for informative sample design was not necessary for the relationships investigated. All analyses were conducted using SAS version 9.2 (SAS institute, NC) using PROC GLIMMIX.

Results

Outcome Rates and Interview Privacy Rates across Countries

Table III.1 presents the different outcome rates in each of the countries included in the analyses. Weighted rates for combined suicide plan or suicide attempt reports differed greatly between countries with the lowest rates, 1.0%, reported in Nigeria, and Romania, and the highest rate, 7.5%, reported in the United States.

Only five out of the nine countries collected ratings of respondents' current marital relationships. High marital ratings were reported by the majority of respondents in all five countries.¹⁹ All rates were higher than 88%.

¹⁹ A high marital rating as defined earlier in the Methods section is a reported value greater than the midpoint of the scale (i.e. 5/10).

All countries included questions on high blood pressure, asthma, and arthritis. Reported rates of asthma were generally low, mainly less than 5.5%, except in the United States where 11.6% of the respondents reported having asthma. Arthritis and blood pressure were more common and rates were more variable across countries. Arthritis rates range from 7.0% (Lebanon) to 33.3% (Romania) and high blood pressure rates range from 4.3% (Lebanon) to 24.1% (United States).

Table III.1: Weighted Outcome Percent (s.e.)

Country	Suicide Plan or Attempt		High Marital Rating ^a		Physical Chronic Conditions ^b			
	N	Percent (s.e.)	N	Percent (s.e.)	N	High Blood Pressure Percent (s.e.)	Asthma Percent (s.e.)	Arthritis Percent (s.e.)
Bulgaria	2,233	1.3 (0.3)	---	---	2,206	20.6 (1.1)	1.9 (0.3)	9.3 (0.7)
Brazil	2,942	6.5 (0.5)	1,836	88.2 (1.1)	2,929	19.9 (1.1)	2.2 (0.3)	7.6 (0.7)
China	1,628	1.5 (0.2)	1,228	93.3 (1.0)	1,625	14.0 (1.1)	3.5 (0.7)	11.6 (1.1)
Japan	1,682	2.8 (0.4)	---	---	1,275	16.1 (1.3)	5.3 (0.9)	9.2 (0.9)
Lebanon	1,031	2.8 (0.4)	696	89.9 (1.3)	601	4.3 (0.9)	1.2 (0.5)	7.0 (1.2)
Mexico	2,362	4.1 (0.3)	---	---	2,362	9.7 (0.8)	2.2 (0.4)	7.5 (0.7)
Nigeria	2,143	1.0 (0.1)	1,411	93.1 (0.9)	2,138	3.0 (0.5)	0.6 (0.2)	16.9 (0.9)
Romania	2,356	1.0 (0.3)	---	---	2,350	17.2 (0.9)	3.0 (0.4)	33.3(1.2)
USA	5,692	7.5 (0.3)	1,601	91.6 (1.0)	5,689	24.1 (0.8)	11.6 (0.5)	27.3 (0.8)

Note. ^a Among married respondents. Dashes (---) refer to unavailable data

^b Differences in sample size between Suicide Plan or Attempt column and this column are either due to subsampling into the chronic condition section or missing data.

s.e.= standard error

On average, 37% of the interviews were conducted in the presence of any third person. Table III.2 presents the rates of third-party presence and duration of stay in each of the nine countries. In most countries, a third party was commonly present for part of the interview rather than the entirety of the interview (the average rate across countries was 26%). On average, only 11% of the interviews had a third party present during *all* of the interview time.

Partners were present in 19% of the interviews across all the countries. Again, partners were mostly present for some parts of the interview rather than the whole interview. Country-specific rates are presented in Table III.3.

Table III.2: Percentage of Any Third Party Presence (s.e.)

Country	N	No Third Party Present During Interview	Any Third Party Present All Interview Time	Any Third Party Present Some of the Interview Time
Bulgaria	2,232	60.7(1.0)	13.2(0.7)	26.1(0.9)
Brazil	2,942	41.0(0.9)	19.7(0.7)	39.3(0.9)
China	1,628	63.9(1.2)	12.8(0.8)	23.3(1.0)
Japan	1,346	87.5(0.9)	2.7(0.4)	9.8(0.8)
Lebanon	1,031	33.6(1.5)	21.5(1.3)	44.9(1.5)
Mexico	2,350	65.4(1.0)	10.7(0.6)	23.9(0.9)
Nigeria	2,141	68.0(1.0)	7.1(0.6)	24.9(0.9)
Romania	2,356	64.6(1.0)	15.2(0.7)	20.2(0.8)
USA	5,304	70.0(0.6)	5.0(0.3)	25.0(0.6)

Note. Values are unweighted estimates of sample % (standard error).

Table III.3: Percentage of Partner Presence (s.e.)

Country	N	No Partner Present during the Interview	Partner Present All Interview Time	Partner Present Some of the Interview Time
Bulgaria	2,232	73.9(0.9)	8.5(0.6)	17.6(0.8)
Brazil	2,942	73.3(0.8)	9.6(0.5)	17.1(0.7)
China	1,628	80.1(1.0)	7.1(0.6)	12.8(0.8)
Japan	1,346	93.8(0.7)	1.5(0.3)	4.7(0.6)
Lebanon	1,031	68.3(1.5)	11.1(1.0)	20.6(1.3)
Mexico	2,350	88.6(0.7)	4.0(0.4)	7.4(0.5)
Nigeria	2,141	91.6(0.6)	1.7(0.3)	6.6(0.5)
Romania	2,356	79.3(0.8)	8.3(0.6)	12.4(0.7)
USA	5,304	82.9(0.5)	2.9(0.2)	14.2(0.5)

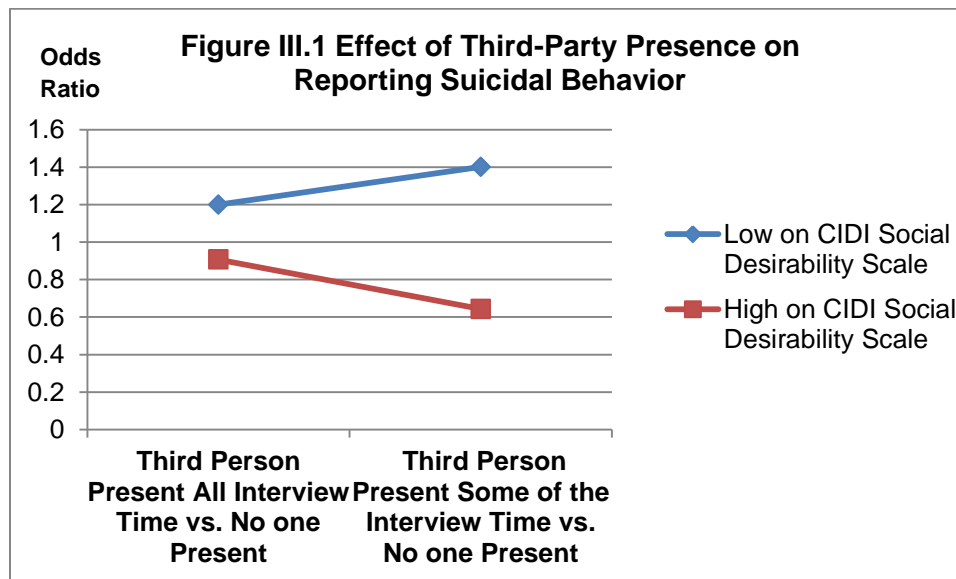
Note. Values are unweighted estimates of sample % (standard error).

Effect of Third-Party Presence on Reporting: Results from Multilevel Logistic Regression Model

Suicidal behavior (Table III.4 and Figure III.1). As hypothesized, the respondent's need for social conformity moderated the relationship between third-party presence and reporting suicidal behavior. Ignoring this interaction would give a misleading picture of the direction of third-party presence effects.

Respondents who scored *high* on the CIDI social conformity scale and who had a third person present during the interview were less likely to report suicidal behavior compared to those who had no one present. When a third person was present all of the interview time or some of the interview time and the respondent scored high on the CIDI social conformity scale, the odds of reporting suicidal behavior was reduced to 0.91 and 0.64 respectively compared to not having a third party present during the interview.

The picture was quite different among respondents who scored *low* on the CIDI social conformity scale. Among those respondents, having a third person present all or some of the time was associated with increased reporting of suicidal behavior compared to being interviewed alone (OR=1.20 and 1.40 depending on whether a third person was present all or some of the interview time).



Contrary to what was hypothesized, however, neither the country’s level of individualism nor the country’s level of wealth moderated the effect of third-party presence on reporting suicidal behavior when all other variables were controlled for.

Table III.4: Odds Ratio and 95% Confidence Interval from Multilevel Logistic Model Predicting Suicide Attempt or Plan (N=21,329)^a

	Main Model 2	Interaction Model
Presence of Third Party		
Third Party Present All of the Time	1.20 (0.99–1.43)	1.20 (0.98–1.46)
Third Party Present Some of the Time	1.31 (1.16–1.45)	1.40 (1.23–1.60)
No Third Party Present	1.00	1.00
Social Conformity Score		
High Score ^b	0.69 (0.59–0.82)	0.82 (0.67–1.01)
Low Score	1.00	1.00
Individualism Standardized Score (IND)	2.72 (2.26–3.28)	2.72 (2.26–3.28)
Country’s GNI per capita		
Low	12.78 (6.54–24.97)	12.74 (6.52–24.88)
Middle	12.49 (6.96–22.41)	12.45 (6.94–22.34)
High	1.00	
Present All of the Time* High Social Conformity	---	0.92 (0.55–1.54)
Present Some of the Time* High Social Conformity	---	0.56 (0.38–0.82)

Note. ^a Significant odds ratios with p<0.05 are presented in bold. Dashes (---) indicate variables not entered in the model. All models control for sex, age, marital status, education level, income level, employment status, household size, and the country’s score on masculinity.

^b High score is greater or equal to one standard deviation above the mean.

Marital ratings (Table III.5). High ratings of marital relationships were positively associated with partner presence during the interview, controlling for other respondent-level characteristics and the country’s level of wealth (see Main Model, Table III.5). However, the effect was only significant among respondents interviewed in a country with middle GNI per capita (compared to high GNI per capita, specifically the United States). Respondents who were interviewed in a middle-income country and in the presence of their partner during the whole interview had more than twice the odds of reporting a high marital rating score compared to those who had no partner present (see Interaction Model, Table III.5). Unlike suicidal behavior and contrary to what was hypothesized, the

association between partner presence and reporting a high marital rating was not moderated by the respondent's need for social conformity.

Table III.5: Odds Ratio and 95% Confidence Interval from Multi-level Logistic Model Predicting High Marital Rating Score (N=6,595)^a

	Main Model	Interaction Model
Presence of Partner		
Partner Present All of the Time	1.59 (1.08–2.35)	0.51 (0.24–1.10)
Partner Present Some of the Time	1.36 (1.07–1.73)	1.51 (0.92–2.50)
No Partner Present	1.00	1.00
Social Conformity Score		
High Score ^b	1.55 (1.17–2.05)	1.53 (1.16–2.02)
Low Score	1.00	1.00
GNI per capita		
Low	1.65 (1.19–2.30)	1.58 (1.01–2.28)
Middle	0.72 (0.52–0.99)	0.68 (0.47–0.98)
High	1.00	1.00
Partner Present All of the Time* Low GNI per capita	---	2.44 (0.78–7.68)
Partner Present Some of the Time* Low GNI per capita	---	1.02 (0.46–2.25)
Partner Present All of the Time* Middle GNI per capita	---	4.43 (1.79–10.92)
Partner Present Some of the Time* Middle GNI per capita	---	0.84 (0.47–1.50)

Note. ^a Significant odds ratios with $p < 0.05$ are presented in bold. Dashes (---) indicate variables not entered in the model. All models control for sex, age, marital status, education level, income level, employment status, and household size.

^b High score is greater or equal to one standard deviation above the mean.

Physical chronic conditions (Table III.6). Unlike suicidal behavior and marital relationship rating, reporting high blood pressure, asthma, or arthritis was not significantly associated with third-party presence (see Table III.6). Moreover, as hypothesized, no significant interaction effects were found between third-party presence and the respondent's need for social conformity or with any of the country-level characteristics for any of the three outcomes.

Table III.6: Odds Ratio and 95% Confidence Interval from Multilevel Logistic Model Predicting Chronic Conditions^a

	High Blood Pressure N=20,482 OR(CI)	Asthma N=20,516 OR(CI)	Arthritis N=20,446 OR(CI)
Presence of Third Party			
Third Party Present All of the Time	1.10 (0.97–1.25)	0.78 (0.61–1.01)	0.95 (0.82–1.09)
Third Party Present Some of the Time	1.07 (0.99–1.20)	1.02 (0.88–1.19)	1.07 (0.97–1.18)
No Third Party Present	1.00	1.00	1.00
Social Conformity Score			
High Score ^b	0.93 (0.84–1.03)	1.03 (0.87–1.22)	0.90 (0.81–1.00)
Low Score	1.00	1.00	1.00
Individualism Standardized Score	2.08 (1.78–2.43)	1.74 (1.40–2.17)	1.28 (1.05–1.55)
Country's GNI per capita			
Low	4.06 (2.41–6.84)	1.07 (0.49–2.33)	1.08 (0.56–2.05)
Middle	6.35 (3.88–10.39)	1.03 (0.51–2.10)	0.47(0.25–0.86)
High	1.00	1.00	1.00

Note. ^a Significant odds ratios with $p < 0.05$ are presented in bold. Gray indicates variables not entered in the model. All models control for sex, age, marital status, education level, income level, employment status, household size, and the country's score on masculinity.

^b High score is greater or equal to one standard deviation above the mean.

Discussion

This chapter is the first to investigate whether the effect of third-party presence on reporting sensitive information is moderated by the respondent's need for social conformity and the respondent's cultural background. Three types of outcomes were investigated: sensitive behavioral outcomes (suicidal behavior), sensitive attitudinal measure (marital rating), and relatively neutral measure (having high blood pressure, arthritis, or asthma). As we hypothesized, although third-party presence affects the reporting of both types of sensitive outcomes, behavioral and attitudinal, it does not affect the reporting of neutral outcomes. Third-party presence effects are moderated by the respondent's need for social conformity and the respondent's cultural background. Though such moderating effects were hypothesized for both outcomes, they did differ depending on whether the outcome is behavioral or attitudinal.

For sensitive behaviors, specifically suicidal behavior, having a third party present during the interview was associated with lower odds of reporting such behavior *only* among respondents who had *high* scores on the CIDI social conformity scale. Having someone present during the interview might create a

contextual stimulus that strengthens the respondent's already existing needs for social conformity and increases their perceived likelihood of experiencing negative consequences upon revealing such behavior in the presence of a third party. To avoid such consequences, respondents might prefer not to disclose such information in their presence. The picture is quite different among respondents who do not have such needs and who scored low on the CIDI social conformity scale. Among those respondents, the likelihood of reporting suicidal behavior is higher in the presence of a third party compared to being interviewed in private. Respondents who scored low on social conformity might not have high concerns about possible negative consequences associated with divulging such information, or they might have already confided in the third person present during the interview. Among such respondents, a third party who is present during the interview might act as a truth control increasing the reporting of such behavior. This has been shown in two studies investigating sensitive undesirable outcomes in the United States (Aquilino, 1997; Hoyt & Chaloupka, 1994). Future research on the effect of information already held by the third party on reporting sensitive information and how it interacts with respondent's need for social conformity is needed. This is further discussed in the conclusion chapter.

The higher reporting of suicidal behavior among respondents who scored low on the CIDI social conformity scale could also be explained by the tendency of household members to look after family who had experienced suicidal behaviors if they are already informed about such behaviors. Respondents who had such suicidal experiences might be more likely to have a family members present during the interview compared to respondents who did not have such experiences. Since the presence of third party during the interview was not randomized and a causal effect can't be posited, we cannot conclude whether- among respondents who scored low on the CIDI social conformity scale- the presence of third party lead to increased reporting of suicide behaviors, or whether respondents who had suicidal experiences were more likely to have a third member present during the interview.

In addition to the respondent's social conformity, we hypothesized that the respondent's cultural background moderates the effect of third party-presence on reporting suicidal behavior. However this interaction effect was not found to be significant. It seems that though third-party presence is more common in countries with a low level of individualism or in less wealthy countries, as has been shown in Chapter II, the effect of third-party presence on reporting suicidal behavior does not change across those dimensions.

The second sensitive outcome investigated is respondents' attitudes toward their current marriages. As hypothesized, partner presence is positively associated with reporting a high marital rating. Respondents who are interviewed in the presence of their partner might be more motivated to provide positive characterization of their relationship compared to those interviewed in private. This finding has been documented by Aquilino (1997). However, when investigating the effect across cultures, the effect was only significant among respondents interviewed in countries with middle GNI per capita (compared to high GNI per capita). Though the direction of the relationship is the same for respondents interviewed in countries with low GNI per capita (compared to high GNI per capita), the effect was not significant. Again, when interpreting the increased reporting of high marital ratings among respondents who are interviewed in the presence of their partner, it is important to keep in mind the non-causal nature of the association. It is possible that in middle and low income countries where family interdependence is high, keeping company of a partner is a sign of a good marital relationship. Thus, respondents who "truly" have good marital relationships might be more likely to have their partner present in general and during the interview compared to respondents whose relationship with their partner is not as good.

The absence of the hypothesized moderating effect of the respondent's high need for social conformity on partner presence and reporting a high marital rating could be explained by the nature of the marital rating measure. Unlike suicide behavior, which is a factual measure, marital rating is subjective. For suicide behavior- and as discussed earlier- respondent who scored low on social

conformity might have already confided in other household members about their suicide experiences. Thus among such a group, respondents might be more inclined to report their “true” suicidal behavior in the presence of third party (compared to being interviewed in private). This however might not be the case for a measure such as marital rating because it is not observable and it does not have a “true” value.

The third set of outcomes investigated in this chapter is chronic physical conditions: high blood pressure, arthritis, and asthma. These outcomes are normally perceived as more neutral compared to suicidal behavior and marital rating measures. Thus, they were chosen for the purpose of contrasting the effect of third-party presence on reporting sensitive information to its effect on reporting more neutral outcomes. As hypothesized, the presence of a bystander during the interview does not significantly affect reporting of such neutral outcomes across the different cultures studied.

In summary, the presence of a third party during the interview is associated with reporting of sensitive outcomes. The effect can go in either direction depending on the respondent’s need for social conformity, the respondent’s cultural background, and the type of question. Whether it is a positive or a negative effect, the presence of a third party during the interview seems to add some measurement variation among respondents belonging to the same sample. This variation is another layer of error that is attributed to the implementation feature of the interview and that needs to be minimized. Researchers sometimes try to minimize the association between third-party presence and reporting by using self-administered modes. Though the use of self-administered modes on reducing interviewer effects has been established in the literature (Tourangeau & Yan, 2007), whether such modes reduce the strength of the association between third-party presence and reporting is still under debate. Only four studies used randomized mode experiments and investigated the effect of third-party presence among each randomized group (Aquilino, 1997; Aquilino et al., 2000; Couper, Singer, & Tourangeau, 2003;

Moskowitz, 2004).²⁰ The results of these studies have been mixed. The presence of a third party during the interview, especially if the third party is a household member, might subtly alter the respondent's frame of mind when answering sensitive questions. Even when using a self-administered mode, the interview might not feel as private when a third party is present compared to when the respondent is interviewed alone. In fact, Aquilino et al. (2000) reported that third party effects are found even when the bystander did not interfere with the interview or communicate with the respondent. Still, due to the limited amount of empirical evidence, further research is needed on the moderating effect of the interview mode on third party presence and reporting sensitive information before any conclusions are made.

Therefore, and as discussed in Chapter II, first of all, survey practitioners need to better train interviewers on how to achieve and maintain interview privacy and the effect of such training interventions on interview privacy need to be measured. Second, researchers are encouraged to collect more data on the interview setting, its dynamics, and information held by the third person to better understand its effect. Third, if such effects are replicated, researchers need to take these measures into consideration when analyzing their data and control for them in their substantive investigations. This becomes even more crucial in comparative cross-cultural research where the possibility of achieving a private interview and its effect on reporting sensitive information could vary from one culture to another.

The findings presented in this chapter should be interpreted with the following considerations in mind. First, the presence of a third party at respondents' home during the interview is very difficult to control by the

²⁰ In one of these studies (Couper et al., 2003), interviews were conducted in a research lab with volunteers, the privacy setting was randomized, and the third person was a stranger. In another study, interviews were conducted over the phone and respondents were asked whether a third person was present (Moskowitz, 2004). In the remaining two studies, interviews were conducted in respondents' homes and interviewers took observations on whether a third person was present during the interview.

researcher and can't be randomized. Thus, causal interpretations of any of the findings can't be made and only associations are reported.

Second, interview privacy measures are based on interviewer observations reported at the end of the interview. Third party duration of stay during the interview is an overall measure for the whole interview and is not section specific. Thus, in instances where the interviewer reported someone to be present during some of the interview time, it was not possible to determine whether the bystander was actually present when the target question was asked. This could explain some of the differences observed between the effect of the third party's duration of stay on reporting the different sensitive outcomes. While information on suicidal behavior is collected toward the first half of the interview, marital rating is collected toward the end of the second half. The third party duration of stay is also collected as an overall measure for all of the bystanders that might have been present during the interview. Thus, when investigating partner presence, if the interviewer indicated that another person was present in addition to the partner, the duration of stay was assigned to all of the different bystanders present. However, this should not significantly impact the findings as the large majority of interviews (83%) where a partner was present had no other bystander also present.

Third, the respondent's need for social conformity is measured using an adapted version of the Marlowe-Crowne Scale. Though this scale has been extensively used in the literature, and though the direction of the association between the respondent's need for social conformity and reporting outcomes is in the expected direction, such scales are not free from measurement limitations. Self-reported scales that measure the respondent's need for social conformity usually are not immune to misreporting, and their accuracy might vary depending on individual as well as cultural factors. Moreover, such scales assume that social conformity is stable from the time of its measurement; however as discussed later in Chapter IV, social conformity is contextual and could vary during the course of the interview.

Fourth, the cultural-dimension indices used in this chapter come from data published in Hofstede et al. (2010), some of which were collected several years ago. One concern is the applicability of those indices to the WMH data. Yet, a number of researchers have shown that while the values of many nations have been changing, the relative positioning of those nations has been retained (Hofstede et al., 2010; Inglehart & Baker, 2000; Schwartz, Bardi, & Bianchi, 2000).

Fifth, and finally, though many factors known to be associated with third party presence and the investigated outcomes have been controlled for, it is conceivable that there could be other unmeasured factors that we did not control for and which could have affected the relationship observed between third-party presence and reporting of sensitive outcomes.

Conclusion

Reporting sensitive information is affected by more than the topic itself. The respondent's personal characteristics and cultural values, the social context in which the topic is communicated, and the players involved in the communication affect the response process during an interview. For a given topic, such factors affect whether the respondents interpret the content as socially desirable or undesirable and whether they decide to misreport information or not. This chapter demonstrates that the effect of the interview privacy setting on reporting sensitive information is moderated by the need for social conformity and the prevailing cultural background.

Thus, survey researchers and practitioners are encouraged to formulate a better understanding of the dynamics surrounding interview privacy, how it is affected by respondents' and third parties' personal characteristics, cultural background, and its potential effects on different types of survey questions. To achieve this, future work on designing measures to capture information already held by the third party, as well as well-defined interviewer privacy observations are highly needed. Such well-defined measures need to then guide both practical interventions on training interviewers to better achieve and maintain interview

privacy, and empirical work on the possible moderating effects of personal, cultural, and other interview setting factors on the response process of sensitive outcomes.

Appendix

Cultural Dimensions Definitions

Individualism: “Individualism pertains to societies in which the ties between individuals are loose; everyone is expected to look after his or her immediate family. Collectivism as its opposite, pertains to societies in which people from birth onwards are integrated into strong, cohesive in-groups, which through people’s lifetime continue to protect them in exchange for unquestioning loyalty” (Hofstede, Hofstede, & Minkov, 2010).

Masculinity: “A society is called masculine when emotional gender roles are clearly distinct: men are supposed to be assertive, tough, and focused on material success, whereas women are supposed to be more modest, tender, and concerned with the quality of life. A society is called feminine when emotional gender roles overlap: both men and women are supposed to be modest, tender, and concerned with the quality of life” (Hofstede et al., 2010).

References

- Anderson, B. A., & Silver, B. D. (1987). The validity of survey response: Insights from interviews of married couples in a survey of Soviet emigrants. *Social Forces*, 66, 537–554.
- Aquilino, W. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly*, 57, 358–376.
- Aquilino, W. (1997). Privacy on self-reported drug use: Interactions with survey mode and respondent characteristics. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (National Institute on Drug Abuse, Monograph 167). Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, & National Institute on Drug Abuse, Division of Epidemiology and Prevention Research.
- Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse*, 35, 845–867.
- Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, 119, 111–137.
- Bradburn, N. M., & Sudman, S. (1979). *Improving interview method and questionnaire design*. San Francisco, CA: Jossey-Bass.
- Cialdini, R.B., and Goldstein, N.J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, 55, 591-621
- Casterline, J., & Chidambaram, V. C. (1984). The presence of others during the interview and the reporting of contraceptive knowledge and use. In J. A. Ross & R. McNamara (Eds.), *Survey analysis for the guidance of family planning programs* (pp. 267–298). Liege, Belgium: Ordina Editions.
- Couper, M.P., Singer, E., and Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly*, 67 (3), 385-395.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354.

- Gfroerer, J. (1985). Influence of privacy on self-reported drug use by youths. In B. A. Rouse, N. J. Kozel, & L. G. Richards (Eds.), *Self-report methods of estimating drug use: Meeting current challenges to validity*. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, & National Institute on Drug Abuse.
- GNI per capita, Atlas method (current US\$). (n.d.). Retrieved July 26, 2012, from <http://data.worldbank.org/indicator/NY.GNP.PCAP.CD>
- Harkness, J., Pennell, B., Villar, A., Gebler, N., Auilar-Gaxiola, S., & Bilgen I. (2008). Translation procedures and translation assessment in the World Mental Health Survey Initiative. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.
- Hartmann, P. (1995). Response behavior in interview settings of limited privacy. *International Journal of Public Opinion Research*, 7, 383–390.
- Heeringa, S. G., Elisabeth, J. W., Hubbard, F., Mneimneh, Z. N., Chiu, W., Sampson, N. A., & Berglund, P. A. (2008). Sample designs and sampling procedures. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.
- Hofhuis, H. (1995). *Invloed van derde aanwezigen tijdens interviews*. Utrecht, Netherlands: ISOR.
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Culture and organizations: Software of the mind* (3rd ed.). New York, NY: McGraw-Hill.
- Hoyt, G. M., & Chaloupka, F. J. (1994). Effect of survey conditions on self-reported substance use. *Contemporary Economic Policy*, 7, 109–121.
- Ingelhart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65, 19–51.
- Johnson, T. P., & Van De Vijver, F.J.R. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. J. R. van de Vijer, & P. P. Mohler (Eds.), *Cross-cultural survey methods*. Hoboken, NJ: Wiley.

- Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M. J., Jin, R., Vega, W. A., Walters, E. E., Wang, P., Zaslavsky, A., & Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 122–139.
- Kessler, R. C., & Üstün, T. B. (2004). The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 93–121.
- Moskowitz, J. M. (2004). Assessment of cigarette smoking and smoking susceptibility among youth: Telephone computer-assisted self-interviews versus computer-assisted telephone interviews. *Public Opinion Quarterly*, *68*, 565–587.
- Pennell, B., Mneimneh, Z. N., Bowers, A., Chardoul, S., Wells, J. E., Viana, M. C., Dinkelmann, K., Gebler, N., Florescu, S., He, Y., Huang, Y., Toma, T., & Saiz, G. V. (2008). Implementation of the World Mental Health surveys. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.
- Podmore, D., Chaney, D., & Golder, P. (1975). Third parties in the interview situation: Evidence from Hong Kong. *The Journal of Social Psychology*, *95*, 227–231.
- Pollner, M., & Adams, R. E. (1994). The interpersonal context of mental health interviews. *Journal of Health and Social Behavior*, *35*, 283–290.
- Pollner, M., & Adams, R. E. (1997). The effect of spouse presence on appraisals of emotional support and household strain. *Public Opinion Quarterly*, *61*, 615–626.
- Puente-Díaz, R. (2011). Context effects: The role of collectivism as a moderator. *International Journal of Psychology*, *46*, 55–62.
- Reuband, K. (1992). On third persons in the interview situation and their impact on responses. *International Journal of Public Opinion Research*, *4*, 269–274.
- Schwartz, S. H., Bardi, A., & Bianchi, G. (2000). Value adaptation to the imposition and collapse of communist regimes in East-central Europe. In Stanley A. Renshon, & John Duckitt (Eds.), *Political psychology: Cultural and cross-cultural foundations*. London, U.K.: Macmillan.

- Silver, B. D., Abramson P. R., & Anderson, B. A. (1986). The presence of others and overreporting of voting in American national elections. *Public Opinion Quarterly*, 50, 228–239.
- Smith P. B., Bond M. H., & Kağıtçıbaşı, Ç. (2006). *Understanding social psychology across cultures: Living and working in a changing world*. London, U.K.: Sage Publications.
- Smith, T. W. (1997). The impact of the presence of others on a respondent's answers to questions. *International Journal of Public Opinion Research*, 9, 33–47.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco, CA: Jossey-Bass.
- Taietz, P. (1962). Conflicting group norms and the “third” person in the interview. *American Journal of Sociology*, 68, 97–104.
- Ting-Toomey, S. (1999). *Communicating across cultures*. New York, NY: Guilford Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- Van de Vijver, F. J. R. (2003). Bias and equivalence: Cross-cultural perspectives. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods*. Hoboken, NJ: Wiley.
- Van Hemert, D. A., van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences*, 33, 1229–1249.
- Welkenhuysen-Gybels, J., & Billiet, J. (2001). The impact of third party presence in survey interviews on the measurement of political knowledge. *Acta Politica*, 36, 287–306.

CHAPTER IV

Bridging Psychometrics and Survey Methodology: Can Mixed Rasch Models Identify Socially Desirable Reporting Behavior?

Introduction

Socially desirable reporting is a response behavior exhibited by respondents as a means to project a favorable image of themselves that maximizes social conformity and minimizes negative judgments by others (Johnson & Van De Vijver, 2003). Respondents usually manifest such behavior when they perceive the information they are requested to provide as sensitive to disclose.²¹

Several factors lead respondents to perceive survey responses as sensitive to disclose. These factors can be divided into four main sets related to the respondent, the question, the study design, and the survey implementation. Chapter I described these sets in detail and how they interact together to affect socially desirable response behavior. One of the respondent-level factors that plays an important role in socially desirable reporting is their need for social conformity. Several scales have been designed to measure respondents' need for social conformity. These scales come from the psychology literature and include the Wiggins Social Desirability Scale (Wiggins, 1959), the Edwards Social Desirability Scale (Edwards, 1957), the Marlowe-Crowne Scale (Crowne & Marlowe, 1960), the Eysenck Personality Inventory Lie Scale (Eysenck & Eysenck, 1964), and the Balanced Inventory of Desirable Reporting (Paulhus, 1984). The Marlowe Crowne Scale remains the most commonly used scale in

²¹ Though the definition of sensitive question is much broader (Tourangeau & Yan, 2007), throughout this chapter we use the term "sensitive" to refer specifically to questions asking respondents about desirable or undesirable information.

the literature (Beretvas, Meyers, & Leite, 2002; Johnson & Van de Vijver, 2003). Several concerns have been raised about the use of these scales, including whether they measure a “pure” representation of response distortion and whether they are comparable across cultures when used in cross-national research (Zickar & Robie, 1999). Most important, the use of these scales requires the assumption that from the time the scale is administered, the respondent’s need for social conformity is stable across all the survey questions. However, as discussed by social-psychologists, one’s social conformity needs may be contextual (Caldini & Goldstein, 2004). Such needs could be activated and strengthened depending on contextual stimuli, such as factors related to the design of the study or the actual interview implementation. If the stimuli are not consistent across the survey interview, then the respondent’s score on such scales might change. One of the important stimuli that could strengthen or dampen such needs affecting the respondent’s response behavior is the survey questions themselves. Characteristics of survey questions could directly and/or indirectly affect the respondent’s perception of the sensitivity of the information to be disclosed, ultimately leading to socially desirable response behavior. Thus, not only might certain respondents be more inclined to respond in a socially desirable manner than others, but the *same* respondent that might exhibit a socially desirable response behavior on one sensitive question might not do so on another. Therefore, investigating socially desirable reporting behavior requires the development of *respondent-level* measures that capture the respondent’s response behavior on the *specific question(s)* that is(are) susceptible to social-desirability bias. Though few authors have *theoretically* addressed the simultaneous effect of question characteristics and respondent characteristics on social-desirability bias (Cannell, Miller, & Oksenberg, 1981; DeMaio, 1984), most *empirical* literature has studied them separately. In fact, the psychological literature is mainly focused on respondent-level characteristics and the survey method literature is mainly focused on question-level characteristics.

The lack of *topic-specific respondent-level* measures of the need for social conformity has recently led researchers to explore new measurement methods.

One method that has been used to study faking behavior is mixture Rasch models.²² This model combines a Rasch item response theory (IRT) model with a latent class model (Rost, 1990). These models were initially proposed in the educational field to study differences in response patterns among examinees for reasons such as the differential use of solution strategies and instructional sensitivity and isolate them from the construct measure. These models are usually applied to scales composed of several items (or questions) measuring the same construct (such as mathematical ability, reading ability, etc.). The central idea is to group respondents into latent classes based on response pattern differences (caused by differential use of solution strategies) that maximize between-class heterogeneity in the probability of endorsing scale items. The probability of endorsing an item is reflected in an item-specific parameter called item difficulty. An item with a lower probability of being endorsed is an item that is more “difficult.” When answering questions about personal traits, behaviors, or attitudes (as opposed to knowledge questions), the probability of endorsing an item (its item difficulty) partially depends on the level of perceived desirability (or undesirability) of the item. The higher the level of desirability of the item, the lower the level of difficulty and thus the easier it is to endorse the item (Zickar & Ury, 2002). Such differences in the perceived desirability or undesirability of the item among respondents could be caused by respondents’ differential needs for social conformity leading to differences in response patterns when answering scales of a sensitive nature. Thus, if mixture IRT models classify respondents into latent classes based on response pattern differences that maximize between-class heterogeneity in item difficulty, and if item difficulty is partially measuring the perceived desirability or undesirability of the item, these latent classes could represent differences in social conformity leading to socially desirable reporting behavior.

²² Faking is “an individual’s conscious attempt to present misleading and deceptive information about his or her personality, interests, experiences, past behaviors, and attitudes with the goal of influencing others” (Kuncel & Borneman, 2007). According to social-psychologists, socially desirable responding could be either a conscious or unconscious behavior exhibited by respondents to maximize social conformity and minimize negative judgments by others. Faking and socially desirable responding could be viewed as regions on a continuum (Kuncel & Borneman, 2007).

In fact, a number of researchers have used this method to study the faking behaviors of different populations on measures of personality and anger expression (Eid & Zickar, 2007; Gollwitzer, Eid, & Jürgensen, 2005; Holden & Book, 2009; Majj-de Mejj, Kelderman, & van der Flier, 2008; Zickar, Gibby, & Robie, 2004). Though faking does not strictly represent socially desirable reporting behavior, both reporting phenomena could be viewed as regions of the same continuum (Kuncel & Borneman, 2007). Thus, the handful of literature on mixture IRT and faking could offer insights into the study of socially desirable reporting among survey respondents.

The greatest concern in using such latent classification is the validation of the nature of classes. As Rost (1990) pointed out, it is important to distinguish between classes that are measuring real structural differences related to the construct itself from those measuring response behaviors. To validate the nature of classes and confirm that they are measuring faking behavior, several methods have been used in the literature. Among those methods is the manipulation of the instructions related to answering interview questions. In this experimental method, respondents are randomized into one of two groups, a group that is instructed to answer honestly and a group that is instructed to fake “good” or fake “bad” (Holden & Book, 2009; Zickar et al., 2004). This design, however, has been criticized for at least two reasons. First, the design could exacerbate the effect of response distortion beyond what would be expected under realistic circumstances (Smith & Ellingson, 2002). Second, it asserts that all individuals who are instructed to answer in a certain way will actually follow those instructions, which is not necessarily true.

Another method that has been used to validate that the latent classes represent true differences in faking behavior is to contrast a group of respondents who have a relatively high motivation to fake reporting (such as job applicants) to those who have a lower motivation to fake (such as incumbents). It is possible, however, that situational influences as well as motivational influences (such as the desire for a job, the need for a job, and satisfaction with the current job) greatly affect whether respondents fake the measures they are asked to

answer (Zickar et al., 2004). This could lead certain groups who are expected to fake (such as applicants) to respond in a relatively honest manner and others who are expected to be honest (such as incumbents) to fake answers (Eid & Zickar, 2007).

Finally, other studies have used ancillary information to explore the nature of respondent classes (Gollwitzer et al., 2005; Majj-de Meij et al., 2008). The most common type of information that has been used is personality scales. The main concern here is that these scales themselves suffer from response distortions and are not necessarily good validation tools. These scales are also designed to measure a respondent trait that is assumed to be constant across the entire interview and is not question or scale specific.

One way to investigate socially desirable reporting behavior among respondents by using mixture item response theory models and that addresses the above limitations is to have a design that : 1) studies socially desirable reporting behavior in a community sample where respondents are interviewed in their natural environment; 2) uses external independent and objective measures of the design and interview implementation that have been established to affect reporting by reducing interview privacy or increasing it; and 3) uses multiple objective survey measures to examine the nature of the latent classes.

This chapter addresses the limitations of prior approaches by using the design described above. Socially desirable reporting behavior was investigated by using mixed Rasch models on a community sample where respondents were interviewed in their own homes. Multiple scales with a wide spectrum of items (desirable, undesirable, and relatively neutral) were used to contrast the results. To explore the nature of the latent classes, their association with interview mode and interview privacy that could interact with social conformity affecting response behavior was investigated.

This chapter aims to answer the following three research questions:

1. On a scale measuring a sensitive substantive topic, is it possible to identify classes of respondents who exhibit different socially desirable response patterns based on their answers to desirable and undesirable scale items?
2. Is class membership associated with the mode of the interview?
3. Is class membership associated with interview privacy?

In investigating these aims, this chapter capitalizes on advances in mixture IRT models to establish topic-specific measures of social conformity that have been lacking in the field of survey methodology. Such an approach could also allow for estimating measures of item desirability (or undesirability). Ultimately, and using mixed Rasch models as the backbone, computer adaptive measurement tools that estimate in real time the respondent's response patterns resembling socially desirable response behavior could be designed. This information could be used to route each respondent to the next scale items that fits their estimated level of construct or to administer intervention messages that aim to reduce their social-conformity concerns, reducing measurement variance (with the former) and measurement bias (with the latter) while also minimizing the number of scale items that need to be asked compared to a static instrument.

Methods

This chapter uses data from the 2002–2003 Lebanon World Mental Health (WMH) study. This study is part of the World Mental Health Initiative, which aims to study the prevalence of mental health disorders, their risk factors, treatment, and burden across different parts of the world (Kessler & Üstün, 2004).

Sample Design of Lebanon World Mental Health Study

The Lebanon-WMH survey was based on a nationally representative multistage clustered area probability household sample of Lebanese adults (ages

18 and over). Detailed information on the sample design is published elsewhere (Karam et al., 2006). The overall survey response rate was 70.0%. To reduce interview length, certain questionnaire sections were administered to subsamples of respondents. The scales that were used in the current analyses were administered to a subsample of 527 (out of 2,857) adults. Weights were used to account for differential probability of selection in questionnaire sections; details are found in the analysis section.

Questionnaire

Two questionnaires were used in the Lebanon-WMH survey. The first was the WHO Composite International Diagnostic Interview (CIDI) version 3.0, which was the standard instrument used in all WMH surveys. The CIDI 3.0 is a fully structured interview designed to generate diagnoses for a wide range of mental-health conditions. Detailed questions on social and family life, employment history, finances, and demographics are also included. CIDI 3.0 was translated into Arabic from the original English-language source using the WHO translation guidelines (Harkness et al., 2008; Kessler et al., 2004; Kessler & Üstün, 2004).

The second questionnaire used was the Lebanese Arabic Temperament Evaluation Memphis, Pisa, and Santiago-Auto Questionnaire (TEMPS-A). TEMPS-A was a local addition to the CIDI 3.0 and was not administered in other WMH surveys. TEMPS-A is composed of “True” or “False” items designed to measure the respondents’ types of affective temperament (Karam, Mneimneh, Salamoun, Akiskal, & Akiskal, 2005). The scale was also translated into Arabic from the original English TEMPS-A version (Akiskal, Akiskal, Haykal, Manning, & Connor, 2005).

Questionnaire Administration

The CIDI 3.0 was administered face-to-face by trained lay interviewers using a paper-and-pencil (PAPI) mode. The TEMPS-A was the last section of the interview and was designed to be self-administered. However, 48% of the respondents requested the interviewers to ask them the questions out loud and

score the scale. Reasons provided by the interviewers for administering the scales were 1) the respondent does not want to read (47%); 2) the respondent cannot read (25%); 3) the respondent is old and has difficulty reading (12%); and 4) other (15%). Other reasons included the respondent being busy with kids, the respondent is tired, and the respondent did not have eyeglasses handy.

Measures Studied in the Current Chapter

The following measures were included in the current analyses:

Affective temperaments. Socially desirable reporting behavior was investigated for each of the following TEMPS-A subscales: hyperthymic, depressive, and anxious. A description and complete list of items included in each temperament subscale can be found in Appendix A at the end of this chapter. Between 2.4%-3.6% of the respondents had a missing answer on any of the items constituting each of the three subscales. These respondents were dropped from the analysis.

The hyperthymic and depressive subscales are each made up of 20 “True” or “False” statements. The anxious subscale is composed of 26 items. Respondents were instructed to read each statement and answer “True” or “False.” A “True” answer meant that the statement reflected how the respondent feels, thinks, or behaves most of the time since reaching adulthood (18 years and over). Statements in the hyperthymic and depressive subscales included a mixture of desirable and undesirable behaviors, thoughts, or feelings. The majority of the hyperthymic statements described desirable behaviors or traits, but the majority of the depressive statements described undesirable ones. Examples of desirable statements included “I have great confidence in myself” and “I am the kind of person you can always depend on.” Examples of undesirable statements included “When I disagree with someone, I can get into a heated argument” and “For as long as I can remember, I’ve felt like a failure.” The anxious subscale, however, included relatively neutral statements, such as

“Many people have told me not to worry so much” and “I frequently have difficulty falling asleep.”

The scales were chosen to reflect a spectrum of questions—desirable, undesirable, and neutral—and to contrast the findings of each of them.

Respondent’s need for social conformity. The CIDI 3.0 included an adapted version of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964). The adapted scale that was used in the current analyses included nine “True” or “False” statements (see Appendix B). Respondents were instructed to choose the answer that first came to their minds and not to take too much time thinking before they answered. Out of 527 respondents, 56 had one or more missing answers to items on the scale. A sequential logistic regression model using IVEWare was applied to impute answers for each missed item from the observed answers for the rest of the scale items.²³ Single imputation was used. For each endorsed (observed or imputed) item, a value of one was given. The sum of these values formed a score ranging from zero to nine. The summary score was categorized into three groups: low social desirability score (less than one standard deviation below the sample mean), medium score (between one standard deviation below and above the mean), and high score (one standard deviation above the sample mean). This scale is referred to as the CIDI social conformity scale throughout the rest of the chapter.

Interview privacy. Interviewer observations on the privacy of the interview setting were collected at the end of the interview. Interviewers were instructed to specify: 1) whether a third person was present during the interview at any time; 2) the relationship of the third party to the respondent (parent, partner, child, youth, or “other adults”); and 3) the duration of the third party’s stay during the interview (all of the time, most of the time, about half of the time,

²³ IVEware was developed by the researchers at the Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan (Institute for Social Research, University of Michigan, 2011)

about one quarter of the time, or less than one quarter of the time). The current analyses investigated the presence of parents, partners, and others (excluding children under the age of 6). The duration of the third party's stay was categorized into all of the interview time, some of the time, or none of the time.

Sociodemographic variables. Sociodemographic characteristics included in the analyses were the respondent's gender, age (18–34, 35–49, 50–64, older than 64 years), marital status (never married, currently married or cohabiting, previously married), education level (high, middle, low, very low), household income level (high, middle, low, very low), and current employment status (employed, studying, taking care of home, others). These measures were controlled for in the analyses to account for their possible confounding effects as many of them are related to the interview mode in the current sample and have been reported to be associated with socially desirable reporting (Fisher & Dube, 2005; Hebert et al., 1997; Klassen, Hornstra, & Anderson, 1975). Moreover, a number of these measures have been found to be significantly associated with scores on the hyperthymic and depressive temperament (Karam et al., 2005). While such associations could reflect true sociodemographic differences in these scores, they could also be confounded by possible misreporting differences due to social desirability concerns. Thus to account for such possible confounding effects, these variables were controlled for in the analyses.

Analysis

Simple item response theory (IRT) model. The simplest IRT model, also called a Rasch or one-parameter model, is built around the central idea that the probability of giving a specific answer to an item (on a scale) is a logistic function of the relative distance between the item location, also referred to as item difficulty (β_i) and the person location on the construct (θ_j) being measured (Hambleton, Swaminathan, & Rogers, 1991) (see Equation IV.1 below). An

example would be a scale with items measuring substantive constructs like mathematical ability, pain, and depression.

$$P(X_{is} = 1 | \theta_s, \beta_i) = \frac{\exp(\theta_s - \beta_i)}{1 + \exp(\theta_s - \beta_i)}$$

item($i = 1, 2, \dots, n$) *subject*($s = 1, 2, \dots, m$) $X \in (0, 1)$ $\theta = \text{construct}$ $\beta_i = \text{item_difficulty}$

Equation IV.1

The “difficulty” parameter represents the extent to which the item taps into the underlying construct being assessed. Items that are less difficult are those that tap into lower levels of the construct and thus have a higher probability of being endorsed. However, items that are more difficult tap into higher levels of the construct and thus are endorsed at lower probabilities, especially at low construct levels. Figure IV.1 below presents a hypothetical item characteristic curve for three items with increasing level of difficulty.²⁴ For example, consider the hyperthymic temperament as the construct of interest. Both of the following items constitute features of the hyperthymic temperament, “I am always on the go” and “I can get enough with less than 6 hours of sleep at night.” The first statement could be thought of as describing a low-difficulty item (the green curve in Figure IV.1) and the second as a high-difficulty item (the black curve in Figure IV.1). Thus, the first item would be shifted to the left side of the x-axis and the second statement would be shifted to the right side of the x-axis as illustrated in Figure IV.1.

²⁴ An item characteristic curve is a monotonically increasing function that describes the relationship between the probability of endorsing an item and the overall construct level (underlying the item performance).

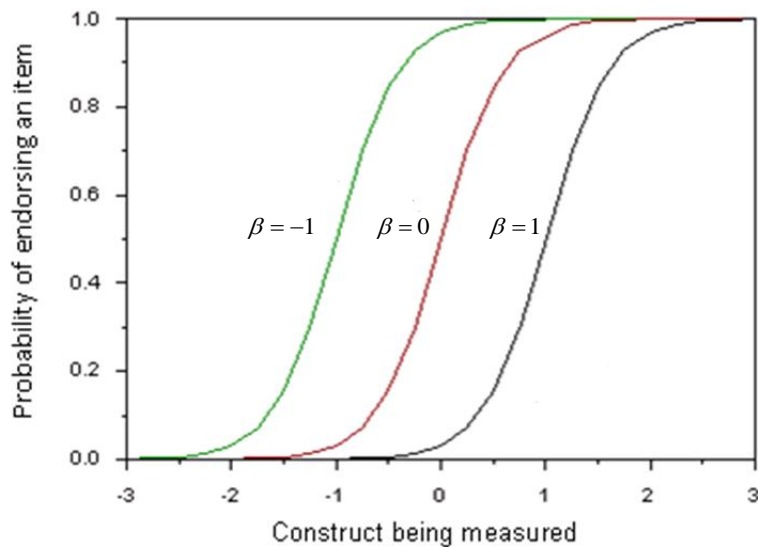


Figure IV.1: Item Characteristic Curves for Three Items with Increasing Level of Difficulty

When dealing with knowledge questions, such as those used in the educational field (the main user of IRT models), the above “classical” definition of item “difficulty” and its estimation process seem to be well suited. However, this definition might be incomplete when dealing with other types of questions (such as behavior, traits, and attitudes). The level of perceived item desirability (or undesirability), which is a possible characteristic of non-knowledge questions, could also shift the difficulty parameter to the right (if the item is relatively undesirable) or to the left (if the item is relatively desirable) irrespective of the level of construct it taps into (Zickar & Ury, 2002). To illustrate, consider two items that reflect different features of the hyperthymic temperament, “I have a gift for speech, convincing, and inspiring others” and “I am the kind of person who likes to be the boss.” Theoretically, these two items tap into *the same level* of the hyperthymic construct and thus should have similar item difficulty estimates.²⁵ The behavior or trait described by these two statements, however, could have different perceived levels of desirability. This difference in the desirability level could shift the estimated difficulty parameter. If the first item is perceived to be more desirable than the second, then the estimated difficulty of the first item is

²⁵ Personal communication with the psychiatrist author of the Lebanese TEMPS-A

expected to be lower than the estimated difficulty of the second item (though theoretically they represent the same level of construct).

For a *specific item*, this shift in the estimated difficulty that is attributed to the perceived item desirability could be occurring in certain respondents more than others. Respondents who believe in the existence of social norms that define desirable attitudes or behaviors and who want to conform to those norms could be more sensitive to the level of item desirability (or undesirability). Therefore, the estimated item difficulty for the group of respondents with such activated needs for social conformity is expected to be higher than in the group who do not have such strong needs. To capture such group differences, a mixture model that *controls for the level of the substantive construct* and combines item response theory with latent class analysis is needed.

Mixed Rasch models. Rost (1990) described a mixed Rasch model (MRM) that identified classes of respondents based on their patterns of responses to substantive scale items in a way that maximizes between-class heterogeneity in item difficulties. These differences in response patterns could be due to differential use of solution strategies (e.g. cross-multiplication algorithm strategy) and instructional sensitivity among respondents answering the same knowledge scales. Two respondents who might have *a similar overall score on the substantive construct* (θ_s) could be assigned to different classes because of differences in their strategies, i.e., pattern of responses (Zickar et al., 2004). The classes are mixed in the respondent pool according to the proportion of each latent class (γ_h); the difficulty measure (β_{ih}) for the i th question is now class-specific. The sum of proportions is equal to one. Rost's (1990) MRM is presented below in Equation IV.2:

$$P(X_{is} = 1 | \theta_s, \beta_{ih}) = \sum_h \gamma_h \frac{\exp(\theta_s - \beta_{ih})}{1 + \exp(\theta_s - \beta_{ih})}$$

item ($i = 1, 2, \dots, n$) *subject* ($s = 1, 2, \dots, m$) *latent_group* ($h = 1, 2, \dots, v$) $X \in (0, 1)$ $\theta = \text{construct}$

$\gamma_h = \text{latent_group_proportion}$ $\beta_{ih} = \text{group_specific_item_difficulty}$

Equation IV.2

Respondent latent classes estimation. As discussed earlier, when answering questions about personal traits, behaviors, or attitudes (as opposed to knowledge questions), the probability of endorsing an item (its item difficulty) partially depends on the level of perceived desirability (or undesirability) of the item. Differences in the perceived desirability or undesirability of the item among respondents could be caused by respondents' differential needs for social conformity leading to differences in response patterns when answering scales of a sensitive nature. In this chapter, such respondent differences are investigated by using MRM on each of the three temperament scales: depressive, hyperthymic, and anxious. These scales were chosen to represent differences in the type of items (desirable, undesirable, and relatively neutral). While the majority of the hyperthymic and depressive scale items described either desirable or undesirable behaviors or traits, the anxious temperament items described relatively neutral items. Thus, although we expect respondent differences in the probability of endorsing desirable or undesirable items for the hyperthymic and depressive temperament, we do not expect such differences for the anxious temperament.

To investigate the existence of latent classes of respondents who differ in the probability of endorsing items on each of the three temperament scales—depressive, hyperthymic and anxious—four models were estimated. These models included a simple IRT model (i.e., one-class), two-class, three-class, and four-class mixed Rasch models.

Goodness of fit indices were compared for the different models, and the best-fitting model was chosen based on the lowest BIC (Bayesian Information Criterion) and CAIC (Consistent Akaike Information Criterion) criteria. BIC and CAIC were chosen as the basis for model selection because they both take the number of cases into account (in addition to the number of parameters), favoring the most parsimonious model. Because of sparse data (i.e., sparse patterns of all possible item endorsement), the model fit was determined based on Pearson Chi-square and Cressie-Read statistics from 400 bootstrap samples. This number of samples is sufficient to allow for a reliable estimation of model fit

(Gollwitzer et al., 2005). Mixed Rasch models were run using Winmira 2001 (von Davier & Rost, 1995). These models were used to investigate the following hypotheses:

Hypothesis 1: Respondents answering the depressive and the hyperthymic subscales exhibit different item difficulty estimates on desirable/undesirable items within each subscale.

Hypothesis 2: Respondents answering the anxious temperament scale that generally describe neutral behaviors or traits do not exhibit differences in item difficulty estimates on the items constituting the scale.

Predictors of class membership. MRM estimate for each respondent the probability of belonging to each of the estimated classes. Respondents were assigned to the class with the highest probability. Class membership was then modeled using design-adjusted logistic regression. The main predictors included in the model were 1) the respondent's score on the CIDI social conformity scale (low, medium, high); 2) the mode of interview (interviewer administered vs. self-administered); and 3) the presence of parents, partners, or "others" during all of the interview time, some of the time, or none of the time. Interactions between mode of interview and third-party presence were tested. All models controlled for the following sociodemographics: the respondent's age, gender, marital status, educational level, employment status, and household income level. The weights used in all the design-adjusted logistic regression models accounted for differential selection probabilities, poststratification to age and gender population distributions, as well as subsampling into the different questionnaire sections. These analyses were conducted using SAS 9.2 (SAS institute, NC) and allowed for testing the following hypotheses:

Hypothesis 3: Respondents who answer the temperament scale using an interviewer-administered mode are *more* likely to belong to the MRM class exhibiting a socially desirable response behavior compared to respondents who fill the scale in by themselves, controlling for all other variables in the model.

Hypothesis 4: Respondents interviewed in the presence of a third party are *more* likely to belong to the MRM class exhibiting a socially desirable response behavior compared to other respondents interviewed without the presence of the third member, controlling for all other variables in the model.²⁶

Hypothesis 5: Third-party presence moderates the effect of the interview mode on class membership. The effect of the interview mode on belonging to the MRM class exhibiting a socially desirable response behavior is more pronounced when a third member is present during the interview than when the interview is conducted in private, controlling for all other variables in the model.

Results

Respondent Classes Exhibiting Socially Desirable Response Behavior (Table IV.1 and Figures IV.2a–IV.2b, IV.3, and IV.4)

After comparing the BIC and CAIC fit indices for all four IRT models for the hyperthymic and the depressive temperament subscales, we decided that the two-class mixed Rasch model had the best fit. BIC and CAIC were lowest for the two-class model compared to the one-class, three-class, and four-class Rasch models (see Table IV.1). Both the Pearson Chi-square and the Cressie-Read

²⁶ No specific hypotheses were made for the different types of relationship between the third party and the respondent (parent, partner, and others). However, all three relationships were investigated to explore any possible differences in their effects as described in the literature.

statistics for the two-class model also displayed good robustness. Based on these results, the two-class MRM was chosen as the best solution for the hyperthymic and depressive temperaments.

For the anxious temperament, though the two-class model had the lowest information statistics, neither the Pearson Chi-square nor the Cressie-Read statistics suggested good robustness. These two statistics behaved very differently for the anxious temperament than for the hyperthymic and depressive temperament subscales. Their p-value was very small, especially when compared to the best fitting model for the hyperthymic and depressive temperaments (see Table IV.1). Therefore, none of the IRT models for the anxious temperament seemed to fit the data well.

The rest of the results focus on the two classes identified for hyperthymic and depressive temperaments. For each respondent, a probability of belonging to each of the two classes was estimated. Respondents were assigned to the class with the highest probability. Figures IV.2a and IV.2b display the distribution of the estimated probability of belonging to class one for each of the two temperaments (hyperthymic and depressive). The majority of respondents had either a high or a low estimated probability for class one. Only a small fraction of respondents (6.2%-6.5%) had mid-range probabilities between 0.40 and 0.60 (see Figures IV.2a–IV.2b).

Table IV.1: Fit Indices and Test Statistics for Mixed Rasch Models with Different Number of Classes

Classes	Hyperthymic Subscale				Depressive Subscale				Anxious Subscale			
	Information fit Indices		Model fit Indices		Information fit Indices		Model fit Indices		Information fit Indices		Model fit Indices	
	BIC	CAIC	P(χ^2)	P(CR)	BIC	CAIC	P(χ^2)	P(CR)	BIC	CAIC	P(χ^2)	P(CR)
One	12883.63	12905.63	0.000	0.000	10806.79	10828.79	0.000	0.000	12500.95	12527.95	0.018	0.005
Two	12808.24	12853.24	0.172	0.125	10612.66	10657.66	0.273	0.255	12362.02	12417.02	0.038	0.02
Three	12838.46	12906.46	0.183	0.030	10657.81	10725.81	0.350	0.073	12410.58	12493.58	0.075	0.033
Four	12883.12	12974.12	0.045	0.010	10743.49	10834.49	0.270	0.038	12504.33	12615.33	0.095	0.03

Note. BIC = Bayesian Information Criterion; CAIC = Consistent Akaike Information Criterion; p=p-value, χ^2 =Chi-square; CR=Cressie-Read. Bold indicate the Model showing the best fit and best robustness.

Figure IV.2a: Distribution of Estimated Probability for Being a Member of Hyperthymic Class I

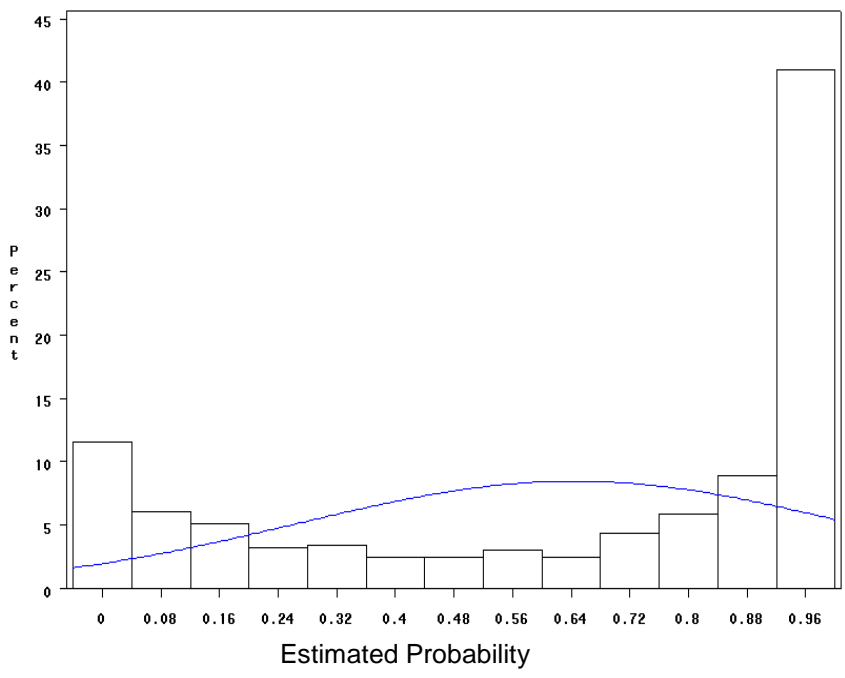
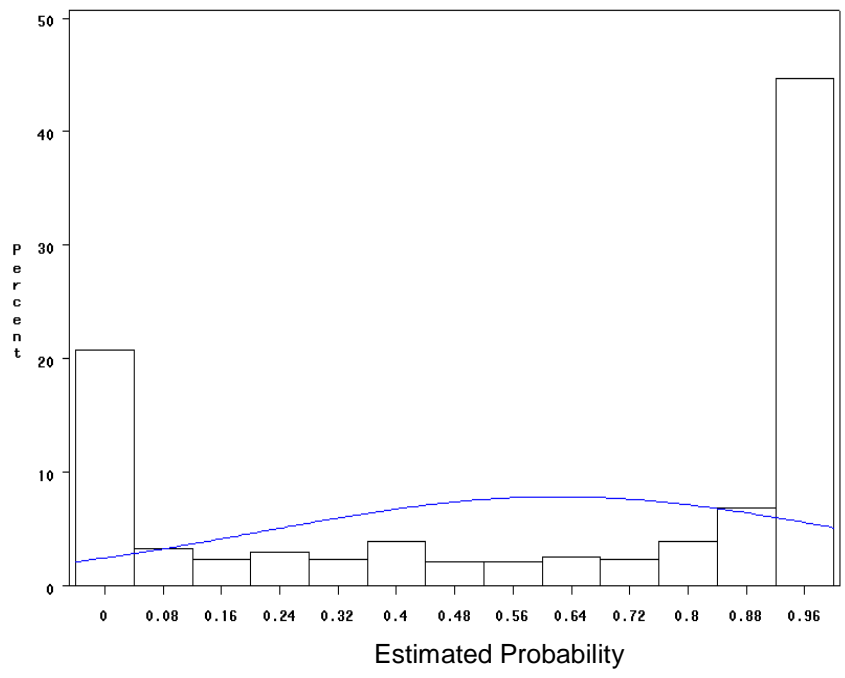


Figure IV.2b: Distribution of Estimated Probability for Being a Member of Depressive Class I



The two identified classes of respondents had different response patterns on the 21 items constituting each of the temperament subscales (see Figures IV.3 and IV.4). These response patterns existed after controlling for the respondent's overall score on the temperament construct (as shown in Equation IV.2 earlier).

For each of the subscales, class one exhibited a more socially desirable response behavior pattern compared to class two. For the hyperthymic temperament scale, respondents belonging to class one endorsed certain items with higher probability (i.e., had lower estimated item difficulties) compared to class-two respondents. Examining the content of these items, it seems that they describe *desirable* traits or behaviors. Examples of such items included "I have great confidence in myself," "I'm usually in an upbeat or cheerful mood," and "I'm the kind of person who believes everything will eventually turn out all right." These items had lower estimated difficulty (i.e., were easier to endorse) among class-one respondents ($\beta_{ih} = -3.62, -1.90, -2.16$) compared to class two ($\beta_{ih} = -1.18, 0.10, -0.24$). The same respondents who belonged to class one, however, endorsed certain items with lower probability (i.e., had higher estimated item difficulties) compared to respondents in class two (controlling for the level of the hyperthymic construct). These items seem to describe *undesirable* traits or behaviors such as "People tell me I often get into things I should not get into," "When I disagree with someone, I can get into a heated argument," and "I am the kind of person who likes to be the boss". These items had higher estimated item difficulties among class-one respondents ($\beta_{ih} = 3.23, 1.69, 0.97$) compared to class-two respondents ($\beta_{ih} = 1.23, -0.07, -0.31$) (see Figure IV.3).

A similar pattern was observed for the depressive temperament subscale. Controlling for the respondent's overall construct score, class-one respondents exhibited a socially desirable pattern of responses on the depressive items. Their estimated item difficulties were higher for items describing undesirable traits and behaviors (such as "For as long as I can remember, I've felt like a failure" and "I give up easily") compared to respondents who belonged to class two.

Figure IV.3: Two-Class Mixed Rasch Model for Hyperthymic Subscale

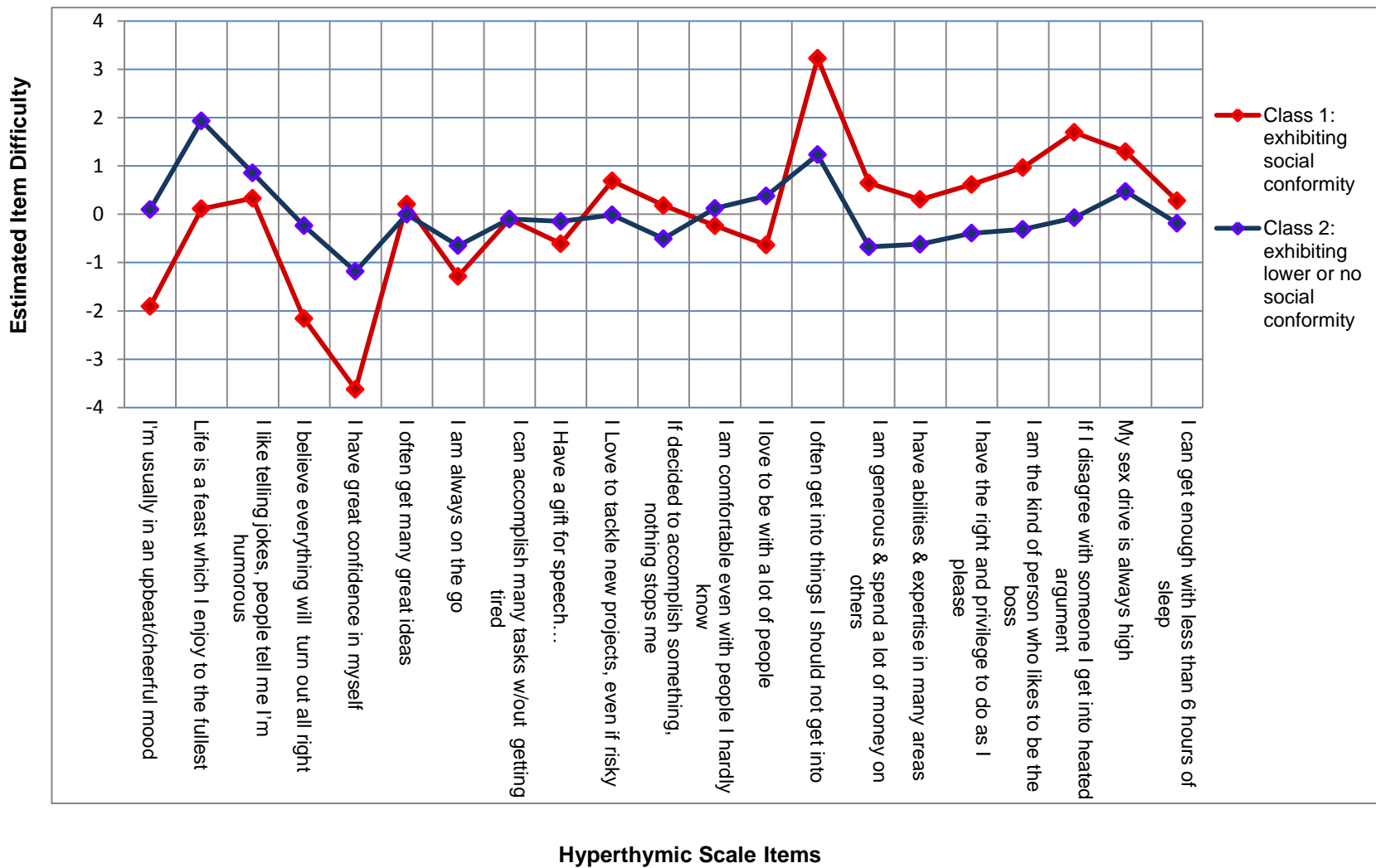
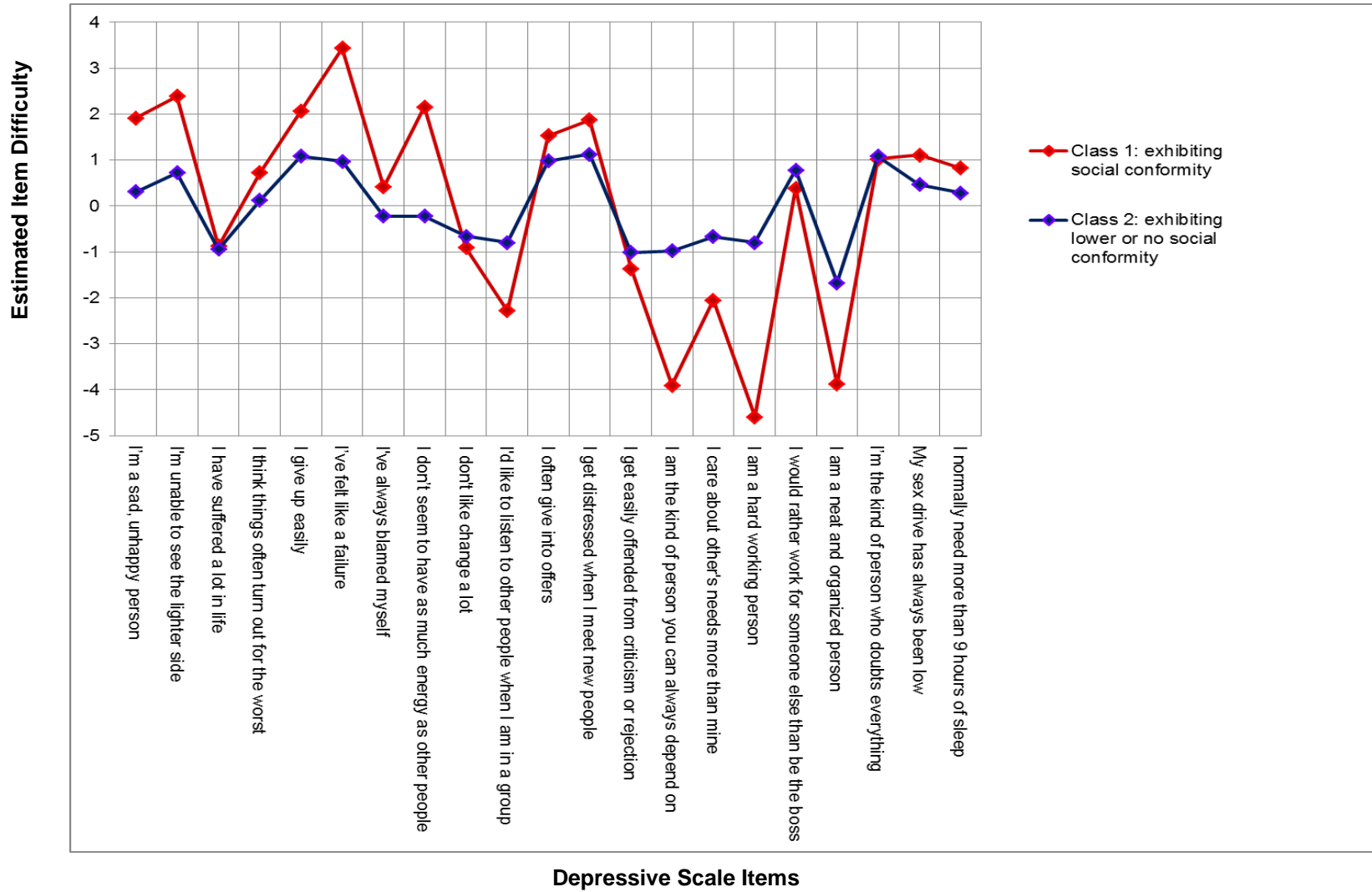


Figure IV.4: Two-Class Mixed Rasch Model for Depressive Subscale



Correspondingly, class-one respondents' estimated item difficulty for items describing desirable traits or behaviors (such as "I am the kind of person you can always depend on" and "I care about others' needs more than mine") were lower (i.e., had higher probabilities of endorsing the items) compared to respondents belonging to class two (see Figure IV.4).

Predictors of Class Membership (Table IV.2)

To validate the nature of these classes, class membership was modeled using design-adjusted binary logistic regression. The association between class membership and objective design and implementation factors, namely interview mode and privacy that could activate and strengthen the respondent's need for social conformity leading to socially desirable response behavior, was tested.

Hyperthymic class membership. Having a partner present all the interview time (vs. no partner present) increased the odds of belonging to class one (which exhibits a more socially desirable response behavior pattern). Respondents interviewed in the presence of their partner had 5.9 times the odds to belong to class one compared to those interviewed without a partner. Moreover, having the interviewer administer the temperament scale (vs. having the scale self-administered) increased the odds of belonging to class one by 1.6 times (Table IV.2, Hyperthymic Class One Main Model). Though this association was in line with the expected direction, it was not found to be statistically significant (95% Confidence Interval=0.8-2.9). Class membership was also found to be statistically associated with the respondent's score on the CIDI social conformity scale. Respondents who scored high on the scale (one standard deviation above the sample mean) had 2.7 times the odds of belonging to class one compared to those with a low score (one standard deviation below the mean). Though this relationship was not part of the main hypotheses, we describe it here so that we can contrast it with the findings related to the depressive temperament class membership in the discussion section.

Table IV.2: Odds Ratio (OR) and 95% Confidence Interval (CI) from Design-Adjusted Logistic Regression Model Predicting Response Class Membership

Predictors	Hyperthymic Class	Depressive Class	Depressive Class
	One	One	One
	Main Model	Main Model	Interaction Model
	OR(95%CI)	OR(95%CI)	OR(95%CI)
Score on CIDI Social Conformity Scale			
High	2.7 (1.2–6.3)	1.0 (0.3–3.4)	1.0 (0.3–3.6)
Middle	1.1 (0.6–2.1)	1.2 (0.4–3.5)	1.2 (0.4–3.7)
Low	1.0	1.0	1.0
Mode of Administration			
Interviewer	1.6 (0.8–2.9)	1.3 (0.6–2.9)	1.0(0.4–2.6)
Self-filled	1.0	1.0	1.0
Partner Presence During Interview			
All the time	5.9 (2.0–17.7)	1.6 (0.4–6.0)	1.2 (0.3–6.1)
Some of the time	1.1 (0.5–2.4)	0.8 (0.3–1.8)	0.4 (0.2–1.00)
None of the time	1.0	1.0	1.0
Parent Presence During Interview			
All of the time	2.1 (0.5–9.2)	2.0 (0.5–8.4)	1.9 (0.4–8.7)
Some of the time	1.2 (0.4–3.4)	2.2 (0.6–7.9)	2.2 (0.6–7.9)
None of the time	1.0	1.0	1.0
Others Present During Interview			
All of the time	0.7 (0.2–2.2)	0.5 (0.2–1.1)	0.6 (0.3–1.2)
Some of the time	0.7 (0.3–1.6)	0.7 (0.4–1.5)	0.7 (0.4–1.4)
None of the time	1.0	1.0	1.0
Interaction Effects			
Interviewer Administered*	---	---	1.4 (0.2–8.7)
Partner Present all of the interview time			
Interviewer Administered* Partner Present some of the interview time	---	---	3.7 (0.9–14.5)

Note. Bold=p-value <0.05; Bold-italics=p-value <=0.06; Dashes “---” indicate variables not entered in the model

Depressive class membership. As shown in the main model, though the direction of the interview mode and partner-presence effects on predicting membership to class one (that exhibited a more socially desirable response behavior) was in the expected direction, both effects were not found to be statistically significant (Table IV.2, Depressive Class One Main Model). The

interaction model, however, showed a marginal interaction effect (p -value=0.06) between these two predictors. When the interview was conducted without the presence of a partner, the odds of belonging to class one were the same for respondents interviewed by an interviewer and those who filled the depressive scale on their own (OR=1.0). The odds ratio, however, increased to 1.5 and 1.7 when a partner was present during some of the interview time (vs. none of the time) and all of the time (vs. none of the time), respectively.

Discussion

Studying socially desirable reporting behavior requires establishing *topic-specific respondent-level* measures that are sensitive to the context of the interview. Unfortunately, commonly used respondent-level measures are not topic specific. These measures are assessed independently of the substantive topic being investigated and suffer from a number of limitations including measurement inequivalence when used across different cultural groups. One possible way to address the limitations of such existing measurement tools and establish measures that are topic specific is to use an extension of item response theory, namely mixed Rasch models (MRM). These models integrate both respondent-level and question-level parameters and are commonly used in the educational field on scales measuring knowledge items. MRM were originally proposed to study differences in response patterns among examinees for reasons such as differential use of solution strategies (e.g. elimination strategy or cross-multiplication algorithm strategy) and instructional sensitivity and to isolate those differences from the examinee's construct scores. This is the first work that extends these methods to study socially desirable response behavior on a community sample and to investigate the nature of response pattern differences using independent and objective design and implementation factors, namely interview mode and interview privacy.

MRM were applied to three subscales: hyperthymic, depressive, and anxious temperament. These scales were chosen because of the differences in their content. Although the hyperthymic temperament generally describes

desirable traits and behaviors, the depressive temperament generally describes undesirable ones, and the anxious temperament describes relatively neutral behaviors or traits. When MRM models were applied to these scales, as hypothesized, we were able to differentiate between two classes of respondents exhibiting different response pattern behaviors on the hyperthymic and depressive items. These models, however, did not fit the data well for the more neutral anxious temperament, which is in line with our hypothesis. Upon exploring the nature of the respondent classes on the hyperthymic and depressive temperaments, one of the classes seems to exhibit a more socially desirable reporting behavior pattern than the other. Respondents belonging to this class generally endorse desirable items at a higher probability while they endorse undesirable items at a lower probability compared to the other class of respondents. To further investigate the nature of these classes and strengthen the conclusion that these classes represent respondent characteristics linked to socially desirable reporting behavior rather than some structural differences in the substantive scales themselves, we tested the association between class membership and two survey variables—interview mode and interview privacy. These two independent, objective design and implementation variables have been studied in the literature and have been proven to affect the reporting of sensitive measures. The more private the mode and the interview setting, the less misreporting observed (Tourangeau & Yan, 2007). For the hyperthymic temperament and as hypothesized, respondents who are interviewed in the presence of their partner are more likely to belong to the class exhibiting socially desirable reporting behavior compared to respondents who are interviewed in private. As for the interview mode, respondents who are administered the hyperthymic scale by the interviewer are also more likely to belong to this class; however, the relationship is not significant. For the depressive temperament, interview mode and partner presence show a marginal interaction effect on class membership. The effect of the interview mode on class membership is intensified when the scale is administered in the presence of a partner. As stated in our hypothesis earlier, respondents who are administered the depressive scale by an

interviewer (compared to those who self-administer the scale) are more likely to belong to the class exhibiting socially desirable reporting behavior when a partner is present. The lack of statistical power (due to the relatively small sample size) might have led to the marginal statistical significance of this association.

Differences in the predictors between the hyperthymic and depressive class membership are not surprising in light of the different nature of the topics covered in each of the temperament subscales. Respondents who perceive the nature of the depressive statements to be generally undesirable and thus might possibly misreport information may be different than those who are sensitive to the nature of the hyperthymic temperament (that generally includes desirable statements). This emphasizes the advantage of using measurement tools that are not only respondent specific but are also topic specific. Such measurement tools avoid locking respondents into groups where the same respondent is expected to have stable response behavior across the range of sensitive topics—an assumption implicit in commonly used social conformity scales.

Furthermore, though respondents who scored high on the CIDI social conformity scale are significantly more likely to belong to the class exhibiting socially desirable reporting behavior on the hyperthymic temperament, this association was not observed for the class membership on the depressive temperament. In fact, upon further exploring the nature of the statements constituting the CIDI social conformity, we observed that all of the nine statements describe desirable traits (generally similar in nature to the hyperthymic temperament). Thus, respondents who scored high on the social conformity scale might be those respondents who are generally more sensitive to similar desirable items. Moreover, the desirable nature of the traits and behaviors described in both the hyperthymic temperament and the CIDI social conformity scale might have unconsciously triggered exaggerated self-views of certain respondents leading them to unconsciously respond in a socially desirable manner on both measures. In fact, as mentioned earlier, the CIDI social conformity scale is adapted from the Marlowe Crown Scale. Marlowe Crowne

Scale items have been shown to load on the dimension of self-deception (Paulhus, 1984).²⁷ This could have led to the significant association between respondents' scores on the CIDI social conformity scale and the hyperthymic class membership. While we cannot warrant which method -MRM method or scores on the CIDI social conformity scale- might be more accurately capturing the unconscious and/or conscious social desirability reporting on the hyperthymic temperament, , we recognize that social conformity is contextual and is activated and strengthened by external stimuli—such as sensitive questions. Thus, a measure that is specific to the substantive topic is expected to perform better than a topic-independent measure.

Lastly, the findings presented in this chapter should be considered in light of the following limitations. First, for each respondent, MRM models estimate a probability of belonging to each of the classes. While the majority of respondents had high estimated probabilities of belonging to one class or the other, a few respondents (less than 6.6%) had probabilities in the range of 0.40–0.60, creating possible uncertainty in class assignments. To address this, one could avoid such assignment by modeling the probability of belonging to a class. Such an approach, however, would make it more difficult to extend the use of mixed Rasch models to real-time measurement adaptation where respondent-specific interventions are implemented based on the respondent's estimated class as discussed later in the conclusion chapter. Second, though the total number of respondents was sufficient to estimate MRM parameters, and the number of bootstrap samples was sufficient to allow for reliable estimation of model fit (Gollwitzer et al., 2005), the test of main and interaction effects between partner presence and class membership might have suffered from low statistical power.²⁸ Third, the assignment of interview mode to the respondents was not randomized. Though many of the factors associated with the administered interview mode

²⁷ Self-deception is defined as the unconscious tendency to give inflated descriptions of one-self (Paulhus, 1984)

²⁸ Only 9.8% and 18.8% of respondents had a partner present during all or some of the interview time, respectively.

were controlled for (such as age, education, marital status, and income), it is possible that there could be other factors that are associated with socially desirable reporting behavior and the mode of administration that were not controlled for. To avoid any possible confounding, a randomized mode experiment is needed. In fact, as part of the initial planning of this dissertation, we designed a fully randomized mode experiment where respondents were assigned to one of two modes: Computerized Assisted Personal Interview (CAPI) or Audio-Computerized Assisted Self Interview (A-CASI). The data collection is expected to begin in January 2013 as part of the Saudi National Health and Stress Survey. The main objective of designing such an experiment is to explore the nature of the MRM latent classes and strengthen the conclusion that they capture socially desirable reporting behavior rather than some structural difference in the substantive construct. Other experiments that could fulfill this objective include randomizing respondents into different question versions. For the same questions that measure the same level of the construct, two versions could be designed. One that loads the question phrasing to make it more sensitive and another that uses more neutral phrasing. Differences in the results of MRM models run on each of the versions could thus be attributed to the increased sensitivity of the loaded items rather than any structural differences in the construct. A fourth consideration to keep in mind when using MRM is that IRT models in general can only be applied to scales with multiple items. Thus, measuring social desirability using single questions and adjusting for it is not possible using IRT models. An alternative approach would require the addition of scales measuring related constructs (keeping in mind that such addition will increase the interview time). For example, researchers interested in studying drug use could add a scale measuring attitudes toward drug use before asking about drug behavior. Respondents with socially desirable reporting behavior on such attitude scales would be expected to be more sensitive toward the drug use behavior questions than those exhibiting a more honest reporting behavior on the attitude scale. The attitude scale, however, needs to be separated from the behavior questions to reduce any possible context effects. Fifth, and finally, MRM

might not be suited to distinguish between the two possible components of social desirability reporting, self-deception and impression management (consciously reporting inflated self-descriptions). While such distinction might not be as important for identifying respondents exhibiting socially desirable reporting behavior, it could impact the effectiveness of intervention messages designed to reduce socially desirable reporting.

Conclusion

Survey research needs to move away from general social conformity scales towards a topic-specific measurement method. Using design and implementation variables that moderate the interaction between the respondent and survey questions, this chapter demonstrates the ability of using psychometric measurement tools to classify respondents into classes exhibiting different socially desirable reporting behavior.

By applying mixed Rasch models to scales measuring desirable or undesirable behaviors and traits on a community sample, it was possible to identify respondents displaying socially desirable reporting behavior based on their response patterns. Such methods will not only provide respondent-level topic-specific measures of socially desirable reporting behavior, but will also save on using an independent measure of social conformity that usually adds to interview time and respondent burden. Moreover, model-generated estimates of item desirability (or undesirability) might become available by calculating (for each item) the difference in the estimated class-specific item difficulty. The nature of classes identified by the mixture models, however, still needs further investigation using randomized experiments of design variables that have been shown to reduce misreporting. Once further validated, the ultimate aim would be to design computer adaptive measurement tools (with mixed Rasch models as a backbone) that estimate in real-time respondents' response patterns resembling socially desirable response behavior as well as their corresponding level of construct. Based on these estimates, the adaptive measurement tool would route each respondent to the next scale item that best fits his/her estimated level of

construct, or would administer intervention messages that aim to reduce his/her social conformity concerns, reducing measurement variance (with the former) and measurement bias (with the latter) while also minimizing the number of scale items that need to be asked compared to a static instrument. Such practical implications are detailed further in the conclusion chapter.

Appendix A

Hyperthymic Temperament

As seen from the symptoms listed below, people with high hyperthymic scores have a temperament that is generally characterized by being optimistic, full of energy, entrepreneurial, and risk taking.

Item Number	Hyperthymic Subscale Items
1	I'm usually in an upbeat or cheerful mood.
2	Life is a feast which I enjoy to the fullest.
3	I like telling jokes, people tell me I'm humorous.
4	I'm the kind of person who believes everything will eventually turn out all right.
5	I have great confidence in myself.
6	I often get many great ideas.
7	I am always on the go.
8	I can accomplish many tasks without even getting tired.
9	I have a gift for speech, convincing, and inspiring others.
10	I love to tackle new projects, even if risky.
11	Once I decide to accomplish something, nothing can stop me.
12	I am totally comfortable even with people I hardly know.
13	I love to be with a lot of people.
14	People tell me I often get into things I should not get into.
15	I am generous, and spend a lot of money on other people.
16	I have abilities and expertise in many areas.
17	I feel I have the right and privilege to do as I please.
18	I am the kind of person who likes to be the boss.
19	When I disagree with someone, I can get into a heated argument.
20	My sex drive is always high.
21	I can get enough with less than 6 hours of sleep at night.

Depressive Temperament

As seen from the symptoms listed below, people with high depressive scores have a temperament that is generally characterized by having a negative attitude toward life, a lack of motivation, and less desire to lead others.

Item Number	Depressive Subscale Items
1	I'm a sad, unhappy person.
2	People tell me I am unable to see the lighter side of things.
3	I have suffered a lot in life.
4	I think things often turn out for the worst.
5	I give up easily.
6	For as long as I can remember, I've felt like a failure.
7	I have always blamed myself for what others might consider no big deal.
8	I don't seem to have as much energy as other people.
9	I am the kind of person who does not like change a lot.
10	I prefer to listen to other people when I am in a group.
11	I often give into offers.
12	I feel very uneasy when I meet new people.
13	I am easily offended by criticism or rejection.
14	I am the kind of person you can always depend on.
15	I care about others' needs more than mine.
16	I am a hard-working person.
17	I would rather work for someone else than be the boss.
18	I am a neat and organized person by nature.
19	I'm the kind of person who doubts everything.
20	My sex drive has always been low.
21	I normally need more than 9 hours of sleep.

Anxious Temperament

As seen from the symptoms listed below, people with high anxious scores have a temperament that is generally characterized by reacting to events and stresses in an anxious way.

Item Number	Anxious Subscale Items
1	I have been a worrier for as long as I can remember.
2	I'm always worrying about one thing or another.
3	I keep on worrying about daily matters that others consider minor.
4	I cannot help worrying.
5	Many people have told me not to worry so much.
6	When stressed, I can't think.
7	I am unable to relax.
8	I often feel jittery inside.
9	When stressed, my hands often tremble.
10	I often have an upset stomach.
11	When I'm nervous, I may have diarrhea.
12	When I'm nervous, I often feel nauseous.
13	When I'm nervous, I have to go to the bathroom more often.
14	When someone is late coming home, I fear they may have had an accident.
15	I am often fearful of someone in my family coming down with a serious disease.
16	I'm always thinking someone might break bad news to me about a family member.
17	My sleep is not restful.
18	I frequently have difficulty falling asleep.
19	I am, by nature, a very cautious person.
20	I often wake up at night afraid that there might be a thief at home.
21	I easily get headaches when stressed.
22	When stressed, I get an uncomfortable feeling in my chest.
23	I'm an insecure person.
24	Even minor changes in routine stress me highly.
25	When I drive a car, even if I did not break any law, I fear that the police might stop me.
26	Sudden noises startle me easily.

Appendix B
CIDI Social Conformity Scale

Item Number	Scale Items
1	I never met a person that I didn't like.
2	I have always told the truth.
3	I always win at games.
4	I have never been bored.
5	I never get lost, even in unfamiliar places.
6	I never get annoyed when people cut ahead of me in line.
7	My table manners at home are as good as when I eat out in a restaurant.
8	I have never lost anything.
9	No matter how hot or cold it gets, I am always quite comfortable.

References

- Akiskal, H. S., Akiskal, K. K., Haykal, R. F., Manning, J. S., & Connor, P. D. (2005). TEMPS-A: Progress towards validation of a self-rated clinical version of the temperament evaluation of the Memphis, Pisa, Paris and San Diego auto questionnaire. *Journal of Affective Disorder, 85*, 3–16.
- Beretvas, S. N., Meyers, J. L., & Leite W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement, 62*, 570–589.
- Caldini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621.
- Cannell, C. F, Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, 389–437.
- Crowne, D., & Marlowe, D. (1964). *The approval motive*. New York, NY: Wiley.
- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. E. Turner & E. Martin (Eds.), *Surveying subjective phenomena*. New York, NY: Russell Sage.
- Edwards, A. L. (1957). *The social desirability variable in personality research and assessment*. New York, NY: Dryden Press.
- Eid, M., & Zickar, M. J. (2007). Detecting response styles and faking in personality and organizational assessments by mixed Rasch models. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models*. New York, NY: Springer Science + Business Media.
- Eysenck, H. J., & Eysenck S. B. G. (1964). *The Eysenck personality inventory*. London, U.K.: University of London Press.
- Fisher, R.J., & Dube, L. (2005). Gender differences in responses to emotional advertising: A social desirability perspective. *Journal of Consumer Research, 31*, 850-858.
- Gollwitzer, M., Eid, M., & Jürgensen, R. (2005). Response styles in the assessment of anger expression. *Psychological Assessment, 17*, 56–69.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (Eds.). (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

- Harkness, J., Pennell, B., Villar, A., Gebler, N., Auilar-Gaxiola, S., & Bilgen, I. (2008). Translation procedures and translation assessment in the World Mental Health Survey Initiative. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health Surveys: Global perspectives on the epidemiology of mental disorders*. New York, NY: Cambridge University Press.
- Hebert, J.R., Ma, Y., Clemow, L., Ockene, I.S., Saperia, G., Stanek, E.J., Merriam, P.A., & Ockene, J.K. (1997). Gender differences in social desirability and social approval bias in dietary self-report. *American Journal of Epidemiology*, *146*, 1046-1055.
- Holden, R. R., & Book, A. S. (2009). Using hybrid Rasch-latent class modeling to improve the detection of fakers on a personality inventory. *Personality and Individual Differences*, *47*, 185–190.
- Johnson, T. P., & Van De Vijver, F. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural Survey Methods*. Hoboken, NJ: Wiley.
- Karam, E. G., Mneimneh, Z. N., Karam, A. N., Fayyad, J. A., Nasser, S. C., Chatterji, S., & Kessler, R. C. (2006). Prevalence and treatment of mental disorders in Lebanon: A national epidemiologic survey. *Lancet*, *367*, 1000–1006.
- Karam, E. G, Mneimneh, Z., Salamoun, M., Akiskal, K. K., & Akiskal, H. S. (2005). Psychometric properties of the Lebanese-Arabic TEMPS-A: A national epidemiologic study. *Journal for Affective Disorders*, *87*, 169–183.
- Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M.J., Jin, R., Vega, W.A., Walters, E.E., Wang, P., Zaslavsky, A., & Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 122–139.
- Kessler, R.C., & Üstün, T. B. (2004). The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 93–121.
- Klassen, D., Hornstra, R.K., & Anderson, P.B. (1975). Influence of Social Desirability on Symptom and Mood Reporting in a Community Survey. *Journal of Consulting and Clinical Psychology*, *43*, 448-452

- Kuncel, N. R., & Borneman, M. J. (2007). Towards a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, 15, 220–231.
- Majj-de Meij, A. M., Kelderman, H., & van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement*, 32, 611–631.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271–282.
- Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, 87, 211–219.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, U.K.: Cambridge University Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883.
- Von Davier, M. V., & Rost, J. (1995). *WINMIRA: Windows mixed Rasch model analysis [Computer software]*. Kiel, the Netherlands: Institute for Science Education.
- Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology*, 23, 419–427.
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, 7, 168–190.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, 84, 551–563.
- Zickar, M. J., & Ury, K. L. (2002). Developing an interpretation of item parameters for personality items: Content correlates of parameter estimates. *Educational and Psychological Measurement*, 62, 19–31.

Chapter V

Conclusion

We conclude with a review of our findings, followed by a discussion of research and design implications, including refining our measures of interview privacy, increasing sensitivity to the differing expectations of privacy across cultures, improving interviewer training, and redesigning adaptive measurement tools to tailor question administration to differing degrees of social conformity among respondents.

Brief Summary of Findings

The earlier chapters demonstrated the importance of investigating the effects of respondent-level, question-level, study design, and interview implementation factors on socially desirable response behavior. The main respondent-level and interview implementation factors studied included the respondent's cultural background, the respondent's need for social conformity, and interview privacy.

Chapters II and III focused on predictors of interview privacy and its effect on socially desirable response behavior. Findings from Chapter II revealed a significant variation across cultures and interviewers in establishing a private interview setting. The country's wealth, the country's level of individualism, and its masculinity significantly affected the privacy setting of the interview either directly or through interacting with several demographic and socioeconomic characteristics of the respondent and the third party present during the interview. Most important, Chapter III demonstrated that having a third party present during the interview interacted with the respondent's need for social conformity and his

or her cultural background to either decrease or increase the odds of reporting undesirable behaviors, such as suicide plan or attempt, or increase the odds of reporting desirable attitudes such as high marital rating. Interaction with the respondent's need for social conformity significantly affected the reporting of suicidal behavior, and interaction with cultural factors significantly affected the reporting of a high marital rating. Ignoring the effects of such interactions could lead to misleading conclusions about third-party presence influences on reporting sensitive information.

Chapter IV used an extension of item response theory, namely mixed Rasch models, to generate a topic-specific measure of social conformity that integrated both respondent-level and question-level parameters when studying socially desirable response behavior. When applied to scale items measuring desirable or undesirable traits or behaviors, respondent latent classes capturing characteristics, such as the need for social conformity, that increase respondents' likelihood of exhibiting socially desirable response behavior were found. Class membership was found to be associated with interview privacy or its interaction with interview mode depending on the scale investigated.

Based on the findings reported in the earlier three chapters, we discuss researcher recommendations as well as design implications in relation to interview privacy measures, the respondent's cultural background, third party's held information, interviewer training, and measures of social-conformity that affect socially desirable response behavior.

Recommendations and Implications

Research Recommendations

Interview privacy setting measures. Findings from Chapters II and III call for collecting qualitative and quantitative data about the interviewer's behavior and attitude toward privacy; the dynamics of the interaction between the respondent, other household members, and the interviewer; and other interview

setting characteristics. This will allow survey researchers to better understand interviewer variations and cultural influences and help them to develop tailored methods to shape interviewers' attitudes and behaviors toward achieving interview privacy. Before designing any quantitative measures, some ground work needs to be done and qualitative data using focus groups and open-ended questions need to be collected from interviewers who vary in their reported rates of interview privacy. Information on how interviewers view the importance of privacy in general and interview privacy in particular, how they request it and negotiate with household members, what techniques they use to convince household members of its importance (if any) are example of topics to be covered during focus groups with interviewers. Once qualitative data are collected and analyzed and based on the findings, a small set of quantitative observation questions could be added to the actual interview. If feasible, also recording the interaction between the interviewer and the respondent would be ideal. Recording the interview setting will allow for exploring the measurement property of the observations taken by the interviewer. Collecting more detailed information on the interview setting and the process through which the interpersonal context of the interview is assembled will also allow survey researchers to better understand the effect of third-party presence on the interviewer's behavior and the respondent's cognitive processes when answering survey questions, specifically sensitive ones. Examples of observation questions that capture such dynamics are provided in the appendix at the end of this chapter. However, when designing such measures, it is important to keep in mind the increased cost associated with such additions. Researchers are highly encouraged to carefully design their interviewer observations questions and include ones that are of value to the survey. Researchers are also encouraged to collect and analyze any time data associated with these measures to be informed about the cost associated with administering them.

Respondent's cultural background. Cultural dimensions seem to play a significant role in determining interview privacy and its effect on reporting.

Researchers are encouraged to investigate further the direct and indirect effects of cultural dimensions like the country's level of individualism, masculinity, power distance, and wealth on interview privacy and reporting. This is especially important in cross-cultural research where comparability is key. While we used measures of cultural dimensions calculated and reported by other authors, researchers interested in cross-cultural comparisons are highly encouraged to collect their own measures on their study respondents themselves and investigate their effect on reporting differences across cultures. Examples of measures to collect come from Hofstede's work and include whether the respondent values personal time and freedom (for the individualism construct), respondent's attitudes towards disagreeing with authority figures, and preference for autocratic versus participative managers (for the power distance construct), and whether the respondent value work recognition advancement, and challenge (for the masculinity construct).

Information held by the third party. Chapter III demonstrated that the effect of third party presence on reporting sensitive behaviors, such as suicide is moderated by respondent's need for social conformity. While third party presence decreases the reporting of such behaviors among respondents with high need for social conformity, it could potentially increase reporting of such behaviors among respondents with low social conformity needs. Respondents with low needs for social conformity might not have high concerns about possible negative consequences associated with divulging such sensitive behaviors, or they might have already confided in the third person present during the interview. To investigate this further, researchers are encouraged to design measures that capture information already held by the third party and investigate its moderating effect on third party presence and reporting sensitive behaviors. Such measures could be self-administered to the third person by either asking them directly about the respondent behavior or asking them about the likelihood that the respondent would confide in them if he/she got engaged in certain sensitive behaviors.

Respondent's need for social conformity. Findings from Chapter IV suggest that applying mixed Rasch models to scales measuring sensitive information seem to be a promising tool for capturing latent classes of respondents exhibiting different social-conformity needs. These findings call for a research paradigm that uses improved and more comprehensive designs to investigate further the nature of such latent classes and confirm the findings. Such future research needs to build on the framework presented in Figure I.1 and investigates independent factors related to the study design, interview implementation, or question characteristics that could interact with the respondents' need for social conformity, affecting their responses. One possible research design would be to conduct a randomized mode experiment in which respondents are randomized into different modes such as CAPI (computer assisted personal interviewing) and A-CASI (audio-computer assisted self-interviewing) and administered a variety of scales measuring sensitive and relatively neutral items. Data collected in each mode would then be analyzed using mixed Rasch models, and the results would be compared. If latent classes from the mixed Rasch model capture the respondent's need for social conformity, we would expect any differences in item difficulty estimates across the latent classes to be smaller (or nonexistent) in the A-CASI subsample compared to the CAPI subsample. The rationale is that lower privacy concerns are expected in A-CASI compared to CAPI; this is evident in previous findings that showed higher reporting of sensitive information in A-CASI (Tourangeau & Smith, 1996). In fact, as part of the initial planning of this dissertation, we have already designed this experiment, and we expect the data to be collected beginning January 2013 as part of the Saudi National Health and Stress Survey.

This design could be supplemented by also randomly administering scale versions with "loaded" items to a relatively small subsample of respondents to increase the items' level of desirability or undesirability. Results from mixed Rasch models run on this version could then be compared to those run on the well-phrased scale version. Again, the expectation is that smaller variation in class-specific item difficulty estimates would be found in the well-phrased scale

compared to the “loaded” scale due to its effect on heightening social-conformity concerns among certain respondents.

Design Implications

Interviewer training practices. The magnitude and significance of interviewer variance in interview privacy measures suggest that more work needs to be done on the different steps that are needed to guarantee interview privacy: requesting, negotiating, achieving, maintaining, and measuring it. Interviewers need to be drilled during training sessions on different scenarios they might face when requesting privacy. Such scenarios include concerns raised by the respondent or third party about the respondent being interviewed in private, questions about the need for privacy, and interruptions during the course of the interview. Practice sessions on how to address these concerns, and most importantly tailor interviewers’ behavior to meet the prevailing cultural norms, need to be incorporated in interviewer training sessions and material. For example, having interviewers remind the third person about the study confidentiality requirements or having them induce the respondent to request privacy might be appropriate in certain cultures. In other cultures, however, interviewers might have to repeatedly satisfy the third person’s curiosity by answering their questions, sidetrack the third person, or create an environment that is less encouraging for third parties to be present, such as gender matching interviewers to respondents or occupying the third party with a self-administered questionnaire designed to collect additional information (possibly on the household) that could benefit the research being undertaken.

Designing and implementing such practices and materials may help to increase interview privacy. Increasing interview privacy among respondents belonging to the same sample will standardize the interview setting across respondents, reducing any layer of *measurement variation* created by the setting effect. Achieving and maintaining a private interview will eliminate one of the contextual stimuli that could activate and enhance the respondent’s need for

social conformity, reducing the impact of such interaction on socially desirable reporting behavior.

Designing adaptive measurement tools. The ultimate aim of using mixed IRT models to capture the respondent's activated need for social conformity when answering sensitive questions is the possibility of using these models as the backbone for adaptive measurement tools designed to reduce measurement error. We first provide a brief summary of how current adaptive measurement tools that rely on simple Rasch models work, and then discuss how they could be adapted to use mixed Rasch models instead to account for possible differences in respondents' socially desirable reporting behavior.

The main motivation for developing and administering computerized adaptive measures is to have not only precise but also efficient instruments for measuring latent constructs. Because of the heterogeneity in the level of estimated construct among respondents, researchers usually strive to design scales with a large number of items to reflect the wide range of construct levels with a good level of measurement precision, yet not too long to burden respondents. Fixed instruments that include long lists of items, despite achieving good precision levels, are nonetheless inefficient for any single respondent. This inefficiency is caused by individual differences in the level of construct being measured. This trade-off between precision and efficiency was addressed by designing computerized adaptive measurement instruments that tailor the list of questions to the construct level of each respondent. Such adaptive tools are based on an algorithm that selects the questions to be administered one at a time. Before selecting the next question, the algorithm calculates a real time estimate of the respondent's construct and an information statistic for each of the scale items that have not yet been administered. The item difficulty greatly

affects the amount of information provided by the item.²⁹ The algorithm then selects the next maximally informative item to be administered to the respondent. The selection and administration of items continues until a specified number of items has been administered or a desired estimate of measurement precision has been achieved. Figure V.1 illustrates the adaptive instrument process.

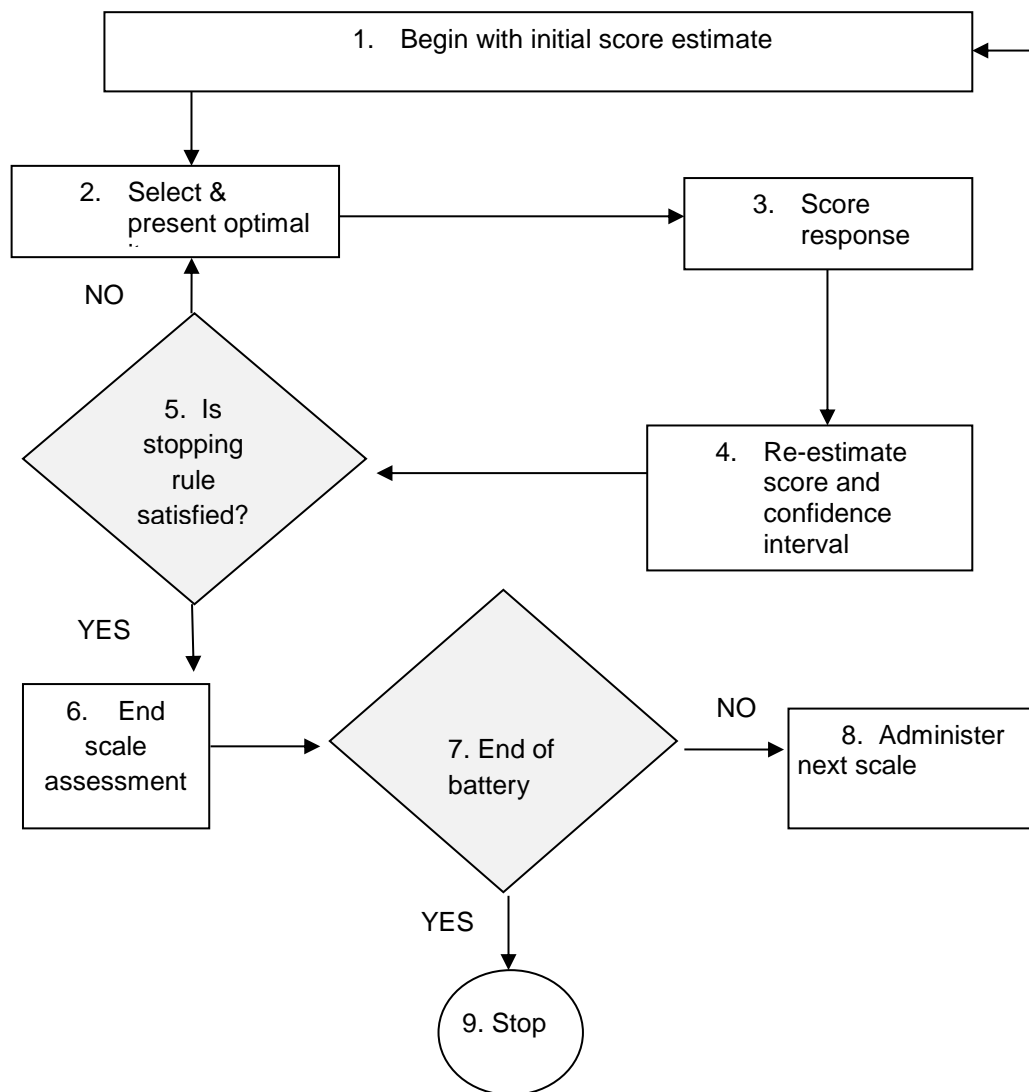


Figure V.1: Computerized Adaptive Testing Logic (Adapted from Wainer et al. [1990])

²⁹ Estimates of item difficulty should be already available before the adaptive instrument is administered. This is usually done in a process called calibration, where a sample of at least 500 individuals are administered the list of items. These items are then calibrated (i.e., their item parameters are estimated) using IRT models.

Such adaptation will shorten the interview length and achieve at least the same level of precision as administering a fixed instrument but with a long set of scale items. Reductions in interview length reported in the literature when comparing adaptive instruments to fixed instruments range from 71% (Ware, Gandek, Sinclair, & Bjorner, 2005) to 95% (Gibbons et al., 2008).

The heterogeneity accounted for in these adaptive instruments refers specifically to respondents' differences in the level of construct being measured. This is an important source of heterogeneity among the population, and this method is appropriate when asking about non-sensitive information. However, such an adaptive measurement approach does not account for possible heterogeneity in socially desirable response behavior when sensitive information is requested. Designing adaptive measurement tools that also address differences in socially desirable reporting behavior are potentially of great value to survey research. One possible design would be to use algorithms that are based on mixed Rasch models instead of the "simple" IRT models that are in place. Mixed Rasch model algorithms would account for differences in socially desirable response behavior by integrating class membership as one component of the tailoring process. Once class membership is assigned, class-specific item difficulty measures would be used to estimate the information statistic for each item that will guide the adaptation process. Such class-specific item difficulties would estimate better the actual level of difficulty of the item (compared to those generated under a simple IRT model) and thus might lead to more accurate estimates for the desired construct. Figure V.2 uses a hypothetical example to illustrate how estimates of item difficulty and the adaptation of the instrument vary under a model that ignores the existence of respondent classes (simple IRT) and a model that accounts for such classes (mixture IRT).

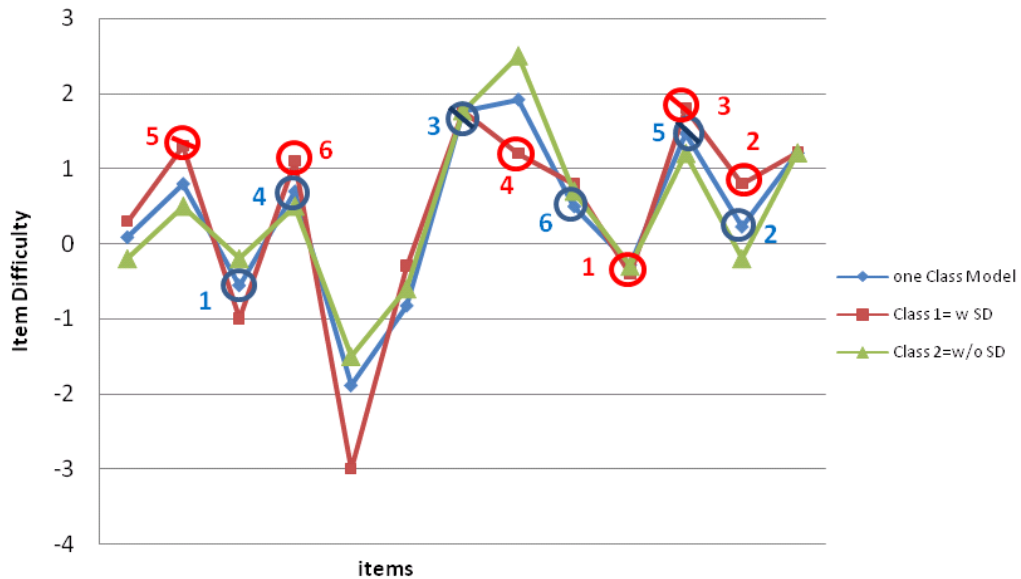


Figure V.2: Hypothetical Example of an Adaptive Measurement Tool Using One-class Model (i.e. Simple IRT) vs. Two-Class Model (Mixed Rasch Model)^a

^a Empty Circles represent a “yes” response on a binary question. Crossed circles represent a “no” response on a binary question. The numbers represent the sequence of administered questions.

W SD: with socially desirable reporting behavior

w/o SD: without socially desirable reporting behavior

In this hypothetical example, the green and red lines represent difficulty estimates for 13 items from a Mixed Rasch model. The red line represents the class of respondents with socially desirable response behavior; while the green line represents the class of respondents with no or minimal socially desirable response behavior. The blue line represents estimates of the item difficulties if a one class model is used (i.e., a simple Rasch model). Estimates of item difficulties are important because they affect the level of information provided by the item and thus they determine the routing of questions in an adaptive instrument and the estimate of the substantive construct. Figure V.2 illustrates how such routing might differ as estimates of item difficulties differ. Suppose a simple Rasch model is used to estimate item difficulties, and the respondent is administered an adaptive instrument based on these estimated difficulties (i.e., the blue line). Based on the responses and on the difficulty estimates from the simple Rasch model, the respondent is administered the following items in the

following order: item 3 (answered yes), item 12 (answered yes), item 7 (answered no), item 4 (answered yes), item 11 (answered no), and finally item 9 (answered yes). If, however, the items are estimated using a mixed Rasch model and if the respondent's estimated latent class exhibits socially desirable behavior (the red line class), the following items in the following sequence would be administered: item 10 (answered yes), item 12 (answered yes), item 11 (answered no), item 8 (answered yes), item 2 (answered no), and item 4 (answered yes). Thus, as illustrated in this example, depending on which model better fits the data, estimates of item difficulties might vary, altering the adaptation of the instrument and possibly leading to different estimates of the construct.

It is important to note that using adaptive measurement tools (whether based on simple IRT or mixed IRT) have raised concerns about possible context effects. Yet, the literature on computerized adaptive tools indicates that changes in item administration order do not compromise the validity of the construct estimates (Roper, Ben-Porath, & Butcher, 1995; Walter et al., 2007). Still, since context effect is scale specific, future studies investigating this issue will be valuable to the application of adaptive measurement in survey research.

So, how would such adaptive data collection tools work in practice?

We propose several measurement adaptation methods. One such method would be to tailor the administration of items within a single construct to the respondent's social-desirability *class* and *level of construct*. This adaptation is illustrated in Figure V.2 by the green and red lines and requires first an estimation of all class-specific item difficulty parameters.

A design similar to Groves and Heeringa's (2006) multiple-phase approach could be used to guide the actual implementation. For illustrative purposes, assume a study is being conducted with a target of 2,000 completed interviews. The study period is divided into three phases per Groves and Heeringa's (2006) suggested responsive design. Below is a description of each phase.

Phase I: A random replicate of the sample is selected and administered a fixed instrument with all the items. The size of the replicate should be large enough to yield about 500 completed interviews (Thissen & Reeve, 2007). This subsample is used to: 1) estimate item difficulty parameters based on a mixed Rasch model and 2) build the predictive model for estimating the probability of belonging to a specific class for the remaining respondents. Respondent-level, design, and implementation factors known to be associated with the need for social conformity (as shown in Figure I.1 and in the findings of Chapter IV) are used as predictors for class membership.

Phase II: For the remaining 1,500 respondents (in this hypothetical example), and based on their specific characteristics, design and implementation factors collected at the beginning of the interview, as well as responses to an initial small set of items, their class membership is predicted.³⁰ Based on this predicted probability, respondents are assigned to specific classes. Class-specific item difficulties (estimated in Phase I above) are then used to route respondents through the sequence of questions that provide the most information for their construct estimate.

It is important to note that for some respondents, the predicted class probability might be borderline (the borderline cutoff could be determined in a simulation study). For these respondents, either more information would have to be collected to revise their predicted class membership before assigning them to a class or no adaptation would be used (i.e., they would be administered all the items in the scale).

Phase III: Respondents interviewed in both phases are combined and their construct level is estimated. For respondents in Phase I, the model used to estimate the item difficulties (in Phase I) is the same one used to estimate respondents' construct level accounting for their class membership.

Other measurement adaptations could include alternating certain dimensions of the instrument, such as commitment or confidential statements, or

³⁰ Such respondent characteristics would include the variables used in the predictive model in Phase I.

using randomized response techniques and tailoring them to respondents exhibiting socially desirable response pattern. In addition to increasing efficiency by administering these dimensions only to the relevant respondents, and possibly reducing measurement bias by dampening social-conformity needs, such methods could be applied to scale-type as well as nonscale-type questions.

In summary, by integrating measures and tools from different disciplines, including social psychology, cross-cultural psychology, psychometrics, and survey methodology, this research advances our understanding of social desirability reporting behavior and opens the door for a new research paradigm that builds on the current findings to come up with new practices that improve the quality of sensitive information collected from respondents.

Appendix

*IWP1. The following list presents types of relationship to the respondent. Select the one(s) that represent the person(s) who were present at any time during this section of the interview.

- 1) Father
- 2) Mother
- 3) Husband
- 4) Wife
- 5) Brother
- 6) Sister
- 7) Son >18 years old
- 8) Daughter >18 years old
- 9) Son 13–17 years old
- 10) Daughter 13–17 years old
- 11) Son 9–12 years old
- 12) Daughter 9–12 years old
- 13) Son 4–9 years old
- 14) Daughter 4–9 years old
- 15) Son under 4 years old
- 16) Daughter under 4 years old
- 17) Father-in-law
- 18) Mother-in-law
- 19) Aunt
- 20) Uncle
- 21) Male Neighbor
- 22) Female Neighbor
- 23) Male Friend
- 24) Female Friend
- 25) Other (Specify)
- 26) No one was present (skip to end)
- 27) Don't Know

*IWP2. Select the appropriate statement:

- 1) The person(s) present could hear the interview questions and answers.
- 2) The person(s) present could hear the interview questions and answers but are too young to understand.
- 3) The person(s) present could *not* hear the interview questions and answers.

*IWP3. Which statement describes the situation that lead to the presence of other person(s) during the interview?

- 1) Person(s) were present in the same room and neither you nor the respondent asked them to leave.
- 2) Person(s) were present in the same room and you asked them to leave but they refused.
- 3) Person(s) were present in the same room and R asked them to leave but they refused.
- 4) Person(s) were present in the same room and you and the respondent asked them to leave but they refused.
- 5) Person(s) were present in the same room but were too young to understand.
- 6) Person(s) were not present in the room and the R invited them to stay and you did not mind.
- 7) Person(s) were not present in the room and the R invited them to stay in spite of the fact that you asked for a private interview.
- 8) Others.

IWP3_Other. Describe the situation that lead to the presence of other person(s) during the interview.

References

- Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., & Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*, 361–368.
- Groves, R. M., & Heeringa, S. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society, 169*, 439–457.
- Roper B. L., Ben-Porath Y. S., & Butcher, J. N. (1995). Comparability and validity of computerized adaptive testing with the MMPI-2. *Journal of Personality Assessment, 65*, 358–371.
- Thissen, D., & Reeve, B. B. (2007). Methodological issues for building item banks and computerized adaptive scales. *Quality of Life Research, 16*, 109–119.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and questions context. *Public Opinion Quarterly, 60*, 275–304.
- Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'anxiety' (Anxiety-CAT). *Quality of Life Research, 16*, 143–155.
- Ware, J. E., Gandek, B., Sinclair, S. J., & Bjorner, J. B. (2005). Item response theory and computerized adaptive testing: Implications for outcomes measurement in rehabilitation. *Rehabilitation Psychology, 50*, 71–78.