

Pragmatic and Dialectic Mixed Method Approaches:  
An Empirical Comparison

A THESIS  
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF MINNESOTA  
BY

Anne E. Betzner

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Frances Lawrenz, Adviser  
December, 2008



## ACKNOWLEDGEMENTS

My graduate experience has been immeasurably enriched by the support and encouragement of my adviser, Dr. Frances Lawrenz. Her critical review of my work was an invaluable learning tool that I will carry forward in my career. Thank you for your insight and responsiveness. I would also like to thank committee members Dr. Michael Harwell, Dr. Robert delMas, and Dr. Debra Ingram for their valuable insights and the dedication of their time. A special thank you to Dr. Stuart Yeh for encouraging me to pursue an evaluation of smoke-free ordinances.

This study embodies the love and support that I have received to conduct and complete it. First, credit is due to my family, especially my mother, Carole Betzner, whose unwavering value for education ignited in me the desire and courage necessary to pursue a doctoral degree. The completion of this study is her accomplishment as well as mine. The love of my husband can also be found on every page in this volume. He will admit that his romantic attention distracted me from my work, what with all the wooing and the marriage. However, no person has been a greater advocate for me to complete my degree. His respect and selfless concern for my work has been an enduring example of how to be the best spouse ever.

This dissertation would not have been possible without the funding of ClearWay Minnesota<sup>SM</sup>, and the braintrust of the ClearWay team, especially Jessie Saul, Ph.D., Barbara Schillo, Ph.D., Ann Wendling, M.D., Marietta Dreher, and Chris Thokles. Many thanks to ClearWay Minnesota.

No funding amount in the world could replace the support that Professional Data Analysts, Inc. provided to me in the completion of this degree, especially from Dr. Michael Luxenberg. He opened the door to the opportunity of conducting this study and provided me with the support and latitude to complete it. Special thanks to Julie Rainey for her assistance in thinking through many issues regarding the telephone survey and focus groups. She also deserves credit for hearing more about this study than she ever wanted. I'd also like to especially thank all the staff at Professional Data Analysts, Inc. who assisted with the evaluation for ClearWay Minnesota, including Traci Capesius, Matt Christenson, Marcy Huggins, Annette Kavanaugh, and Laura Stopyro.

I could not have completed my degree without the support and friendship of Drs. Jeffrey Haring and Ann Ooms. During our time together as students, they were all you could ever wish for as friends. The example of their scholarship and completed degrees was an inspiration to me long after our paths diverged. An extra special thanks to Dr. Andrew Zieffler who went above and beyond the call of duty as a fellow student and friend to see me through to the completion of my degree.

It is with sadness that I remember my father, David Betzner, who completely believed me when I blithely stated that anyone could complete a Ph.D. if they worked hard enough. I have since found the extent to which that statement is both true and false. But the part of me that inherited my dad's obsessive nature surely helped in the process. I'd also like to thank my grandmother, Mary Prokay, R.N., for always talking to me like an adult. I know she would be proud of my accomplishment.

## ABSTRACT

Mixed methods are increasingly used in the fields of evaluation, health sciences, and education in order to meet the diverse information needs of funders and stakeholders. However, a consensus has yet to develop on the theoretical underpinnings of the methodology. A side-by-side assessment of two competing theoretical approaches to mixed methods, the dialectic and pragmatic, can assist researchers to optimize their use of mixed methods methodology and contribute to the growth of mixed methods theory.

This study empirically compares the dialectic and pragmatic approaches to mixed methods and probes key issues underlying the methodology, including unique yield from mixed method studies, the importance of paradigmatic divergence between methods, and the financial demands of mixed method studies. A secondary analysis of a real-world evaluation, this study explores five research questions regarding the convergence, divergence and uniqueness of single method findings; the extent to which mixed methods produce unique findings over and above single methods presented side-by-side; the extent to which studies meet key criteria for validity; stakeholders' perceptions of the utility and credibility of the studies; and the cost of single methods.

The pragmatic mixed method study was developed by integrating a post-positivistic telephone survey with weakly interpretive focus groups at the point of interpretation using pragmatic criteria. The dialectic study mixed the same post-positivistic telephone survey with strongly interpretive phenomenological interviews

using a Hegelian-inspired dialectic format. All three single methods were examined by a method expert in the field who affirmed the methodologies used.

Findings suggest that both mixed method approaches produced unique conclusions that would not have been available by presenting single methods side-by-side. However, the dialectic method produced more complex convergence and more divergence, leading it to be more generative than the pragmatic method. The use of stronger as compared to weaker interpretive methods contributed to the generative quality of the dialectic approach. Overall, the dialectic method appears more suitable to exploring more complex phenomenon as compared to the pragmatic approach. However, these conclusions are drawn from one study of one real-world evaluation. Much more scholarship is needed to explore the issues raised here.

## TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION .....	1
1.1. Statement of the Problem .....	1
1.1.1. The Need for Research on Mixed Methods .....	1
1.1.2. Two Competing Theories of Mixed Methods .....	4
1.1.3. Conclusions .....	6
1.2. Purpose of the Study .....	7
1.3. Conceptual Framework .....	8
1.4. Research Questions .....	8
1.5. Research Design .....	12
1.6. Overview of the Dissertation .....	12
CHAPTER 2 LITERATURE REVIEW .....	14
2.1. Paradigms .....	14
2.1.1. Defining Three Key Paradigms along a Continuum .....	15
2.1.2. Defining Terms .....	18
2.1.3. The Relationship between Methods and Paradigms .....	21
2.2. An Historical Review of Mixed Methods .....	27
2.2.1. Mixed Methods from a Post-Positivist Perspective .....	28
2.2.2. The Growth of Qualitative Methods and the Emergence of Mixed Methods .....	29
2.2.3. Triangulation and Role of Qualitative Methods in Mixed Methods .....	30
2.2.4. The Development of the Dialectic Approach to Mixed Methods .....	33
2.2.5. An Emerging Pragmatic Approach to Mixed Method Studies .....	41

2.2.6. Dialectic and Pragmatic Approaches as Generative Endeavors .....	42
2.3. Applying Pragmatism to Mixed Methods .....	43
2.4. Applying Dialectics to Mixed Methods .....	48
2.5 Validity.....	56
2.5.1. Trustworthiness: A Unified Framework .....	57
2.5.2. The Standards for Program Evaluation .....	71
 CHAPTER THREE RESEARCH METHODS.....	 74
3.1 Overview of Methodology .....	74
3.2 Timeline for Secondary Data .....	76
3.3 Research Approach and Rationale .....	78
3.4. Research Questions .....	79
3.4.1. Research Question 1: Comparison of Single Method Results to One Another .....	80
3.4.2. Research Question 2: Comparison of Mixed Method Results to Each Other and Single Methods.....	82
3.4.3. Research Question 3: Examination of Validity of Inferences for Studies.....	88
3.4.4. Research Question 4: Stakeholder Views of Credibility and Utility .....	89
3.4.5. Research Question 5: Cost of Single Methods.....	91
 CHAPTER FOUR RESULTS.....	 93
4.1 Research Question 1 .....	93
4.1.1. Summary of Single Method Findings .....	93
4.1.2. Convergence, Divergence and Unique Findings across Methods.....	103
4.2. Research Question 2.....	116
Findings of Pragmatic and Dialectic Mixed Method Studies .....	117



Comparison of Mixed Method Process and Findings .....	140
What unique information do the mixed method findings produce over and above single methods?.....	147
4.3. Research Question 3 .....	150
4.3.1. Trustworthiness / Validity .....	152
4.3.2. The Joint Committee's <i>Program Evaluation</i> <i>Standards</i> .....	171
4.3.3. Expert Review of Single Methods .....	179
4.4 Research Question 4.....	189
4.4.1. Respondent Experience with Single Methods.....	190
4.4.2. Perceptions of Credibility of Single Methods.....	191
4.4.3. Perceptions of Credibility of Mixed Methods.....	194
4.4.4. Reported Utility of Mixed Methods .....	196
4.5. Research Question 5 .....	199
Cost in Billable Researcher Dollars .....	199
Cost in Subject Hours.....	202
 CHAPTER FIVE CONCLUSIONS AND DISCUSSION.....	 204
5.1 Research Questions Answered .....	206
5.1.1. Research Question 1 .....	206
5.1.2. Research Question 2.....	210
5.1.3. Research Question 3.....	212
5.1.4. Research Question 4.....	216
5.1.5. Research Question 5 .....	217
5.2 Conclusions .....	217
5.3. Limitations .....	229
5.4. Recommendations .....	232
 REFERENCES .....	 221

## APPENDIXES

Appendix A. Stakeholder Survey .....	235
Appendix B: Single method findings by topic area .....	250
Appendix C: Convergence, divergence and uniqueness of all single method findings by topic area.....	263
Appendix D: Extent to which single methods meet the Joint Committee's <i>Program Evaluation Standards</i> .....	276
Appendix E. Expert critique of single methods .....	284

## LIST OF TABLES

Table 1.	Validity and Trustworthiness for Post-positivist and Interpretive Paradigms .....	66
Table 2.	Format to Apply Hegel’s Dialectic Approach to Divergent Mixed Method Findings .....	85
Table 3.	Format to Apply Pragmatic Criteria to Pragmatic Single Method Findings .....	87
Table 4.	Frequency of Findings in Each Mixed Method Relationship Group.....	114
Table 5.	Frequency of Findings in each Mixed Method Relationship Group by Method .....	115
Table 6.	Pragmatic Mixed Method Decisions .....	120
Table 7.	Dialectic Mixed Method Decisions.....	131
Table 8.	Key Conclusions by Study .....	148
Table 9.	Summary of Lincoln and Guba’s (1985) Unified Theory of Validity / Trustworthiness.....	153
Table 10.	Stakeholders’ Previous Experience with Methods.....	191
Table 11.	“Other” Comments on Experience with Methods by Method .....	191
Table 12.	Rating Points and Labels for Stakeholder Survey.....	192
Table 13.	Stakeholder Ratings of the Validity of Findings by Method and Respondent .....	192
Table 14.	Stakeholder Ratings of Method Mixes.....	195
Table 15.	Comments on the Impact and Potential Impact of the Studies and the Findings that Lead to them, by Method Mix .....	197
Table 16.	General Comments on Mixed Methods’ Use.....	198
Table 17.	Total Billable Researcher Hours Expended by Method.....	200

Table 18. Dollars Worked and Percent of Budget for Study Tasks by Method .....	201
Table 19. Cost in Subject Hours for Recruitment and Completion.....	203
Table B1. Survey Study Findings by Topic Area .....	250
Table B2. Focus Group Study Findings by Topic Area.....	253
Table B3. Interview Study Findings by Topic Area .....	258
Table C. Convergence, Divergence, and Uniqueness of All Single Method Findings by Topic Area .....	263
Table D1. Extent to which Single Methods Meet Utility Standards .....	276
Table D2. Extent to which Single Methods Meet Feasibility Standards .....	278
Table D3. Extent to which Single Methods Meet Propriety Standards .....	279
Table D4. Extent to which Single Methods Meet Accuracy Standards .....	281

## LIST OF FIGURES

Figure 1. Paradigm Continuum .....	16
Figure 2. Study Methods Mapped to Paradigm Continuum.....	27
Figure 3. A Continuum of Triangulation Design (Jick, 1979) .....	31
Figure 4. Timeline for Key Intervention Dates, Evaluation Data Collection, and Reporting .....	77

## CHAPTER 1

### INTRODUCTION

First, this section describes the problem addressed by this research study and the purpose of it. Next, the conceptual framework used in the study is described. The research questions and design are presented next. Finally, an overview of the dissertation chapters is provided.

#### Statement of the Problem

##### *The Need for Research on Mixed Methods*

Local, state, and federal governments invest considerable public resources to address the education, health, and welfare of residents of the United States. The government bodies that disperse these resources and the citizens that fund them have a compelling interest in understanding the extent to which funds are used efficiently and effectively. Mixed method research and evaluation is a tool commonly used by researchers and evaluators to investigate program or policy merit and worth (Creswell, Trout, & Barbuto, 2002; Teddlie & Tashakkori, 2003).

Mixed method methodology is frequently used to meet the needs of multiple stakeholders or the individuals or groups who comprise the audience for evaluative and research endeavors (Chelimsky, 1997; Smith, 1997). Especially in the field of evaluation and in research on public policy, an investigator seeks to gather information in the service of clients or constituents who wish to make programming or policy

decisions based on the results. Frequently, decisions are made not by one executive, but by multiple individuals who represent diverse interests and hold unique perspectives on what kind of information is accurate and credible (Patton, 1997). Like researchers, stakeholders hold beliefs about what kinds or types of information are most accurate and authentic, and will best support decision-making. Either implicitly or explicitly, some stakeholders believe quantitative studies provide the most reliable information, while others view qualitative data as the most authentic representation of reality, and thus the best source of information for decision-making. In many situations, a combination of types of information provides multiple stakeholders the type of information that they have the most confidence in for use in decision-making (Patton, 1997; Chelimsky, 1997; Benofske, 1995).

Additionally, researchers turn to mixed method methodology to address the practical challenges and resultant uncertainty of conducting any single method (Datta, 1997; O’Cathain, Murphy, & Nicholl, 2007). Both post-positivist and interpretive methods have serious limitations. For example, Carole Weiss (1995) described the challenges to conducting the gold standard of post-positivistic research – the randomized controlled experiment – on complex community initiatives. First, likely too few communities could be marshaled for randomization in treatment and control groups for interventions administered at a community level. Second, controlling for key factors in community initiatives, such as a dynamic interplay of government, funder, and grassroots support for issues, would be difficult and ethically questionable. Finally, due to external factors, such as economic and political situations, community

initiatives and policies may be enacted or altered or repealed during the course of an investigation, making positivist approaches less effective.

The difficulties in conducting post-positivist research to investigate complex phenomenon, such as community initiatives or policies, may suggest that interpretive approaches would be more effective. However, in these settings, interpretive methodologies also face significant challenges. A key strength of interpretive approaches is the ability to understand a phenomenon in depth. However, the impact of large-scale interventions is frequently so wide as to make the sole use of interpretive approaches formidable. Additionally, the mechanisms to identify causal mechanisms in interpretive research require further development (Smith, 1994; Johnson & Onwuegbuzie, 2004). By using multiple, diverse methods, researchers may corroborate findings to increase confidence in the inferences drawn from them. This rationale applies equally to smaller or more bounded objects of study whose complexity lies in the content of the phenomenon being studied.

Finally, researchers use mixed methods in order to achieve findings unavailable to single method studies conducted independently. Greene (2007) described that mixed method studies may be generative, as paradox and contradiction are engaged and “fresh insights, new perspectives, and original understandings” emerge (2007, p. 103). Other mixed method authors share this belief in the promise of mixed methods. For example, Abbas Tashakkori and Charles Teddlie (2003) used the term *gestalt* to indicate the how inferences from mixed methods may be greater than the single method components.



Rosaline Barbour (1999) described mixed methods as a whole greater than the sum of its parts”.

### *Two Competing Theories of Mixed Methods*

The predominant theory of mixed method evaluation is the dialectic approach developed by Jennifer Greene and Valerie Caracelli (1997). The purpose of the dialectic approach is to gain insight by juxtaposing methods conducted using clearly defined and diverse research paradigms (for example, post-positivistic, phenomenological, ethnographic, etc.). Given the importance of paradigms in the dialectic approach, an unstated assumption is that differing paradigms may increase the variance between types of evidence, thus increasing the utility of findings and the validity of inferences drawn from them. Thus, the “distance” between paradigms of diverse methods may be critical to mixed method studies. However, an enhanced focus on paradigms within the dialectic approach faces three primary challenges.

First, the empirical literature on the use of a dialectic approach to mixed methods is sparse. The bulk of the literature on mixed methods and on the use of paradigms in mixed methods develops typologies for types of mixed method studies (Creswell, Trout, & Barbuto, 2002; Greene & Caracelli, 1989; Greene & Caracelli, 1997; Rossman & Wilson, 1985). Only one empirical study published in a refereed journal was uncovered. Jennifer Greene and Charles McClintock (1985) asked if methods that differ paradigmatically might be equal when combined in a triangulated mixed method evaluation: an interesting question, but not one directly related to the rationale and practice of mixed method research and evaluation.

Second, the dialectic approach appears alternately poorly understood or misunderstood. In his 2001 assessment of 22 evaluation models, noted evaluator Daniel Stufflebeam rated a mixed method model as having restricted though beneficial use in program evaluation. However, his understanding of mixed methods did not include the work of leading theorists, Greene and Caracelli, nor their dialectic approach. Stufflebeam cited only more minor mixed method theorists in his monograph. In terms of the dialectic stance itself, research reviews suggest the dialectic stance is both misunderstood (Mark, Feller, & Button, 1997) and infrequently used (Riggin, 1997; Creswell, Trout, & Barbuto, 2002; Patton, 1985).

Third, a nascent yet growing body of work is focusing on a pragmatic approach to mixed methods. Lois-ellen Datta (1997) and Spencer Maxcy (2003) articulated a pragmatic stance to mixing methods that has its roots in the philosophic writings of John Dewey and William James (among others). Datta (1997) outlined the essential criteria for making pragmatic design decisions as (1) *practicality*, which implies one's experience and knowledge of what does and does not work; (2) *contextual responsiveness* to the demands, opportunities, and constraints to an evaluation situation; and (3) *consequentiality*, or making decisions based on practical consequences.

Although the pragmatic theory is evolving, researchers commonly employ a pragmatic stance in mixed methods. Michael Quinn Patton's (1985, 2008) utilization-focused evaluation is implicitly pragmatic in that it judges the merit of an evaluation by the extent to which it was useful to the clients. John Creswell (2003) reported that

pragmatism appears to be the dominant paradigm employed by mixed method researchers. Leslie Riggin (1997) found a pragmatic stance to be almost exclusively employed when she reviewed all examples of mixed method evaluations presented in a volume of *New Directions in Evaluation* dedicated to the subject. More recently, R. Burke Johnson and Anthony Onwuegbuzie (2004) suggested that “the time has come” for mixed method research, and that investigators do whatever is practical. However, dialectic and pragmatic practitioners of mixed methods alike conceded that pragmatic theory requires further development (Teddlie & Tashakorri, 2003; Morgan, 2007; Greene, 2007). None the less, the evolving theory of pragmatism challenges the primacy of Greene and Caracelli’s dialectic theory and deserves further examination.

### *Conclusions*

While stakeholder and research considerations suggest a strong need for mixed method research and evaluation, methodological literature on mixed methods is nascent. The rationale that mixed methods yields unique insight from qualitative or quantitative studies conducted independently requires further investigation. Given the still evolving dialectic and pragmatic approaches to mixed methods, researchers and evaluators would benefit from additional guidance in how to optimize the design, implementation, and interpretation of mixed method studies. Additionally, mixed method research and evaluation frequently requires more resources to implement than single method studies. The additional cost of mixed method research warrants a more explicit assessment of the rationale for mixed methods and its optimal design and implementation.

Empirical as opposed to theoretical investigations of dialectic and pragmatic approaches to mixed methods is especially needed in order to legitimize and optimize mixed methods. Evidence that the theoretical rationale for varying paradigms in mixed methods is or is not justified would support or challenge the dialectic approach's status as the predominant theory. An empirical comparison of the dialectic and pragmatic approaches – even in one context – would arm mixed method practitioners with valuable information for practice and would advance the field's understanding of mixed methods and hopefully lead to higher quality mixed method studies.

#### Purpose of the Study

The purpose of this study is to assist researchers in optimizing their mixed method research designs by examining two real-world mixed method studies, each representing one of two competing theories of mixed methods methodology. In this quest, this study also probes underlying assumptions and rationales of mixed methods: that mixed method studies yield findings over and above single methods presented side-by-side, that the paradigmatic divergence of methods is a critical factor in mixed method studies, that mixed method studies can better meet the demands of multiple stakeholders with differing opinions on the usefulness and credibility of qualitative and quantitative research methods, and to examine the increased financial demands of mixed method studies.

## Conceptual Framework

The conceptual framework guiding this dissertation is the two competing theories of mixed methods. The predominant theory of mixed method evaluation is the dialectic approach developed by Greene and Caracelli (1997), as described above. Likewise, the pragmatic approach to evaluation as articulated by Datta (1997), Maxcy (2003), Teddlie & Tashakkori (2003), and Johnson and Onwuegbuzie (2004) is described above. The two approaches differ primarily in their treatment of paradigms. While the dialectic stance prioritizes consciously choosing and engaging paradigms in the conduct of mixed methods research, pragmatically based mixed methods respond not to philosophical tenets, but to a grounded reality of practicality, contextual responsiveness, and consequentiality.

## Research Questions

Additional empirical research in five areas would probe the rationale for using paradigms in mixed methods and provide information on the optimization of mixed methods approaches. First, an underlying assumption of the dialectic approach to mixed methods is that paradigms matter and that more paradigmatically diverse methods may result in more generative findings. However, no empirical information is available on comparing findings from paradigmatically similar and dissimilar methods. While researchers assume a common knowledge of differences in findings based on paradigms, the literature appeared to provide few studied comparisons outside of theoretical argument. Additionally, both William Shadish (1993) and Melvin Mark and

Lance Shotland (1987) emphasized the importance of understanding the direction of bias among single methods. Systematically recording the convergence and divergence of findings from single methods that comprise mixed methods in a specific research context would produce a foundation for further insight into the impact of dialectic versus pragmatic mixed method approaches, and would provide evidence to help researchers optimize their choice of single methods in a mixed method study.

Second, the dialectic and pragmatic approaches to mixed methods represent two primary approaches to mixed methods. Literature on the implementation of practice of the dialectic approach is scant (Riggin, 1997; Mark & Shotland, 1997), while researchers commonly recognize the need to further develop the pragmatic approach in both theory and practice (Greene, 2007; Tashakkori & Teddlie, 2003). Examining the extent to which dialectic and pragmatic mixed method studies differ substantively in a specific research context would provide researchers with empirical evidence to optimize mixed method practice. A comparison of the approaches would also illuminate the importance of differing paradigms as a part of mixed method design and implementation, and at the point of interpretation and use of findings. Finally, comparing mixed method findings to single method findings to mixed method findings would illuminate the extent to which mixed method findings yield unique insights over and above the presentation of single method findings side-by-side.

Third, inferences from mixed method studies can only be as legitimate as the inferences from the single methods upon which they are based (Greene, 2007; Teddlie & Tashakkori, 2003). Therefore, to assess the validity / trustworthiness of the

inferences from the mixed method studies produced, the validity / trustworthiness of the three single methods' inferences will be examined. The topic of validity within mixed methods is nascent, in terms of both how validity should be conceptualized in mixed methods and in criteria for judging it (Greene, 2007; Dellinger & Leech, 2007; Creswell & Plano Clark, 2007; Teddlie & Tashakkori, 2003). However, examining mixed method findings against key criteria for validity in a specific research context is a first step. Patton (2002) suggested alternative criteria for validity that may be applied to mixed methods; the Joint Committee's *Program Evaluation Standards* (1994) will be examined.

Fourth, a key rationale for mixed methods is that it meets the multiple information needs of diverse stakeholders (Benofske, 1995; Patton, 1997). Examining stakeholders' perceptions of the credibility and utility of mixed method findings would provide additional empirical evidence to support this claim. Stakeholder perceptions of the credibility and utility of mixed method evaluation findings can also contribute to an understanding of the validity of dialectic versus pragmatic mixed method evaluations. Therefore, examining stakeholders' views of the credibility and utility of individual methods and dialectic versus pragmatic mixed method evaluation findings in a specific research context would provide valuable information about the rationale and validity of mixed methods.

A final foundational consideration in mixing methods is quintessentially pragmatic and rooted in the Joint Committee's *Program Evaluation Standards* (1994) and the standard of feasibility. Little empirical literature exists on the financial

feasibility of mixed methods. Understanding the cost of single methods that comprise mixed methods in a specific research context in terms of researcher and subject hours and resources expended provides additional valuable information to investigators and they choose between dialectically- or pragmatically-driven approaches.

Taken together, these five research questions provide background and evidence to probe the dialectic versus pragmatic approaches in real-world mixed methods research and evaluation, and to explore several of the assumptions, rationales and issues that under gird mixed methods. The research questions include the following:

1. What are the substantive findings of single methods? What findings converge and diverge? What findings are unique?
2. What are the substantive findings of pragmatic versus dialectic mixed method studies? How are the two mixed method study findings similar and different from one another? What unique information do the mixed method findings produce over and above single methods?
3. To what extent are the inferences drawn from single method findings meeting key criteria for validity / trustworthiness? To what extent are inferences drawn from single method and mixed method findings valid / trustworthy according to *The Program Evaluation Standards* (Joint Committee, 1994)?
4. How do stakeholders view the credibility and utility of single method findings and mixed method findings? What do they see as the advantages and disadvantages of mixing? What are their prior beliefs about the credibility of diverse methods and paradigms?



5. What are the costs of the single methods in terms of researcher and subject hours?

### Research Design

This dissertation is a secondary analysis of data from a real-world evaluation of local, smoke-free ordinances on Minnesotans trying to quit and enrolled in a specific stop-smoking program. The evaluation involves three single methods that are combined into two mixed method studies for this dissertation. The three studies include an 18 month follow-up telephone survey with comparison groups, phenomenological interviews, and focus groups. The specific methodology for each method was reviewed by an expert in that methodology to provide evidence of content validity. The first mixed method study represents a dialectic approach to mixed methods and combines the survey and phenomenological interviews. The second mixed method study represents a pragmatic approach and combines the survey and focus groups. The substantive findings of single and mixed method studies are compared via content analysis of evaluation documents. The studies are also examined for the extent that they meet key criteria for validity. Stakeholders' views on the utility and credibility of single and mixed methods is examined. Finally, the cost of single methods is considered.

### Overview of the Dissertation

Chapters 2 through 5 comprise the remainder of this dissertation. Chapter 2 reviews the pertinent literature of paradigms and mixed methods, probing the rationale,

theories, and methodological research in the area. Chapter 3 describes the methodology of the dissertation. Chapter 4 presents the results. The conclusions and limitations of the research are discussed in Chapter 5.

## CHAPTER 2

### LITERATURE REVIEW

This literature review comprises five sections representing concepts critical to the conduct of this dissertation. First, the three key paradigms relevant to mixed methods are defined, and the relationship of the paradigms to the three methods used in this dissertation are specified. Second, an historical review of the development of mixed methods is presented, leading to a description of the two primary approaches promoted within the field of mixed methods today. The literature review describes how further consideration of these two approaches, called the pragmatic and dialectic stances, is needed, despite their ascendancy. To explore how a pragmatic mixed method study would be conducted, the third section presents an in-depth review of pragmatism. Likewise, the fourth section describes an in-depth review of a dialectic approach. Finally, the concept of validity or trustworthiness is explored within the context of mixed methods studies in the fifth section.

#### Paradigms

The purpose of this section is to define the paradigm continuum and the relationship of specific research methods to paradigms, especially for the three research methods used for this dissertation. Paradigms refer to a worldview that guides decision-making. Popularized by Thomas Kuhn (1962), they encompass one's views on the nature of reality and of knowledge, its origins and foundations (Greene & Caracelli, 1997). A paradigm is essentially philosophical in nature, and may be specified by its

ontological, epistemological, and axiological tenants. For researchers, one's paradigm informs the research questions one chooses, and how one collects information and interprets it.

### *Defining Three Key Paradigms along a Continuum*

This section defines three key paradigms and situates them on a continuum. Logical positivism, post-positivism, and interpretivism are examples of paradigms, which may be considered on a continuum (see Figure 1, below). Anchoring one end is logical positivism. Introduced by French philosopher August Comte (Yu, 2006), logical positivism holds that truth is represented by measurable, naturally occurring phenomenon. In fact, logical positivism asserts that measurement is proof of existence, so if a phenomenon cannot be measured, than it does not exist (Potts, 1998). Further, logical positivism argues that all naturally occurring phenomenon can be broken down into measurable moments, which when considered together form the whole of the phenomenon of interest and reproduce "truth". Logical positivist researchers use deductive reasoning to generate theory from which specific hypotheses evolve and are tested. Inferences from experiments are then employed in theory construction and the development of natural laws (Yu, 2006; Benofske, 1995). Contemporary researchers universally agree that logical positivism consists of numerous irreconcilable fallacies, and is dead (Reichardt & Rallis, 1994a; Shadish, 1998).



Figure 1. Paradigm Continuum

Post-positivism is a softening of the logical positivist position that has been evolving since the 1930s (Popper, 1959, as cited in Reichardt & Rallis, 1994). The post-positivist philosophy asserts that truth may be discovered, and is best understood through objectivity, standardization, deductive reasoning, and control within the research process (Yu, 2006). Causality is a central concern of post-positivist research techniques, and is established by research design, statistical hypothesis testing, and energetically assessing alternative possible explanations for findings. The randomized controlled experiment is considered the ideal. Validity of inferences from findings is assessed by internal validity, external validity, reliability, and objectivity. The strengths of post-positivist research are precision, generalizability, reliability, and replicability. Post-positivist research focuses on addressing causality in research questions and is commonly considered to be well suited for confirmatory research (Shadish, Cook & Campbell, 2002).

Mary Lee Smith (1994) carefully considered the shortcomings in both quantitative (post-positivistic) and qualitative (interpretive) research and suggests areas for substantive improvement in each realm. She argued that quantitative research could be enhanced and refined by considering its applicability in highly complex,

unpredictable situations. She also urged quantitative researchers to consider the impact of research over-design on program development, and to avoid technically correct but sterile studies. Perhaps most persuasively, she pointed out the need for additional consideration of over-analysis (inflation of Type I error rates) and the impact of attrition on study findings. I would argue that another area for consideration in Educational Psychology is the assumption of independence of subjects that underlies all inferential statistics, especially in studies where students in classrooms are considered to be the object of study. While the emergence of hierarchical linear modeling techniques and the use of mean studies address this concern, much quantitative educational research focuses on student-level outcomes but fails to address inter-dependence of subjects.

The third paradigm considered here is pure interpretivism, which may be found on the opposite end of the paradigm continuum from positivism. In its most extreme form, interpretive research contends that reality is constructed and that no universal truth exists. More broadly, interpretivism asserts that multiple truths exist, as determined by individuals' unique perspectives on the world. Interpretive research, then, illuminates individuals' perspectives and experiences. In interpretive approaches, truth is best understood through research conducted in natural settings where the researcher is close to the research participant, and through critical subjectivity and inductive reasoning (Bednarz, 1985; Smith & Heshusius, 1986). Interpretive approaches emphasize thick description, and utilize the researcher as the chief instrument in data gathering and analysis (Van Manen, 1990). Interpretive researchers

prioritize exploring the ontological, epistemological, and axiological aspects of their inquiry; as such, specific methodologies (i.e., phenomenology, social interactionism, ethnomethodology) may be considered to have their own paradigms (Creswell, 1998; Bednarz, 1985). Validity may be assessed by the credibility, transferability, dependability, and confirmability of findings (Lincoln & Guba, 1985). The strengths of interpretive research include a strong understanding of context, rich detail, and flexibility to address emerging issues. Interpretive research is commonly considered to be well suited for exploratory research, especially uncovering the “how” and “why” of phenomenon.

Smith (1994) also considered areas for improvement in qualitative researchers. She urged researchers to carefully consider and differentiate between subjectivity and bias, and to carefully plan and interpret findings in light of sample selection. Qualitative research is increasingly being used to make policy and programming decisions. As such, Smith called for qualitative researchers to develop methods to determine attribution, influence and causality, as well as methods to synthesize findings and conduct secondary analysis of data. Perhaps most persuasively, she called for increased methodological transparency among qualitative researchers.

### *Defining Terms*

The literature of mixed methods often uses the terms “paradigm,” “quantitative / qualitative” and “method” interchangeably (Jick, 1979; Rossman & Wilson, 1985; Reichardt & Rallis, 1994a). This section defines those terms in order to reduce the confusion of imprecise terminology, to advance more nuanced distinctions in this

discussion, and to increase the the precision with which the mixed method methodology may be discussed.

As described above, paradigms refer to a worldview that guides decision-making, encompassing one's views on the nature of reality and of knowledge, its origins and foundations. Greene (2007) viewed paradigms to be primarily philosophical, however, Leonard Bliss (2008) argued that paradigms includes a researcher's practical experiences and subjective predispositions. Despite the difference, both authors agree that paradigms dictate a comprehensive worldview that guides decision-making.

Different researchers use the use of the terms *quantitative* and *qualitative* in fundamentally different ways. For example, John Creswell, Shirley Trout, and John Barbuto (2002) and Kate Lynch (1983) described quantitative data as including numbers, whereas qualitative data include words, symbols, pictures and other non-numeric data. This is the common understanding of these terms in texts that broadly review research design in the social sciences (Johnson & Christensen, 2000; Simon & Burstein, 1985) and in evaluation (Wholey, Hatry, Newcomer, 2004; Patton, 2008).

However, many mixed method authors used the terms quantitative and qualitative in a way that conflates the concepts of data type, methods and paradigm. For example, a closed-ended survey conducted under a post-positivist paradigm would be referred to as *quantitative*, while an interview conducted under an interpretive paradigm would be referred to as *qualitative* (Jick, 1979; Rossman & Wilson, 1985;



Reichardt & Rallis, 1994a; Smith, 1997). This use of the terms quantitative and qualitative act as a rough proxy for the concept of paradigm.

The undefined way in which the terms quantitative and qualitative are used reflects, in part, the broad nature of mental models as described by Greene in her recent (2007) book on mixed methods. Greene described mental models as “a set of assumptions, understandings, predispositions, and values and beliefs with which all social inquirers approach their work” (p. 12). These models are shaped by a wide range of factors including educational and professional experiences and personal beliefs and values. For Greene, mental models “thus subsume philosophical paradigms, as well as substantive theories, disciplinary perspectives, and a whole host of more personalized experiences, values and ways of knowing” (p.13). The models exist even before an evaluator defines and formalizes her theoretical approach to a project .

In contrast to the complexity of paradigms and the multiple meanings of *quantitative / qualitative*, the term *method* is much more straight-forward. Methods are most commonly defined simply as strategies for collecting data, such as interviews, paper and pencil surveys, participant observation, secondary data analysis, etc. (Lynch, 1983; Greene & Caracelli, 1997; Creswell, 2003). It is important to note that methods are not synonymous with data type. For example, a paper and pencil survey may collect numerical data as well as open-ended comments in the form of anecdotes or stories, and interviews may collect stories as well as numerical information.

### *The Relationship between Methods and Paradigms*

This section defines the relationship of methods to paradigms. Next, the relationship is applied to the three single methods in this dissertation in order to establish the particular relationship of each method to its corresponding paradigm.

The relationship between methods and paradigms is complex. On one hand, some data types and methods strongly lend themselves to best use under a specific paradigm. For example, the context, thickness of description, and richness of data required by phenomenological interviewing (an interpretive research strategy) is unobtainable through closed-ended surveys. Likewise, the reliability, validity and objectivity required by post-positivist research is impossible to obtain through phenomenological interviewing. However, the relationship is not constant. For example, a structured interview may be conducted using closed ended items under a positivist paradigm. Similarly, under a phenomenological (interpretive) paradigm a researcher may augment experience-driven interviews with a survey that includes both closed- and open-ended items. The key point is that neither method nor data determines paradigms, and vice versa (Greene & Caracelli, 1985; Shaffer & Serlin, 2004).

### *The Relationship of Dissertation Methods to Paradigms*

Three single methods are used in this dissertation: an 18-month follow-up telephone survey, focus groups and phenomenological interviews. The relationship of method to paradigm is discussed for each of these methods next.

*Telephone Survey.* The 18-month follow-up telephone survey used in this dissertation was administered to treatment and comparison groups at enrollment in the intervention and again at six and eighteen months after enrollment. The survey was guided by a post-positivistic paradigm. The goal of the survey was to obtain valid and reliable estimates of participants' tobacco use given the constraints of the context of the study. It also assessed associated behaviors and attitudes in a standardized format. The aim was to draw conclusions about the impact of smoking regulations on subjects who try to quit smoking. The survey sought to infer from a sample of participants in a statewide stop-smoking intervention to all tobacco users in Minnesota who have attempted to quit using a similar intervention. Logistic regression analyses were conducted.

*Focus Groups.* Focus groups were designed to gather the stories, opinions and personal testimony from a larger group of participants than could be obtained through interpretive methodologies such as interviews. The groups were conducted and analyzed according to the pragmatically-oriented methods of Richard Krueger (Krueger & Casey, 2000; Krueger, 1998). The groups were designed to address the extent to which local smoke-free ordinances impacted quitting behaviors among those enrolled in a stop-smoking intervention. Participants were recruited from the 7-county Twin Cities, Minnesota metropolitan area. In a completely crossed design, participants were sampled based on their geographic location (Hennepin County, Ramsey County, 5 County metropolitan area) and their tobacco use (current smokers, former smokers).

Only participants with the same geographical location and tobacco use status participated in the same focus group with each other.

Several factors suggest that the focus groups were conducted using an interpretive paradigm. First, the data gathered by the focus groups is qualitative (words, body language and intonation), not quantitative (numbers). The format was open-ended, so participants could speak in their own words. The information gathered by the focus groups provided important context, which is a feature prioritized in qualitative methods. Where post-positivistic methods seek to infer from a sample to a population using hypothesis testing and statistical techniques, the focus group approach clearly does not meet this standard. Finally, in terms of design, the facilitator actively guided the group in a relationship closer than would be optimal in post-positivistic frameworks.

At the same time, several factors suggest that focus group method is only weakly interpretive in approach. The goal of reaching a larger number of participants to achieve greater representation from the population reflected a more post-positivist approach, and the use of a crossed design borrowed heavily from the post-positivist tradition. Perhaps most importantly, however, the focus groups brought six participants, on average, together for a two-hour conversation that lacked the depth and richness necessary for an intensely interpretive approach. Taken together, the evidence suggests that the focus group method is indeed interpretive, but only weakly.

*Phenomenological Interviews.* The phenomenological interviews were designed to gather in-depth qualitative data about participants' experiences quitting and of

smoking ordinances in two variations: for those who have quit and those who continued using tobacco after a failed quit attempt. Interviews were guided by a descriptive phenomenological paradigm (Van Manen, 1990; Giorgi, 1997; Polkinghorne, 1989) and addressed the following question: What is the experience of the smoking ban for people who have quit or have tried to quit? Data were analyzed according to descriptive phenomenological strategies outlined in Giorgi (1997) and Polkinghorne (1989).

Like focus groups, the interviews sampled from the 7-county Twin Cities, Minnesota metropolitan area so that participants were approximately equally distributed between Hennepin County, Ramsey County, and the 5-County suburban metropolitan area. Additionally, former smokers were over sampled to adequately represent their experience, and only those reporting attending a bar or restaurant one or more times per month were eligible to participate.

The phenomenological approach is clearly an intensely interpretive approach. Considered to be its own paradigm, phenomenology encompasses epistemology, ontology, and axiology; it is a research practice and a way of understanding the world. Richard Parker (1969) described that phenomenology is the science and art of understanding what being human means through language and is a quest to understand the meaning of a phenomenon for those who experience it (Van Manan, 1990). As being human is to be concerned with meaning, phenomenology is best poised to answer inquiries into the meaning and the significance of phenomena. Phenomenological inquiry is designed to more sensitively orient one to a phenomenon,

so that one might act “more thoughtfully and more tactfully in certain situations” (Van Manan, 1990, p.23).

The Human Science of phenomenology developed in reaction to the tradition of positivist philosophy and science that flourished and expanded to encompass both physical and social sciences in the nineteenth and twentieth centuries. According to Van Manan, the quality of positivist science is judged by its precision and exactness, and the extent to which the research may be judged as objective, or removed from influencing the object of study. Phenomenology provides an opportunity to re-conceptualize the criteria for research quality and validity. In phenomenology, precision and exactness are fulfilled by interpretive description that exacts the “fullness and completeness of detail” that is discovered through closeness to the phenomenon of interest – a subjectivity realized by acute perception, insight and discernment that reveals the phenomenon in its “full richness and greatest depth” (Van Manan, 1990, p. 17). In phenomenology, objectivity refers not to the researcher’s degree of influence on the phenomenon of study, but on the extent to which he or she remains true to the phenomenon (Van Manan, 1990).

The phenomenological interviews were conducted using several techniques specific to phenomenology and allied with interpretive methods more broadly. Specifically, I sought to adopt a scientific attitude by placing myself in a state of reduction, and bracketing my preconceptions (Giorgi, 1997). I sought to consider the phenomenon of smoke-free ordinances “precisely as it is given” (Giorgi, 1997, p. 237), unmediated by reflection, intellectualization, desire to please the researcher, or

explanation. I sought to conduct the interview in a posture of wonder and openness, seeking concrete details with an eye toward action (Van Manan, 2006) in order to uncover the invariant meaning or essence of the ordinances for people trying to quit smoking (Giorgi, 1997).

Finally, two overlapping yet distinct forms of phenomenology may be practiced: descriptive and hermeneutic. Descriptive phenomenology was practiced in the interviews for this dissertation and is designed to elicit lived experiences, usually through interviews or written accounts, and explores the meaning of a phenomenon through a general structure and its constituents (Giorgi, 1997). While the goal of descriptive phenomenology is to discover meaning through deep and profound description, hermeneutic phenomenology uses a more interpretive approach that includes exploration of other texts, including but not limited to etymological inquiry, literature, and the experiences and reflections of the author. Despite the differences between descriptive and hermeneutic phenomenology, the two approaches share many features and are not mutually exclusive. Indeed, the primarily descriptive phenomenological approach used for the interviews in this dissertation included an etymological analysis more typical of a hermeneutic approach.

*Summary of Mapping Methods to Paradigms.* Although the relationship of methods to paradigms is not constant, the paradigm-guided implementation of the specific methods used in this dissertation, as described above, may be mapped onto the paradigm continuum shown in Figure 1, above. The 18-month follow-up telephone survey embodies a primarily post-positivistic paradigm, while the phenomenological

interviews embody an intensive interpretive paradigm. In contrast, the focus groups express a weaker interpretive paradigm that is closer to post-positivism on the paradigm continuum. For a graphic depiction of these relationships, please see Figure 2, below.

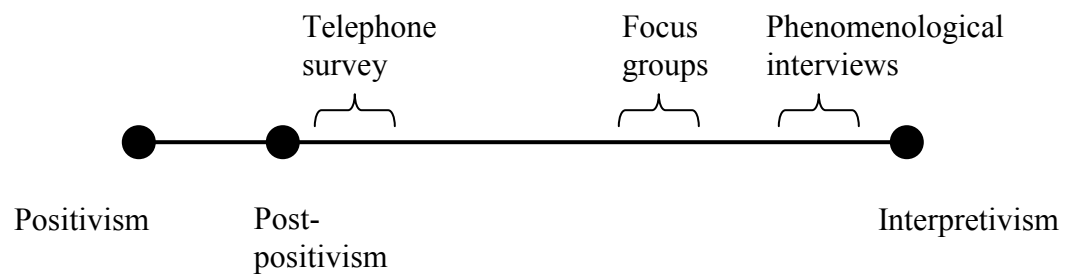


Figure 2. Study Methods Mapped to Paradigm Continuum

### An Historical Review of Mixed Methods

The purpose of this section is to introduce the historical development of mixed methods methodology in order to illustrate the role of paradigms in the field and the development of the dialectic and pragmatic approaches to mixed method research and evaluation.

Teddlie and Tashakkori (2003) provided a brief history of mixed methods. They revealed that a significant amount of mixed method research occurred in the traditional positivistic period of 1900-1950. For example, the famous Hawthorne and Yankee City studies, among other research projects, employed methods from both qualitative and quantitative research traditions. Teddlie and Tashakkori's review, in



addition to the review of other authors (Datta, 1994) suggested that researchers employing mixed method in the early to mid 20<sup>th</sup> Century did not identify their work as belong to the field of mixed method research, nor did they critically reflect on how or under what circumstances methods from different research traditions might be mixed. No overt discussion of paradigms is evident during mixed method research in this early time period.

### *Mixed Methods from a Post-Positivist Perspective*

The first explicit discussion of mixed methods research was initiated by Donald Campbell and Donald Fiske (1959) in their scholarship on a multitrait-multimethod matrix to strengthen validity of a trait. The authors employed correlational analysis on multiple traits gathered by multiple methods in order to demonstrate the independence of methods and traits. Convergent results suggest an enhanced belief that results are valid and not a methodological artifact. The concept that convergence suggests increased validity would be marshaled by later researchers as a rationale for triangulation and mixed methods.

Campbell and Fiske's (1959) conception of multi-method research differs considerably from a more current understanding of mixed methods research because it required two or more quantitative data sources and excluded the possibility of using any kind of qualitative data. As such, Campbell and Fiske's approach is of limited usefulness to a modern conception of mixed methods. However, it does point out some key dimensions along which mixed methods research may be judged.

First, Campbell and Fiske's analysis also required a construct from which alternate or parallel traits may be drawn. Additionally, Campbell and Fiske privilege concurrent and independent implementation of measures, to rule out threats to internal validity due to history and investigator bias. Their work also suggested that methods should have equal priority in terms of resources and time as it contributes to the overall methodological quality of results.

#### *The Growth of Qualitative Methods and the Emergence of Mixed Methods*

The evolution of mixed methods research was fueled by a resurgence among qualitative researchers in the 1960s and 1970's who questioned the dominant positivist paradigm. Positivist, logical positivist and interpretive philosophies, their practitioners, and cultural critics became engaged in enlightening, passionate, and ultimately divisive debates coined as "paradigm wars". In academia, the paradigm wars are philosophical, methodological, and political debates about the best ways to conduct research, including what research questions to ask, what information to gather, and how to collect and interpret the findings (Greene, 2007; Reichardt & Rallis, 1994b).

The critical discourse of the paradigm wars has expanded many researchers' understanding of their own approach, and the approach of others. The debates have highlighted the need for researchers to articulate the assumptions that underlie their work. On the other hand, the debates of the paradigm wars have also been contentious and divisive, sometimes vicious in nature. Many feel that in-fighting among colleagues and peers has become more destructive than constructive (Reichardt & Rallis, 1994a). The debates were so divisive that an entire arm of mixed method theory, discussed later

in this paper, debated the legitimacy and feasibility of mixing methods from different paradigms at all.

Scholars of mixed methods responded and in some cases sought to advance the debate between researchers practicing within interpretivist and post-positivist paradigms (Reichardt & Rallis, 1994a; Johnson & Onweugbuzie, 2004; Greene, 2007). The paradigm wars and researchers' contemporary understanding of mixed methods are tightly bound together, as the former fueled the development of the latter.

#### *Triangulation and Role of Qualitative Methods in Mixed Methods*

Norman Denzin and Todd Jick spearheaded an early expansion of mixed method literature in the late 1970's. The primary contribution of both Denzin and Jick was to more fully articulate the need for triangulation of data sources that mixed methods provides, and to embrace and for use of qualitative methods within a mixed methods approach.

The term triangulation was coined, according to Sandra Mathison (1988), by Charles Webb in his 1966 work on non-reactive measures, and refers to results of discrete methods converging to suggest increased validity of findings. In his 1978 article on triangulation, Norman Denzin broadened the concept to include triangulation on data (across time, space and persons), investigators, theory, and methods. Denzin is the first researcher to consider the role of paradigm (or theory, as he named it) in mixed method literature, however briefly. He suggested that triangulation of theory was problematic at best and potentially impossible due the constraints of a researcher's biases. No empirical evidence was provided. Denzin's argument mirrors researchers

adopting the stance that methods of diverse paradigms may not be mixed because their underlying assumptions are orthogonal to one another (Bednarz, 1985; Smith & Heshusius, 1986).

Perhaps more importantly, Denzin was the first researcher to encourage triangulation at the level of methods. He discriminated between within method and between method triangulation. Within method triangulation is exemplified by Campbell and Fiske (1959) who employ two similar data gathering methods (i.e., two quantitative surveys). In contrast, between methods triangulation would employ two or more diverse methods (i.e., a survey and an interview), holding theory (i.e., paradigm) constant.

Todd Jick (1979) advanced the discussion of triangulation by more fully incorporating qualitative methods and by more clearly applying the assumptions of Campbell and Fiske (1959) to a qualitative and quantitative mixed method approach. First, Jick developed a continuum of triangulation that attempted to outline purposes for triangulation and areas for further exploration (see Figure 3, below). He argued triangulation exploits the strengths of qualitative and quantitative methods, and counterbalances or compensates for their weaknesses. Implicit in the discussion

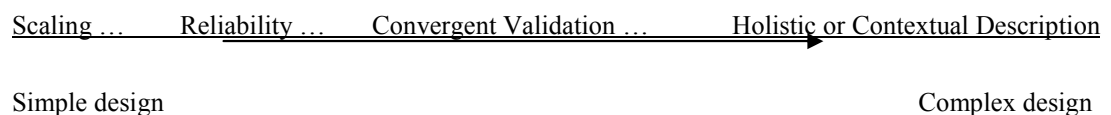


Figure 3. A Continuum of Triangulation Design (Jick, 1979)

is that the weaknesses of both quantitative and qualitative methods constituted their bias. Therefore, the key to mixed methods triangulation is to ensure that methods complement each other in strengths and that they do not share the same weakness or bias. Otherwise, bias would be compounded.

Based on a discussion of strengths and weaknesses of methods and the function of triangulation, Jick argued that mixed methods can uncover both overlapping variance (convergence) and unique variance (divergence). In so doing, Jick shifted Campbell and Fiske's conception of triangulation to define divergence not as methodological artifact, but as information with the potential to uncover new insights, or potentially a new layer of truth revealed by the use of a different method. Indeed, Jick grappled with how to analyze data and draw conclusions. He asked, how does a researcher know that results converge? In addition to calling for more research to answer this question, Jick responded that convergence is apparent. He placed the researcher as a builder and creator of plausible frameworks to explain findings, and argued that the ultimate claim to validity rests on judgment.

Lance Shotland and Mel Mark (1987) grappled with these same questions of interpretation of multiple methods; however, they assumed that all divergence represents methodological bias, which may be resolved by examining threats to validity. They proposed, in part, a long-term solution of meta-analysis that examines the direction and magnitude of bias both within and across methods, within specific research content areas. Shotland and Mark's rejection of divergence suggests that an

underlying assumption of their research paradigm may include a greater belief in a single, as opposed to multiple truths.

#### *The Development of the Dialectic Approach to Mixed Methods*

As a part of the paradigm wars, researchers were debating not just the validity of their own methodologies and techniques, but their superiority over their rivals. This led to a galvanization of thought and a questioning of researchers' very ability to authentically use methods from diverse paradigms together. Gretchen Rossman and Bruce Wilson were the first to articulate a typology of stances formalizing the differing perspectives that researchers could take towards conducting mixed method research.

#### *Gretchen Rossman and Bruce Wilson*

The three stances Rossman and Wilson (1985) developed are differentiated in two respects: by the ability to use mixed methods (yes or no) and by the extent to which the methods are integrated in the study or evaluation (from little or none to fuller integration).

The first stance is *purist*, which is rooted in paradigmatic concerns. Researchers taking a purist stance would argue that methods from different paradigms could not be mixed because the epistemological and ontological assumptions that undergird them are irreconcilable. Excepting the practical concerns of limited funding and insufficient researcher training, the most consistent arguments against mixed methods have been philosophical. For example, Dan Bednarz (1985) highlighted the post-positivist, social interactionism and ethnomethodology traditions. He argued they are orthogonal to one

another and cannot be mixed, in part, because of their differing conceptions of quantitative / qualitative, causation, reliability and validity, which he defines. Combining paradigms, he argued, only results in critical aspects of one method's findings being ignored, or one being misinterpreted or reinterpreted by either the historically dominant paradigm (argued to be post-positivism) or the researcher's own paradigm. Norman Denzin (1978) also questioned the researcher's ability to triangulate across theory.

Bednarz' and Denzin's concerns are important; Greene and McClintock appear to have designed their study on triangulation in response to them. Indeed, can researchers successfully step outside of their paradigm if they are well trained and intentional? In 1985 Greene and McClintock published an important study in triangulated mixed method design. The work is foundational because its design addressed concerns of purist researchers such as Bednarz (1985) and implemented more closely Campbell and Fiske's conception of triangulation. Additionally, they advanced their interest in exploring paradigms, a component of mixed method research not previously addressed outside of philosophic argument.

The authors sought to answer the question, can the paradigms that guide methods be integrated at analysis or must one be subsidiary to another? Each co-author implemented a one-method evaluation of the same aspect of an adult and community education organization. The resulting mixed method evaluation was a concurrent and independently administered paper-and-pencil survey designed from a post-positivist perspective, and an unstructured interview designed from an interpretivist perspective.

Descriptive results were reported to be complementary across survey and interview findings. However, major findings differed in substance and form because the qualitative results were primarily descriptive. When descriptive results were mapped onto major findings, the survey illustrated a linear one-on-one relationship between the two, while the interview illustrated a web-like relationship. The authors argued that the difference in relationships indicates the different analytic strategies necessary for each method. At the point of recommendations, however, the authors found the methods complementary. The authors concluded that mixed methods provides “structure, substance, and strength” to findings (p. 540). They also argued that in the end, one method must be subsidiary to another, and therefore advocate a purist position against cross-paradigm triangulation.

The strength of this article is it advances the paradigm debate with the rigor of independently conducted evaluations, addressing concerns about researcher bias (Campbell & Fiske, 1959; Denzin, 1978; and Bednarz, 1985). However, the authors may have concluded differently: that because the findings from the two methods were complementary at the point of recommendations, and because the two data sources enriched each other, cross-paradigm triangulation was successful. This might reflect a pragmatic stance toward mixed methods. Instead, Greene and McClintock appear to have sought strict equality between methods when this expectation may not be appropriate. For example, Louise Kidder and Michelle Fine (1987) pointed out that qualitative researchers’ research questions often evolve as the study progresses, which poses a key challenge in triangulating methods. Shotland and Mark (1987) suggested



that use of multiple methods may affect the level of generalization of inference, particularly if methods address slightly different questions. Indeed, I would argue that a criterion of strict equality is not sufficient for dismissing the potential advantages of cross paradigm mixed method inquiry.

The last two of Rossman and Wilson's (1985) stances allow a mixed method study to be conducted, but integrate the diverse methods at different levels. The second stance is *situationalist*, which focuses on methods and the situations in which they may best be used. Researchers employing this stance would mix methods, and analyze and interpret the results side-by-side with little or no integration. The final stance is *pragmatist*, in which a researcher focuses on the relative strengths of differing methods and seeks to integrate them at analysis and interpretation.

*Jennifer Greene and Valerie Caracelli*

In a 1997 volume of *New Directions in Evaluation* dedicated to mixed methods, Greene and Caracelli present a reformulated trio of stances to reflect their interest in incorporating paradigms more intentionally into mixed methods practice. The first stance is *purist*, and mirrors Rossman and Wilson's definition. The second is *pragmatic*, a revision of Rossman and Wilson's conceptualization. For Greene and Caracelli, a researcher employing a pragmatic stance would welcome the opportunity to choose the method to best address the research question at hand. The pragmatist recognizes the epistemological differences of paradigms, but "prioritizes situational responsiveness and an empirical perspective" (p. 9). Patton (1985) argued that pragmatism – and a focus on evaluation clients – allows researchers to draw from the

full complement of research methods and paradigms available. Drawing on his experience as an evaluator, Patton argued that evaluators can make mindshifts back and forth between paradigms. He also deemphasized the importance of paradigms as he describes how clients from different perspectives can agree on an evaluation design and find the results convincing and useful. He urged evaluators to listen to their clients' worldviews and center them on an empirical perspective.

Datta (1997) and Maxcy (2003) articulated a pragmatic stance to mixing methods that has its roots in the philosophic position of John Dewey and William James (among others). Datta (1997) outlined the essential criteria for making design decisions as (1) *practicality*, which implies one's experience and knowledge of what does and does not work; (2) *contextual responsiveness* to the demands, opportunities and constraints to an evaluation situation; and (3) *consequentiality*, or making decisions based on practical consequences. She suggested that consequentiality be judged by answers to the following questions: can salient evaluation questions be answered?; can the design be successfully carried out?; are design tradeoffs optimized?; and are the results usable?.

In contrast to the arguments of Bednarz (1985), the pragmatic stance appears disarmingly practical – to the possible detriment of methodological principles – and somewhat inelegant. However, it has wide appeal. John Creswell, Vicki Plano Clark, Michelle Gutman, and William Hanson (2003) reported that it appears to be the dominant paradigm employed by mixed method researchers. Leslie Riggin (1997) found a pragmatic stance to be almost exclusively employed when she reviewed all

examples of mixed method evaluations presented in a volume of *New Directions in Evaluation* dedicated to the subject.

Greene and Caracelli's final stance is the *dialectic*, in which the researcher would prioritize paradigmatic and methodological differences and seek to integrate them in a dialogic manner. The authors described the dialogically developed research results as reflecting the object of study, as well as being descriptive of the paradigms employed. In a dialectic stance the researcher would use both sources of information from each method to directly inform the other in a spiraling manner. Greene and Caracelli (1997) argued that the purpose of the dialectical stance is to gain insight. Leslie Riggin (1997) pointed out that the dialectic stance assumes that methods act as carriers for unique information about paradigms. I would add the stance also assumes methods act as carriers for unique information about findings due to paradigms.

The dialectic stance is intriguing, but it is also very complicated. A researcher would examine the results of a method and consider how the paradigm impacted the results. Then he or she would analyze how the results and paradigmatic concerns interact with a second or third method – each with their own results and potentially different paradigm.

In the same journal volume in which Greene and Caracelli presented their stances, Melvin Mark, Irwin Feller and Scott Button (1997) presented an extensive evaluation and attempt to categorize their stance as purist, pragmatic and dialectic. They described several components of their evaluation as potentially dialectic because the mix did not result in new insight, an implicit criticism of the conceptualization of

the stance. In actuality, the authors misinterpreted the dialectic stance because the dialog between methods described in the stance is the responsibility of the authors, not the vagaries of the data or situation. In another test of the stances, Riggin (1997) reviewed every example presented in the same volume of *New Directions in Evaluation* edited by Greene and Caracelli, and attempted to categorize them by stance. She found every example to be pragmatic.

In her 2007 book on mixed methods, Greene asked the question:

So how can paradigms be simultaneously respected, honored, and understood, on the one hand, as vitally important anchors for inquiry findings and their warrants, and on the other, as open, dynamic, and inviting of dialogue with other perspectives and stances? (p. 53).

Greene argued this can be done by broadening the philosophic paradigm to the more grounded, intuitive notion of mental model which are not rigid and incommensurate as paradigms can be. Conflicts may occur between mental models, but those conflicts can be generative because mental models seek conversation and understanding. She also guided researchers to avoid questions of incommensurable attributes such as objectivism-subjectivism and realism-idealism, and to direct themselves to distinct attributes on a continuum, such as distance-closeness and generality-particularity, etc. In sum, Greene turned from paradigms to mental models in order to support a position that warrants mixing methods of diverse paradigms.

Greene revised her stances to include six stances towards mixing methods and paradigms or mental models. The six stances reflect the nature of a philosophic

paradigm for social inquiry and the role paradigms play in social research<sup>1</sup>. In her revision of the dialectic stance, paradigms are seen as importantly different, but because they are socially constructed, their mixing is not inviolable. Methods are guided by paradigms and context and “important paradigm differences should be respectfully and intentionally used together to engage meaningfully with differences and, through the tensions created by juxtaposing different paradigms, to achieve dialectical discovery of enhanced, reframed, or new understandings” (p. 69, 2007). While Greene’s (2007) focus on intentionally implementing methods within explicit paradigms is strongly promoted, the evidence suggests the dialectic stance requires further clarification to be useful in the field (Mark, Feller, & Button, 1997; Riggin, 1997).

Greene also updated her conceptualization of the pragmatic stance in her 2007 book, by re-categorizing it as one of three possible alternative paradigm stances (pragmatism, scientific realism and the transformative or emancipatory paradigm). Within alternative paradigm stances, paradigms may be mixed because an alternative paradigm is employed that welcomes or even requires mixed methods, but that rejects the notion of incommensurability. This alternate paradigm offers its own internal

---

<sup>1</sup> The first stance is purist and reiterates that mixing paradigms/models is impossible because the assumptions of the paradigms are irreconcilable, and paradigms guide all inquiry decisions. The second stance is a-paradigmatic and states that paradigms are logically independent from each other so can be mixed based on the practical characteristics and demands of the inquiry context and problem at hand, not paradigms. The third stance is substantive theory, where the substantive issues and conceptual theories relevant to the study matter most in guiding practical inquiry. The fourth stance is complementary strengths, where paradigms are viewed as different in many ways, but not fundamentally incompatible. Inquiry is guided by paradigm as well as by inquiry context. An important goal is to maintain paradigmatic and methodological integrity.

coherence and integrity, instead. Greene viewed pragmatism as appealing in a mixed method context because it embraces scientific inquiry and social construction and its practical, consequential character.

In the late 1990s, the dialectic stance towards mixed methods represented the best thinking in the field for how to conceptualize and approach the use of two methods from different research traditions. However, since the publication of the book *The Handbook of Mixed Methods in Social and Behavior Research* (Tashakkori & Teddlie., 2003), the pragmatic approach first described by Rossman and Wilson (1985) has gained popularity.

#### *An Emerging Pragmatic Approach to Mixed Method Studies*

In the last five years, mixed method scholars have begun exploring a pragmatic approach to mixed method studies with concerted interest. Johnson, Onwuegbuzie and Turner (2007), Tashakkori and Tedlie (2003), and Morgan (2007) have argued that pragmatism represents the single most appropriate approach to mixed method studies. Pragmatism's core contributions to mixed method research are two-fold. First, it provides a rationale for combining the methods from diverse paradigms, undercutting the incommensurability thesis. According to Johnson & Onwuegbuzie (2004), pragmatism finds a middle ground between paradigmatic incommensurability and presents a workable solution for combining diverse paradigms. Second, pragmatism promotes the use of research approaches to best answer important research questions. Pragmatist mixed methods research would use empirical and practical consequences to judge the merit and worth of combining methods. Pragmatism suggests that researchers

adopt a needs-based or contingency approach to selecting methods and approaches (Johnson & Onwuegbuzie, 2004).

Despite the efforts of many mixed method scholars to promote a pragmatic perspective, the summary of pragmatism that they provide glosses over important differences between the three major American pragmatists, Charles Pierce, William James and John Dewey. This lack of specificity casts doubt on the suitability of pragmatism within mixed methods. Greene (2007) pointed out this problem, and posited that John Dewey's thinking is best suited to mixed methods. Based on this she argued that pragmatism does not set aside paradigms embedded within methods but attends to "transactions and interactions; to the consequential, contextual and dynamic nature of character of knowledge; to knowledge as action; to the intertwinement of values with inquiry; and so forth" (p. 85). She also questioned how precisely these philosophical commitments get enacted within mixed methods research and that more scholarship is necessary in this area.

#### *Dialectic and Pragmatic Approaches as Generative Endeavors*

Both pragmatic and dialectic theorists and practitioners see an important benefit to mixing methods. Greene (2007) described that mixed method studies may be generative, as paradox and contradiction are engaged and "fresh insights, new perspectives and original understandings" emerge (p. 103). This is especially suited to a dialectic approach that prioritizes the consideration of paradigms. At the same time, proponents of a pragmatic stance use the term *gestalt* to indicate the how inferences

from mixed methods may be greater than the single method components (Tashakkori & Teddlie, 2003).

To examine the claim that mixed method studies produce a unique yield beyond single method studies conducted independently, Alicia O’Cathain, Elizabeth Murphy and John Nicholl (2007) analyzed mixed method studies in the field of health services research through research funded by the United Kingdom Department of Health and identified on the Department’s research website. The authors developed two indicators of yield, the level of integration between methods and publications of mixed method findings. Of 81 studies examined, they found that 21% integrated the findings, and an additional 28% integrated findings but that more was possible. However, evidence of method mixing was not commonly seen in publications. An important area for future research is the extent to which mixed methods produce unique insights, although this endeavor would not be easy in practice because the transparency necessary in reporting how insights emerged from studies.

### Applying Pragmatism to Mixed Methods

As described in previous sections, pragmatism is an emerging approach to the conduct of mixed method evaluation and research. However, a structured application of pragmatism to mixed methods requires further definition. This section briefly reviews the history of pragmatism in order to propose a method for operationalizing a pragmatic strategy to mixed findings from two paradigmatically diverse methods.



Pragmatism is a uniquely American philosophical tradition rooted in the Academic Sceptics who studied in Plato's Academy from the third to the early first century B.C.E. This group of philosophers believed that authentic understanding and knowledge is impossible due to humans' epistemic limitations. Instead, plausible information and understanding must suffice and judgment must be suspended (Rescher, 2001). Eighteen centuries later, Immanuel Kant's scholarship provided a further basis for pragmatism. Kant believed that humans are fundamentally limited in their ability to achieve understanding in totality, which he considered a theoretical endeavor. Instead, humans must function with understanding that is merely sufficient or practical in nature. Kant argued that the need to prioritize the practical over the theoretical is a part of the human condition (Rescher, 2001).

One hundred years after Kant, pragmatism was most fully developed by Charles Sanders Pierce, William James and John Dewey. In their scholarship, pragmatism is primarily concerned with meaning or epistemology as measured by its consequences. Modern pragmatist Nicholas Rescher (2001) described that in pragmatism, what works in practice becomes the standard for the truth of assertions, the rightness of actions and value of appraisals. Gathering evidence is a method of estimating the truth of a thing; and if sufficiently warranted, the evidence stands in for the truth. An implication of pragmatism is that the best we can do to gather evidence is always good enough. Despite the challenge of the inquiry, one's best evidence gathering is never insufficient to determine truth.

Despite the popularity of pragmatism, the field has been somewhat fragmented. Pierce coined the term “pragmatism” and later renamed it “pragmaticism” to differentiate his work from that of William James (Rescher, 2001). Pierce’s approach was deeply influenced by his work as an experimental scientist in cosmology. His influence was wide-ranging and includes innovations in the areas of probability, induction, and logic. From this background, he described himself as a “laboratory philosopher” and believed that the scientific method be adopted for inquiry. For Pierce, this scientific approach assumed that real things exist beyond our opinions of them and that those things stimulate our senses according to laws. By observation, we cannot see real things, but may ascertain how those things really are (Hookway, 2008). Therefore, Pierce understood truth to be an opinion that withstands scrutiny and evidence (Brent, 2001). He sought to generate meaning (opinions) by impersonal and objective standards of what is observed. To do so, one would list predictions for a thing’s character and test if they are true. Once additional evidence comes to light to refute a truth, rigorous methods of inquiry should be applied to refine it (Hookway, 2005).

William James altered and popularized Pierce’s conception of pragmatism. James expanded the evidence necessary to indicate truth. Not only is truth proven by objective observation of predicted behaviors, but also by the affective benefits an object produces for idiosyncratic individuals. For example, an idea might be proven by its ability to encourage a valuable emotion or behavior. For James, an assertion is true if it works satisfactorily in the widest sense of the word (Blackburn, 1996). James’ most famous application of pragmatism is in his discussion of god. James argued that

god exists because he provides comfort and hope to believers (Walker, 2008). James was deeply influenced by his own bouts of deep depression and personal faith, as well as by his training as a psychologist. His popularity was fueled by his colorful and populist presentation of the “cash value” of truth and that all truths share the quality “that they pay” (Thayer, 1968). Additionally, James’ work substantially influenced Husserlian phenomenology. James believed that conceptual thought only reflects reality in a sliver of its fullness. Instead, one must plunge into an experience in all its flux and contradictions, and provide an account of that world in progress. Even through this account is likely to contain some contradictions, lived experience is more likely to truly capture the essence of reality than Pierce’s coolly scientific method (Sprigge, 2005).

John Dewey adopted Pierce’s scientific approach to pragmatism, but acknowledged the importance of social context, as highlighted in James’ writing. For Dewey, truth was learned through a self-corrective process of experimentation where new evidence required a revision of established conclusions. However, Dewey understood this scientifically-oriented process to take place within and be influenced by a social milieu of community-based moral and psychic values. Dewey’s emphasis on community is consistent with his focus on advancing democracy through reform in civic process and education (Hanson, 2005). As described by Greene (2007), Dewey contended that the meaning of human experience is found in the transaction or relation of meaning or knowledge between the objective reality and the internal mind of the knower, in each of which knowledge and meaning resides. Truth, then, may be found

in the consequences of the interaction, so truth is contextual and related to action, which is constituted by truth, meaning and knowledge.

Pragmatism has been widely criticized, most thoroughly by European philosophers who have disregarded pragmatism as exemplifying a crass concern with success over truth, a naïve democratism, and a “go getter” spirit overly invested in the practical. Specifically, F. H. Bradley argued that the practical interests at the heart of pragmatism are not exhaustive, so should not dictate truth. Similarly, G. E. Moore pointed out that utility changes over time, so is an inaccurate measure of truth. Most famously, Bertrand Russell argued that beliefs can be both useful and plainly false (Rescher, 2001).

Regardless of the criticisms, pragmatism has been very influential in America, and embodies a specific stance toward mixed methods. Mixed methods researchers such as Tashakkori, Johnson and Morgan support the adoption of a pragmatic approach in the field. Brief reviews of pragmatism are provided by the authors but they concede that the philosophy of science behind their use have yet to be developed. Mixed method scholar Riggin (1997) reviewed a group of mixed method studies and determined them all to be pragmatic in nature. However, Riggin’s scholarship on mixed methods seems to equate a pragmatic approach to a common-sense one. Indeed, “common-sense” is frequently used as a shorthand for a pragmatic approach among many mixed method scholars.

One marked departure from this trend is Datta (1997) who outlined the essential criteria for making mixed method decisions. The criteria include (1) practicality, which

implies one's experience and knowledge of what does and does not work, (2) contextual responsiveness to the demands, opportunities and constraints of an evaluation situation, and (3) consequentiality, or making decisions based on practical consequences. Datta suggested that consequentiality be judged to answer questions such as: can salient evaluation questions be answered?, can the design be successfully carried out?, are the design trade-offs optimized?, and are the results reliable? Interestingly, Datta defined that consequentiality be used as a criterion for mixed methods exclusively in the planning stages of the evaluation. However, practicality (in terms of knowledge and experience) and contextual responsiveness may be employed during planning as well as analysis.

### Applying Dialectics to Mixed Methods

As described in previous sections, Greene & Caracelli developed the dialectic stance towards mixed methods, in which a researcher mixes results from diverse paradigms. The purpose of the stance is to gain insight by integrating the paradigmatic and methodological differences in a generative and spiraling manner (Greene & Caracelli, 1997). However, I was unable to find examples in the literature of how a dialectic stance might be implemented. This section briefly reviews the history of dialectics in Western culture in order to operationalize the construct preliminarily. Next, I propose a method for operationalizing a dialectic strategy to mix findings from two paradigmatically diverse methods.

Dialectic is a term derived from Greek meaning to converse or discuss. The term originates from Aristotle who attributes it Greek philosopher Zeno of Eliatic, but is explicated most thoroughly by Plato (Howatson & Chilvers, 2003). Both Aristotle and Plato used the dialectic to illuminate the Greek concept of *elenchus*, a method of cross-examination for the purpose of enlightenment (Robinson, 1953).

Plato used the dialectic in two ways. First is through his Socratic dialogs depicting Socrates' methods. Here, the dialectic is a form of reasoning and argument designed to uncover truth. The dialectic process is one of question and answer, like ordinary conversation, where the philosopher seeks to uncover what is known. She tests arguments to expose contradictions, rendering a position as defective (Hare, 1983). In Greek, this helpless position is termed as *aporia*, whose benefit is the knowledge of ignorance. A criticism of this form of dialectic is that the method is destructive as opposed to constructive (Howatson & Chilvers, 2003). In *Theaetetus*, for example, Plato presented Socrates as a barren midwife, who draws out the ideas of others but generates none himself. This form of the dialectic survives today in the fields of logic and analytic philosophy, where inconsistencies prove falsehood (Sedley, 2004).

Plato expanded the meaning of dialectic in his middle dialogs such as *The Republic* and *Phaedrus*, where Socrates remains the chief figure, but begins to expound on positive ideas. Plato described the dialectic as one way to reach the knowledge of the supreme good (the form of the Good), which describes not only a method of philosophy but the goal of philosophy itself (Roberts, 2007).

In the Middle Ages, dialectics was considered to be one of the seven liberal arts. Specifically, it is one of three “arts of language” that comprise the *trivium*. In practice, dialectics was synonymous with logic during the Middle Ages (Jolivet, 2001).

In the 18<sup>th</sup> Century, Immanuel Kant used the concept of dialectic to illustrate that principals of science have diametrical aspects, but Georg Friedrich Hegel’s treatment of the dialectic is much more comprehensive. Hegel used the dialectic as a way to understand history and freedom. For Hegel, dialectics is a process by which human history unfolds. History is a critical force on human development, shaping our nature and choices. As stated in *Lectures of the Philosophy of History*, “History of the world is none other than the progress of consciousness towards freedom” (as cited in Singer, 1983). For Hegel, an important dimension of freedom is the ability to take control of the forces of history (Singer, 1983).

Hegel illustrated this by tracing human history from despotically led ancient empires whose subjects, he believed, did not understand their capacity to judge right from wrong. Hegel traced the drive of human freedom through history, sparked by Greek civilization, and reaching a pinnacle at the Reformation, where each person could achieve their own salvation. Hegel sought to advance history beyond the Reformation by positing the possibility for the existence of a rational community (Beiser, 1993). This community of pure idealism married Kant’s purely rational morality of duty versus personal interest with Hegel’s own belief in the importance of community-specific ethical customs. Hegel’s lengthy historical exegesis in *History*

illustrated at length the principal that the course of history is driven by the need to transform the world (Singer, 1983).

This transformation is termed in *The Science of Logic* and is translated as “sublation” or “overcoming,” and is the key concept that redefines the dialectic from its Socratic roots. Like Socrates as depicted by Plato, Hegel’s dialectic is concerned with contradictions. However, for Hegel, truth is not discovered by identifying falsehood through contradiction. Instead, a position is challenged by an argument, and the two points are united by a third that transcends and subsumes both. This transcendent concept then becomes subject to challenge, until the final transformation is perfected (Singer, 2005).

Philosopher Heinrich Mortiz Chalybause interpreted Hegel’s dialectic as having three parts: thesis, antithesis and synthesis, although Hegel never used these terms. Instead, they were developed by Hegel’s colleague Johann Gottlieb Fichte (Beiser, 1993). In an historical example, ancient Greek customary morality is the thesis and the Reformation is the antithesis. Hegel’s conception of the rational community is the synthesis. In more abstract form, dialectics may be seen in Hegel’s discussion of Being and Nothing that is described in *The Science of Logic*. Hegel argued that pure Being has no content, so cannot be anything. Therefore, being is both being and its antithesis, Nothing. However, these opposites must be brought together, just as one’s living is also one’s dying. This synthesis is Becoming, a transformation that encompasses both Being and Nothing and overcomes it with greater truth.



In his day, Hegel was criticized for selling himself to the Prussian Monarchy. Hegel's contemporary, Author Schopenhauer decried Hegel in *Philosophy of Right* for paralleling Hegel's ideal of a rational community of pure idealism to the Prussian Monarchy, Hegel's employer. Karl Popper also sharply criticized Hegel as the precursor to the modern totalitarian state. Popper argued that Hegel gave dictators' license to "force" their subjects to freedom through fascism. Preeminent Hegel scholar Peter Singer, however, argued that this claim misreads Hegel's support for constitutional monarchy, the rule of law, trial by jury, and freedom of expression very generous for his day. However, Singer did fault Hegel for a perhaps overly optimistic belief in humans' ability to overcome conflict and build a rational and harmonious community (Singer, 2001).

Karl Marx and Friedrich Engels are the last scholars to substantially contribute to the concept of dialectics. Both Marx and Engels studied Hegel and were deeply influenced by his work. In fact, their treatment of Hegel's dialectic fundamentally supports the notion of thesis, antithesis and synthesis. However, in *Capital*, Marx described Hegel as "standing on his head" (as cited in Calhoun, 2000) and Marx announced his quest to right him. Whereas Hegel viewed the thinking process as creating the real and physical world, Marx argued the opposite – that the material world creates thought. For Marx, the material base determined the world of the mind. Based on this distinction with Hegel, Marx develops dialectical materialism in *Capital*. For Marx and Engels, the contradiction inherent in class struggle, between mental and

manual labor and between town and country are critical to their central theories of surplus value and their materialist conception of history (Calhoun, 2000).

In summary, a brief review of dialectics in Western culture reveals three main conceptualizations: (1) Plato's dialectic where contradictions in arguments identify falsehood; (2) Hegel's dialectic where a thesis faces an antithesis and is transformed by a synthesis that encompasses both positions; and (3) Marx and Engel's dialectical materialism that interprets thesis, antithesis and synthesis in the material world. Which of these conceptions may be most productively applied to a dialectic strategy to mixing methods?

This question is examined by considering the ability of each approach to handle a mixed method study where findings from phenomenological interviews and post-positivist surveys contradict one another. First, consider Plato's dialectic, which would attempt to argue that the findings from one method are logically superior to the other method. This is plausible if one method was well administered according to its own standards of methodological quality while the second method fell short of its standards for high-quality implementation. In this case, it is possible to argue that the findings from one method are superior to another. However, in the case where both methods are conducted with an acceptable level of quality, it is unclear to me how to determine which method is logically superior to another. One might attempt to consider the two divergent findings of the study and determine the most logical of the two. However, both sets of evidence were derived from carefully conducted studies. The premise of mixed method literature is that diverse paradigms provide valid information for

research and decision-making. The theory of mixed methods provides no support for the decision to determine that one method or set of findings is logically superior to another.

If the researcher honors the integrity of both methods, it is possible to conclude that the contradictory findings of both methods are wrong. This mirrors the non-generative quality of Plato's dialectic, where ideas are proven false but no alternative is presented. Greene & Caracelli defined their dialectical approach as one that is generative in nature. Therefore, it appears that using Plato's conception of dialectics would result in either a biased result or one that opposes' Greene and Caracelli's vision of the approach.

Next, consider Hegel's approach to the dialectic. Two conflicting findings would be examined side by side, and a synthesis that encompasses but transcends them would be sought. Unlike that of Plato, Hegel's approach allows for a generative conclusion when the findings from two different methods diverge from each other. While Plato's use of the dialectic is focused on proving arguments false through contradictions, Hegel's dialectics seeks to generate new truths that transcend the old.

Hegel's approach also meets a second of Greene and Caracelli's criteria, that methods be combined in a spiraling manner. The spiraling is manifest because the synthesis created could itself turn into a thesis, which may then be challenged by another antithesis, until the final synthesis is perfected. This may be especially important in mixed methods because as new syntheses are generated, they may conflict with one another and require resolution.

One issue in applying Hegel's method of thesis, antithesis and synthesis is that the assignment of thesis and antithesis is arbitrary. Just as it is problematic to define one method's finding as logically superior in Plato's conception of the dialectic, it is problematic to assign one method's finding as a thesis or antithesis for any non-arbitrary reason. However, Hegel himself did not use the terms thesis and antithesis and many of his examples of dialectic do not demand that one concept be placed first. For example, in his dialectic exploration of Being and Nothing, either term could have been used first. Therefore, an inability to purposively assign one finding to thesis and the other to antithesis does not appear problematic.

Finally, consider Marx's dialectical materialism, which affirms Hegel's use of thesis, antithesis and synthesis, but grounds the work of the mind in a material basis. It is not obvious to me how Marx's variation of Hegel's dialectic improves on Hegel's own conceptions as applied to mixed methods. It is possible that the prioritization of the material world is less compelling when considering non-material information or evidence from studies of diverse paradigms.

In sum, Hegel's conception of the dialectic appears to be best suited for use in mixed methods studies. The approach is generative, as desired by Greene and Caracelli (1997) and findings may be combined in a spiraling manner. Marx fundamentally supports Hegel's use of thesis, antithesis and synthesis but his focus on materialism does not obviously improve on Hegel's conceptions. On the other hand, because Plato's brand of dialectics focuses on identifying falsehood through contradiction, the

approach is fundamentally non-generative. Therefore, I will use Hegel's conception of the dialectic to guide the development of the mixed method studies in this paper.

### Validity

Validity is a construct developed to assess the truth value of inferences made from study findings. The validity of mixed method findings and the single methods upon which they are based is a critical component of rigorous mixed method research. Unfortunately, the literature on validity in mixed method studies is nascent, likely because of the formidable challenge it represents. Mixed methods can comprise single methods guided by paradigms whose epistemological, ontological and axiological assumptions diverge wildly from each other. Gareth Morgan (1983) has convincingly argued that the criteria for judging the quality of a research method is derived from the paradigm that undergirds that method. Therefore, using quantitative criteria to judge qualitative inquiry has been argued to be inappropriate (Hammersly, 1992), to result in assessments that lack meaning and that are awkward and confusing (Leininger, 1994). In short, the quality of the inferences from a study should be judged by the terms of the paradigm in which the study is situated. None-the-less, in the context of mixed methods studies, it would be helpful to judge validity using standards that may be translated among and between methods of diverse paradigms.

The concept of validity was developed within the positivist tradition and a rich literature honoring its complexity has emerged. Likewise, a concern for validity is held with equal seriousness by practitioners of the interpretive tradition. In the interpretive

literature, the appropriate criteria for the validity or trustworthiness of inferences made from study findings is still contended. This is especially true as phenomenological, feminist, grounded theory, and ethnographic methodologies, among others, have claimed their own unique paradigms with corresponding validity criteria (Whittenmore, Chase, & Mandle 2001; Patton, 2002). However, a gold standard appears to have emerged for an interpretive methodology broadly considered (Whittenmore, Chase, & Mandle, 2001; Patton, 2002), which holds promise for evaluating validity within a mixed method framework.

#### *Trustworthiness: A Unified Framework*

Drawing on concepts of validity via research design (Campbell & Stanley, 1966; Cook & Campbell, 1976), Lincoln and Guba conceptualized validity for constructivist or naturalistic inquiry and termed it broadly as “trustworthiness” (1985). For Lincoln and Guba, trustworthiness of findings involved answering the question, “How can the inquirer persuade his or her audiences (including self) that the findings of an inquiry are worth paying attention to, worth accounting of? What arguments can be mounted, what criteria invoked, what questions asked, that would be persuasive on this issue?” (p. 290, 1985). To answer this question, Lincoln and Guba (1985) identified four major concepts of validity within a positivist tradition whose fundamental concern applies to both positivist and interpretive inquiry and reframes them for use in interpretive methodologies. These four concepts include “truth value,” applicability, consistency and neutrality. Using these concepts as a template, the

authors specify the criteria that may be best used to judge them within interpretive inquiry.

The next four sections describe each of Lincoln and Guba's (1985) abstracted concepts, the positivistic validity criteria they are based on, and the interpretive criteria that they develop.

### *"Truth Value"*

According to Lincoln and Guba (1985), "truth value" refers to how to establish confidence in the "truth" of a study's finding for its subjects, given the particular context in which the study took place. Lincoln and Guba's concept of "truth value" has its roots in internal validity as explicated by Cook and Campbell (1979), or the extent to which a relationship between two variables is caused by the intervention or by other factors. The authors discussed Cook and Campbell's eight criteria. Positivist methodologists Pedhazur and Schmelkin (1991) reformulated them as follows: (1) history, or events that took place during the study that may effect the results; (2) maturation, or changes due to the passage of time; (3) testing, or differences in test performance due to multiple exposures to a test; (4) instrumentation, which refers to changes in the instrument or its administration that would effect outcomes; (5) regression to the mean, which indicates regression or reversion in scores that occurs when the correlation of two variables is imperfect, and which is most crucial when subjects are chosen because of extreme qualities or characteristics; (6) selection, or difficulty in attributing differences between groups to the intervention and not participant characteristics of those groups; (7) mortality, or the self-selected factors of

those drop out of a study; (8) diffusion or imitation of treatments, which refers to the extent to which participants learn of a study condition they are not in and participate in their unassigned treatment to a degree; and (9) compensatory rivalry or resentful demoralization, where participants react to the fact they are in a different treatment in which they wish to be, which may result in higher or lower behaviors.

The threats to internal validity listed above are not exhaustive, but is representative of the major issues presented in the discussion of internal validity. Within positivist inquiry, the use of tight study controls in the research design are the best protection against threats to validity. Namely, random assignment to treatment and comparison groups and the manipulation of independent variables are key design considerations (Pedhazur & Schmelkin, 1991).

The positivist conceptualization of internal validity is merely a starting point for Lincoln and Guba, however. While they described positivist efforts to establish a one-to-one relationship between inquiry findings and reality via control groups and manipulated independent variables, these goals and strategies do not apply to interpretive efforts. Instead, interpretivists believe in multiple, constructed realities, and seek “truth” for each individual subject encountered. Therefore, the interpretive analog to internal validity is credibility via participant experience.

Lincoln and Guba outlined several methods that would increase the probability that credible findings would be produced. These methods included prolonged engagement or persistent observation with participants, triangulation of data sources, efforts to build trust with participants, peer debriefing after data gathering, negative



case analysis, referential adequacy or documentation of data, and member checks, or presenting findings to participants for confirmation and revision of understanding.

### *Applicability*

Lincoln and Guba (1985) described applicability as the extent to which findings from one study are relevant to other contexts or subjects. This concern was drawn from the positivist conceptualization of external validity. Like applicability, external validity refers to an ability to generalize findings to or across populations, locations, settings, times, etc. Generalizing to a population refers to the ability to extrapolate findings to the population from which the sample was drawn. Random sampling with negligible attrition best accomplishes this goal. Generalizing across findings concerns the ability to generalize to populations, locations or times outside of the population to which the sample is presumably representative is more problematic. Cook and Campbell suggested that a number of smaller studies with haphazard samples may contribute more to external validity (Pedhazur & Schmelkin, 1991).

Bracht and Glass (1968) contributed substantially to Cook & Campbell's conception of internal validity by developing the distinction between ecological validity and population validity. Ecological validity refers to the extent to which a study's methods, materials, and setting approximate the real-life situation under investigation. However, ecological validity is not always necessary for generalization. For example, a study involving a mock jury trial would not meet the criteria for ecological validity because the trial is controlled and pre-determined, not real, and may take place in a non-courthouse setting. However, it is still possible that this study could

be generalized to a real trial none-the-less (Shadish, Cook & Campbell, 2002).

Population validity, in part, represents the ability to generalize to other populations, taking into consideration interactions between the treatment and other factors. Threats to external validity include (1) treatment-attributes interaction, or the extent to which subjects with varying levels of an attribute respond differently to the treatment; (2) treatment-settings interaction, or the condition in which participants in different settings (i.e., a laboratory versus in the field) respond differently to the treatment; (3) multiple treatment interference, which refers to subjects for whom treatments administered at the same time may interact with each other, and to subjects for whom treatments are administered sequentially but may have crossover effects; (4) pre-test sensitization, where completing a pre-test may sensitize a subject to the phenomenon of study and influence his or her responses; and (5) post-test sensitization, during which a post-test measure may sensitize subjects to a latent or incomplete effect that the treatment itself would not have elicited. Testing and accounting for these interactions enhances external validity.

While positivist researchers often seek to generalize across populations, interpretive researchers do not believe that it is desirable or possible to generalize beyond situations with similar contexts. Therefore, Lincoln and Guba described that interpretive researchers are concerned with the transferability of findings to other contexts. To achieve transferability, thick description of the context of the study is necessary so that the researcher can determine if a transfer to another context is necessary. Likewise, thick description of the “receiving” context is also necessary.

Thick description may be accompanied by a database that provides information about the minimum elements necessary to make a decision about transfer.

### *Consistency*

For Lincoln and Guba (1985), exploring consistency determined the extent to which findings from a study would be repeated if the study were replicated with the same or similar subjects, or in the same or similar context. The concept of consistency is an analog to reliability in the positivist tradition. Although reliability is discussed separately from validity, reliability is a necessary but insufficient condition of valid inferences from a positivist or post-positivist study. This is seen, for example in criteria-related validation. The underlying rationale of criteria-related validation is that the truth value of study inferences may be supported if the study findings are statistically associated with an external criterion that is a standard for the phenomenon of interest. The reliability of both the test measure in question and criterion measurements are positively related to criterion-related validity (Sax & Newton, 1997; Pedhazur & Schmelkin, 1991).

Reliability is derived from classical test theory and indicates true score variance divided by variance observed in the measure in question. Reliability assesses the stability of tests over time and their equivalence to similar test forms. Strategies for assessing reliability include test-retest reliability, where the same test is administered to the same group of people two times and the results are correlated with one another. Increasing the time interval between administrations reduces test-wisness, but allows subjects to change on the phenomenon being measured. Another strategy is to develop

parallel or equivalent forms of a test that measure the same phenomenon and administer them within several days of each other. Finally, the internal consistency or homogeneity of a test may also indicate reliability. Using split-half reliability, the items on a test would be split and correlated with one another. Whether assessed through test-retest, equivalent forms or internal consistency procedures, reliability is expressed as coefficient alpha and represents the true score variance divided by observed score variance.

Obviously, the testing procedures associated with reliability are a poor match for interpretive methods, which rely on the investigator as the primary instrument in a study. From this perspective, Lincoln and Guba argued that instrumental unreliability may be a concern, where the instrument is the investigator conducting the study. Humans may become careless, or bend to fatigue and make mistakes. However, variation due to the emergent designs of interpretive methods is a fundamental characteristic of interpretive research and does not indicate unreliability that would threaten the inferences of a study. However, monitoring the study context and emerging designs for instability in findings that may negatively impact the inferences made from study findings is an important strategy in assessing consistency.

According to Lincoln and Guba (1985), consistency may be assessed through examining dependability. In turn, dependability may be assessed by overlap of methods (triangulation); stepwise replication, where teams of researchers analyze the same data; and via audits that examine both the inquiry process and accuracy of findings.

*Neutrality*

Lincoln and Guba (1985) described neutrality as “the degree to which findings of an inquiry are determined by the subjects (respondents) and conditions of the inquiry and not by the biases, motivations, interests, or perspectives of the inquirer” (p. 290). This conception of neutrality overlaps with the positivist conception of objectivity, but differs in several respects. In a positivist tradition, objectivity implies that one reality or truth exists to be discovered, and that this is best accomplished through a strong distance between the observer and the observed. Additionally, the positivist tradition holds that it is possible for researcher to take measurements without impacting findings by values.

Conversely, in an interpretive tradition, neutrality reflects the quality of the data that eschews the influence of bias or opinion. In this way, neutrality is not a property of the investigator (as discussed in the discussion of consistency above), but is instead a property of the data. For Lincoln and Guba (1985), neutrality is tested by the question, are the data confirmable? The authors specify the development of an audit trail and an audit process to test confirmability. The audit trail involves providing raw data; data reduction and analysis products; data reconstruction and synthesis products, such as categories, findings and a final report; process notes; materials relating to intentions and dispositions; and instrument development information. The audit process follows five stages that include pre-entry, determination of auditability, formal agreement, determination of trustworthiness, and closure.

### *Summary of Trustworthiness*

Taken together, Lincoln and Guba (1985) examined the rich literature of validity in the positivist tradition. Despite the inability to replicate criteria to the interpretive paradigm, the authors glean the conceptual kernels of positivist validity to develop parallel criteria for interpretive sciences. These criteria included specific tools and procedures that may be employed to determine the extent to which they are met. Table 1 below lists the conceptual content of validity identified, the positivist validity criteria and attendant tests, and the interpretivist criteria and attendant tests.

### *Criticisms of Trustworthiness*

Lincoln and Guba's conception of trustworthiness can be criticized by the standards of both positivist and interpretive traditions. From the positivist perspective, trustworthiness fails to account for several key concepts of test validity. Within the field of educational psychology, test validity has historically been understood to comprise three inter-related components: criterion-related validity, construct validity, and content-related validity. More recently, Samuel Messick (1989, 1995) and Michael Kane (1992) have advocated a unified theory of validity. The historical and reformed notions of validity are each discussed in turn.

*Criteria Related Validity.* The underlying rationale of criteria-related validation is that the truth value of study inferences may be supported if the study findings are statistically associated with an external criterion that is a standard for the phenomenon of interest. In a predictive validity test, the test measurement is associated with a

Table 1

*Validity and Trustworthiness for Post-positivist and Interpretive Paradigms*

Trust-worthiness Concept	Positivist Analog		Interpretive Analog	
	Concept	Test / Method	Concept	Test / Method
“Truth Value”	Internal validity	Research design – randomized control groups and manipulation of independent variables	Credibility	Prolonged engagement or persistent observation, triangulation, building trust, peer debriefing, negative case analysis, referential adequacy, and member checks.
Applicability	External validity	Random sampling and testing for interactions between the treatment and key factors	Transferability	Thick description of context and dataset of minimum elements needed to assess transferability
Consistency	Reliability	Coefficient alpha via test-retest, equivalent forms or split-half reliability techniques	Dependability	Overlap of methods, stepwise replication, and inquiry audit
Neutrality	Objectivity	Distance between observer and observed, value-free measurement	Confirmability	Audit trail and audit process

predictor collected at a later time than the test. Conversely, in concurrent validation procedures, the test and criterion measurement are collected at the same time.

*Content Related Validity.* Content-related validity refers to the extent to which an instrument measures the domains of a specific content area. Test invalidity is demonstrated by a test that fails to fully assess all relevant content, or that includes items that measure constructs unrelated to the phenomenon of interest. Content-related validity may be assessed by gauging agreement among subject matter experts on how essential a particular item is (Lawshe, 1975). This is achieved through calculating a Content Validity Ratio (CVR) for each item in an instrument that ranges in value from -1.00 to 1.00. A  $CRV > 0.00$  would indicate that more than half of the subject matter experts endorse an item as essential, which Lawshe viewed as establishing content validity. A second quantitative approach to content validity was developed by Jacob Cohen (1960). Cohen's methodology asks subject matter experts to sort measurement items into pre-defined and mutually exclusive measurement scales representing different constructs. Cohen's *kappa* assesses the degree of inter-expert agreement on the placement of items into their measurement scales. Like Lawshe's CVR, Cohen's *kappa* may range from 1.00 to -1.00, where *kappa*  $> 0.00$  represent observed agreement beyond chance.

It is important to note that outside of the fields of Education and Educational Psychology, some researchers do not consider content-related validity a measure of validity at all because content-related validity is not seen to affect the inferences made



from test scores, but is instead considered to assess the content of the instrument (Pedhazur & Schmelkin, 1991). Conversely, educational psychologists argue that in measuring achievement especially, test content is critical to the inferences that may be made from test scores.

*Construct-Related Validity.* Construct-related validity represents a third component of validity that assesses the extent to which test scores measure or account for some unobserved, theoretical construct on the basis of observed variables that are presumed to indicate the construct. Sax and Newton (1997) outlined at least six steps by which construct validation may occur. These steps include explicitly justifying the construct; distinguishing the hypothesized construct from those that may be similar; ensuring the measurability of the hypothesized construct; demonstrating convergent validity by correlating the measure in question with other independent measures of the same or similar trait; demonstrating discriminant validity by correlating the test measure with irrelevant factors; and modifying the nature of the construct based on the evidence gathered.

Within the positivist tradition, scholar Samuel Messick (1989; 1995) and Michael Kane (1992) have postulated a unified theory of validity that centers on construct validity and informs the quality of interpretations. Messick argued that all three types are really one because they are all necessary for validation (Sax & Newton, 1997). Instead, he defined six aspects of construct validity: content-relevancy (extent to which the domain is relevant, representative and socially desirable), substantive (theoretical rationales for test responses and confirmation participants are engaged by

them as measured by think-alouds, for example), structural (evaluation of scoring criteria), generalizability (extent to which score properties and interpretations generalize across the construct domain), external (correlations with test scores that are expected to have positive and negative relationships) and consequential (consider score bias, fairness and social consequence of test scores).

For Messick (1995), “almost any kind of information about a test can contribute to an understanding of test score meaning” (p. 725). The task of the investigator is to interpret information and develop an argument about the validity of interpretations of findings. The clarity of the interpretive argument, its coherence and the plausibility of assumptions that under gird it are criteria for its evaluation (Kane, 1992).

While positivist researchers may criticize Lincoln and Guba’s conceptualization of trustworthiness for failing to account for the specifics of test validity or the comprehensiveness of Messick and Kane’s argument-based unified theory of construct validity, researchers from interpretive traditions have even bigger fish to fry. First, Lincoln and Guba (1986) themselves argued that the full range of criteria that may be used to determine the trustworthiness of conclusions drawn from interpretive studies is truncated by using positivist criteria as a template. As described in Whittemore, Chase, and Mandle (2001), Hammersley (1992) argued that criteria based on positivist conceptualizations are inappropriate. Leininger (1994) found such efforts awkward, confounding and confusing. Kahn (1993) viewed the translation of positivist criteria as a procedural charade.

Of equal importance, a wide variety of interpretive evaluators from different traditions (ethnographic, grounded theory, phenomenological, feminist, Marxist, connoisseurship, and artistically-driven constructivist, for example) base their criticism of Lincoln and Guba's conception of trustworthiness on Lincoln and Guba's (1985) very contention that studies of diverse paradigms should be judged by their own criteria. These evaluators have all claimed their own paradigms and right to determine the criteria that best judge their own quality. Therefore, Whitemore, Chase, and Mandle (2001) argues that validity criteria must be sufficiently flexible to support interpretive methods that are defined by their fluidity, emergence and uncertainty. Whitemore and colleagues proposed primary criteria of credibility, authenticity, integrity and criticality, as well as secondary criteria of explicitness, vividness, creativity, thoughtfulness and congruence. Practitioners of diverse methods would place emphasis on the criteria that best reflects the paradigm of their inquiry. Likewise, Patton (2002) matched criteria for judging the quality and credibility of interpretive inquiry by approach or paradigm. For example, he proposed constructivist evaluations be judged on criteria such as praxis and particularity, while artistic and evocative paradigms be judged by aesthetic quality, stimulating, provocative, and critical change evaluations (such as feminist, queer theory and Marxist) be evaluated by clear historical and values context and consequential validity.

#### *Trustworthiness: A Final Assessment*

Despite the trenchant criticism of trustworthiness from both positivist and interpretive quarters, Lincoln and Guba's conceptualization remains the gold standard

(Whittemore, Chase, & Mandle, 2001; Patton, 2002). It is especially well-suited to mixed method studies because validity arguments from studies guided by diverse paradigms may be examined together. Additionally, Lincoln and Guba's criteria are sufficiently universal in interpretive inquiry that they may be used to judge the trustworthiness of inferences from interpretive paradigms as diverse as phenomenology, constructivism and grounded theory. However, it is important to note that critical change evaluations (feminist, queer theory, Marxist, etc.) would be severely underserved by Lincoln and Guba's conception of neutrality. Critical change evaluations prioritize the interests of disadvantaged groups and would be penalized for their advocacy of those groups under Lincoln and Guba's criteria, when that advocacy is an integral part of their paradigm. Therefore, mixed method designs that include critical change or advocacy research should not employ Lincoln and Guba's conceptualization of neutrality. Alternate criteria as described by Whittenmore, Chase, and Mandle (2001) and Patton (2002) would be more appropriate, despite the fact that the criteria may have no match with trustworthiness or validity criteria from other paradigms in a mixed method study.

### *The Standards for Program Evaluation*

In the context of mixed methods evaluation, it may be useful to consider alternative criteria by which the quality and rigor of mixed method studies may be judged. Patton (2002) suggested *The Program Evaluation Standards* (Joint Committee, 1994) as a criterion. Because the three single methods and two mixed method studies utilized in this dissertation were conducted as policy evaluations for a non-profit

organization, the *Standards for Program Evaluation* are relevant to the studies explored in this dissertation.

Patton (2002) described the development of the standards from the inception of the field of evaluation in the social programs developed by the Great Society in the 1960's. In the 1970's evaluation was solidifying as a field of professional practice and findings were judged by criteria of post-positivist research that was favored at that time. However, large sophisticated evaluations failed to inform policy (Patton, 2002; Chelimsky, 1997; Greene, 2007). Funders and evaluators themselves began to ask the question, who is evaluating the evaluators?

Headed by noted evaluator Daniel Stufflebeam, the first set of standards for program evaluation were developed over a period of five years by representatives from 12 partner organizations and input from hundreds of evaluators (Joint Committee on Standards for Educational Evaluation, 1981). Entitled *Standards for Evaluations of Educational Programs, Projects, and Materials*, the volume selected the criteria of utility, feasibility, propriety and accuracy. As described by Stufflebeam, the order of the criteria was intentional. An evaluation without a prospect for utility is ill-fated. Likewise, without practicality an evaluation is inappropriate. The same argument was made for ethics and accuracy, in that order (Joint Committee on Standards for Educational Evaluation, 1981). In 1994, Standards were revised, resulting in updates to the four key criteria and 30 sub-standards that specify them. Currently, the Standards are under review again. Draft standards are currently being assessed in field trials and hearings.

The Standards are the most widely used criteria for evaluation quality within the field. However, in 1995, the American Evaluation Association developed a set of five guiding principals for evaluators. These principals include systematic inquiry, evaluator competence, integrity / honesty, respect for people and responsibilities for general and public welfare. These principals are often used jointly with the Standards.

## CHAPTER THREE

### RESEARCH METHODS

This study investigates mixed methods in order to provide evidence to assist evaluators and applied researchers to optimize their designs of mixed method research and evaluation. This chapter describes the research methods employed in the study in four sections. First, an overview of the methodology will be described. Second, a timeline for the secondary data is presented. Third, the overarching methodology and rationale for this dissertation study is discussed. Last, the specific research methods utilized to answer research questions 1 through 5 are presented.

#### Overview of Methodology

This study is a secondary analysis of a mixed method evaluation conducted for ClearWay Minnesota, a non-profit organization dedicated to enhancing life for all Minnesotans by reducing tobacco use and exposure to secondhand smoke through research, action and collaboration. As one part of its mission, ClearWay Minnesota funds a variety of QUITPLAN<sup>SM</sup> programs to assist Minnesotans to stop using tobacco.

The purpose of the evaluation conducted for ClearWay Minnesota was to determine the extent to which local smoking regulations impacted QUITPLAN program participants in their efforts to stop smoking. The evaluation consisted of three single method sub-studies: an 18-month follow-up telephone survey with comparison group, a series of focus groups, and a series of phenomenological interviews. A total of

nine (9) staff from Professional Data Analysts, Inc., (PDA) the evaluation firm contracted to conduct the evaluation, contributed to conducting the evaluation. I was the Principal Investigator for the evaluations; as such, I designed, managed and contributed to collecting data for all three of these studies in order to: (1) to conduct an evaluation for ClearWay Minnesota, and (2) to conduct a dissertation regarding mixed methods.

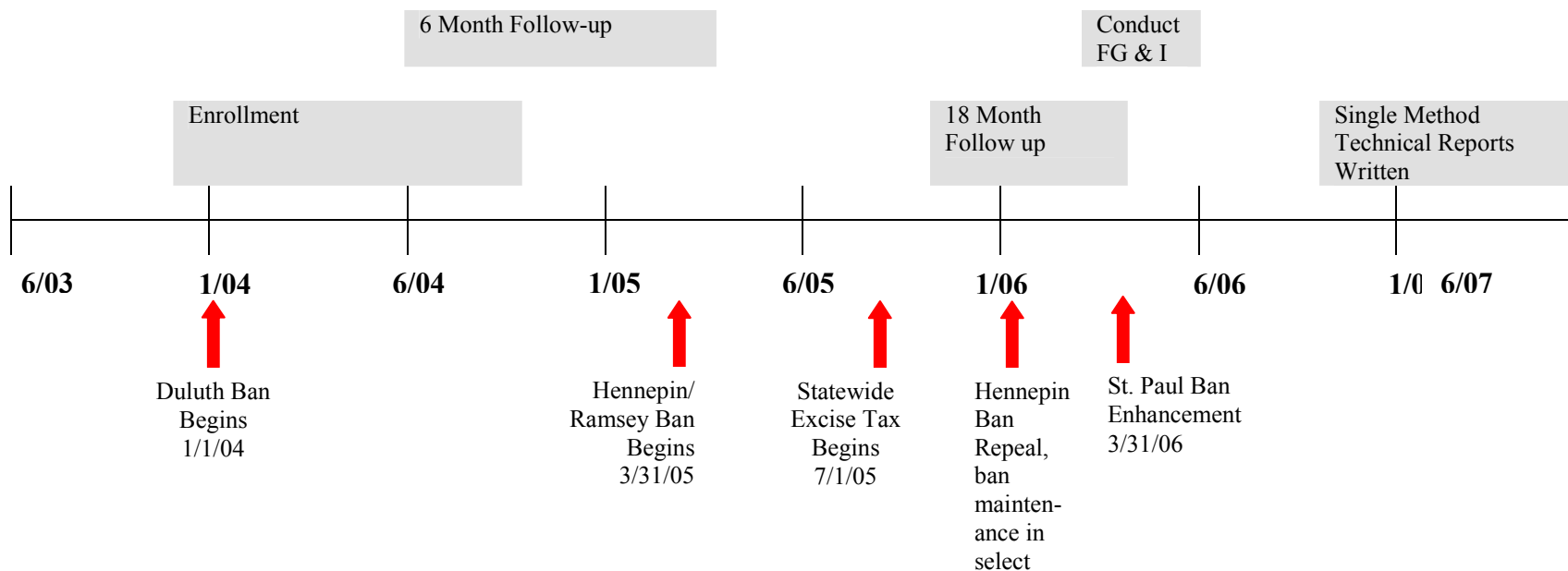
The purpose of this dissertation study is to address five research questions, the results of which are designed to provide evaluators and applied researchers with empirical evidence so that they may optimize their mixed method designs. Content analysis will be used to compare the substance of single method findings to each other in order to identify convergence, divergence and uniqueness of findings. In order to create pragmatic and dialectic mixed method studies, dialectic and pragmatic mixing processes will be developed and utilized. The substantive findings from the two mixed method studies will be compared to each other and the substance of single methods to determine convergence, divergence and uniqueness of findings. Additionally, validity concerns will be addressed. Single and mixed method studies were be assessed according to a framework for validity that is inclusive of post-positivist and interpretive methods. Mixed method findings will also be assessed. Criteria include standards of validity from the literature, an assessment of *The Program Evaluation Standards* (Joint Committee, 1994), and expert review of the single methods. Finally the cost to conduct each single study in billable research dollars and subject hours will be computed and analyzed using Quickbooks accounting software.



### Timeline for Secondary Data

The evaluation data collection period spans from December 2003 to May 2006 and involves three data collection methods: a telephone survey that includes data collection at intake, 6-months post-enrollment and 18-months post enrollment; focus groups and interviews. The time periods for each of these is presented in Figure 4 below. To fulfill the evaluation contract, a technical report on each of the methods was written from October 2006 to May 2007.

Figure 4 below also illustrates key dates of the intervention for the evaluation, local smoke-free ordinances. Such ordinances may be considered “full” if they prohibit smoking in all indoor public places or “partial” if they allow exceptions for facilities such as bars, bowling alleys, pool halls, etc. On January 1, 2004, the city of Duluth enacted a partial ordinance. Over a year later, on March 31, 2005, Hennepin County enacted a full smoke-free ordinance and Ramsey County enacted a partial ordinance. Over the next year, both Hennepin County and the City of St. Paul changed the strength of their ordinance. On January 1, 2006, Hennepin County repealed its full ordinance and instituted a partial one. However, the full-strength ordinances enacted by the cities of Minneapolis, Golden Valley and Bloomington superseded the Hennepin County repeal and these three Hennepin County cities remained under a full smoking ban. The city of St. Paul strengthened its ordinance from partial to full on March 31, 2006. Finally, the State of Minnesota enacted a \$.75 excise tax on a pack of cigarettes effective July 1, 2005.



Dates for Evaluation Data Collection and Reporting

Enrollment	12/03 – 10/04
6 Month Follow up	6/04 – 4/05
18 Month Follow-up	10/05 – 4/06
Single Method Technical Reports	10/06 – 5/07

Figure 4. Timeline for Key Intervention Dates, Evaluation Data Collection, and Reporting

## Research Approach and Rationale

Paradigms are central to the conduct of this dissertation; the telephone survey was conducted using a post-positivist paradigm, while the focus groups were weakly interpretive and the phenomenological interviews were more strongly interpretive. Because paradigms are a central concern to this study, it is pertinent to explore what paradigms will be employed in the secondary data analysis that constitutes this study's methods.

A primary method that will be utilized is content analysis of documents, which may be either post-positivist or interpretive in its administration. As Guba and Lincoln (1981) pointed out, content analysis has historically been considered a quantitative endeavor as consensus developed between researchers Bernard Berelson (1952), F. Earle Barcus (1959), and Ole Holsti (1969), among others. They defined content analysis as a rule-guided and systematic process that aims for generality by focusing on the manifest as opposed to latent content of document, particularly at the coding stage. However, such content analysis is only weakly post-positivist, and incorporates judgments and qualitative coding that are descriptive of interpretive measures. Therefore, if the content analysis is considered interpretive, it is only weakly so. Taken together, the content analysis for this dissertation reflects a paradigm that hovers in the center of the continuum, neither strongly post-positive nor interpretive in nature.

In addition to content analysis of documents, this dissertation will compute the cost of single methods in researcher billable dollars and subject time. Each of these is a

quantitative method and post-positivist in approach. Finally, this dissertation employs a paper and pencil stakeholder survey, the content of which was based on the theory under girding the study. The survey includes both closed-ended and open-ended response opportunities and falls on the post-positivist divide of the paradigm continuum.

While methods tend to be weakly post-positive, I hold an interpretive approach to synthesizing findings and drawing conclusions overall. I have chosen this approach because mixed methods and interpretivism complement each other. While an interpretive approach explicitly allows for the discovery and discussion of multiple truths, the theory underlying mixed methods is that diverse data collection methods and approaches more fully capture the latent truth(s) of a phenomenon, and provide better evidence for decision making. In the synthesis of results from this dissertation, I will search for the most meaningful evidence that explicates the multiple truths that I believe to underlie a more optimized practice of mixed methods.

### Research Questions

The research methods used to answer each of the following research questions is discussed in the next five sections.

1. What are the substantive findings of single methods? What findings converge and diverge? What findings are unique?

2. What are the substantive findings of pragmatic versus dialectic mixed method studies? How are the two mixed method study findings similar and different from one another? What unique information do the mixed method findings produce over and above single methods?
3. To what extent do inferences drawn from single method findings meet key criteria for validity / trustworthiness? To what extent are inferences drawn from single method and mixed method findings valid / trustworthy according to *The Program Evaluation Standards* (Joint Committee, 1994)?
4. How do stakeholders view the credibility and utility of single method findings and mixed method findings and why? What do they see as the advantages and disadvantages of mixing? What are their prior beliefs about the credibility of diverse methods and paradigms?
5. What are the costs of the single methods in terms of researcher and subject hours?

This section outlines the subject, apparatus and procedure used to answer each research question, in order from first to last.

*Research Question 1: Comparison of Single Method Results to One Another*

The first research question concerns the substantive finding of the single method studies and the extent to which the findings converge, diverge or are unique. The object of study for this research question will be the full technical reports written for each single method of the evaluation. The content analysis method will be practiced

through a moderate post-positivist lens. The content analysis will be rule-guided in order to reduce subjectivity and increase the replicability of the study.

First, topic areas will be identified. Topic areas reflect the major content domains for which data was collected, for example, the impact of ordinances on relapse, travel to communities with different ordinance conditions, social norms, etc. Next, within each content area, findings will be extracted. Findings are defined as information units bounded by substantive content about a specified topic area. More than one finding may populate a topic area, and a finding may include more than one related idea. Findings may also include conclusions from the single study findings.

Findings for each topic area will be compared across single methods, and will be categorized as being convergent, divergent, divergent by degree or unique. A convergent finding is the same finding in a particular topic area across two or more single method studies. A divergent finding indicates that the substantive finding in a particular topic area differs between two or more single method studies. A finding that is divergent by degree is a finding that contains both convergent and divergent elements across two or more studies. A unique finding is a finding in a topic area not addressed in other reports. These categories follow criteria for category development (Holsti, 1969). They reflect the purpose of the research, are exhaustive, mutually exclusive, independent and reflect a single classification principal or level of analysis. The coding categories also reflect the nature of the data, which is narrative and content oriented in terms of findings.

I will read each single study report and create one list of study findings for each. The findings will be presented in a two column table. Column headings will include Topic Area and Finding. After all three single method study findings have been tabled, the tables will be merged and sorted by Topic Area. Findings will be condensed as appropriate. A two column table will be constructed that describes the finding and the relationship between methods for that finding. Findings will be organized by topic area. Findings will be coded a “C” for convergent findings, a “D” for divergent findings, a “U” for unique findings, and “Dd” for findings that diverge by degree. The studies that exhibit the relationship will also be specified. “S” refers to telephone survey; “FG” refers to focus group, and “I” refers to phenomenological interviews. The table will be examined for patterns of convergence, divergence and uniqueness of findings within each study, across single studies and by topic areas. Conclusions about convergence, divergence and unique findings across studies will be drawn.

*Research Question 2: Comparison of Mixed*

*Method Results to Each Other and Single Methods*

In the analysis of Research Question 1 above, the results from single methods were examined side-by-side. However, no integration of findings occurred, and no method for mixing at the stage of implementation (e.g., pragmatic or dialectic) was applied. The side-by-side comparison of single methods provides the basis for more sophisticated mixed method studies. Research Question 2 concerns the development of pragmatic and dialectic mixed method studies, and the extent to which their results converge, diverge or are unique from each other and from single methods.

This research question will be answered in three steps. First, the dialectic and pragmatic mixed method studies must be developed. The dialectic study will mix the telephone survey and phenomenological interviews. These two methods best represent the dialectic mix because Greene and Caracelli (1997) stipulated that paradigms should be prioritized in the dialectic approach. One way of prioritizing paradigms during design is to select two methods that differ the most from each other paradigmatically. The post-positivistic telephone survey and strongly interpretive phenomenological interviews represent this arrangement. Additionally, an unstated assumption of a dialectic approach to mixed methods is that differing paradigms may increase the variance between methods, thus increasing the utility of findings and the validity of inferences drawn from them. Thus, the “distance” between paradigms of diverse methods may be critical to mixed method studies. The combination of the post-positivistic telephone survey and strongly interpretive phenomenological interviews would best test this unstated assumption.

The dialectic study involves mixing at three points. First, during the implementation of the phenomenological interviews, mixing occurred during sampling: participant characteristics gathered by the survey were used to select the sample for the interviews. Second, the data on participant characteristics gathered during the telephone survey were also used to analyze the interview data; therefore, the methods have been mixed during analysis. However, the primary stage at which mixing will occur in the dialectic study is during the interpretation of findings. During all stages of



research, all single methods were weighted equally in terms of the researcher attention and importance to the study.

To mix the telephone survey and phenomenological interviews at the point of interpretation, I will compare the single method tables of findings for the telephone survey and the phenomenological interviews. Findings will be categorized as being convergent, divergent, divergent by degree or unique. However, the literature on mixed methods provides little instruction on how to conduct the intellectual exercise of mixing findings at the point of interpretation in a specifically dialectic manner. I turn to my literature review on dialectics for guidance on this subject. In sum, Hegel's conception of the dialectic appears to be best suited for use in mixed methods studies. The approach is generative, as desired by Greene and Caracelli (1997) and findings may be combined in a spiraling manner. Therefore, I will use Hegel's conception of the dialectic to guide the development of the mixed method studies in this paper. In a dialectic approach, divergent findings will be examined using the format outlined in Table 2 below. It is important to note that all syntheses should be compared to one another. If a contradiction between them exists, they should be subjected to the same form of Hegel's approach. After the dialectic mixed method study is developed, a narrative will be written and findings from the study will be tabled.

Table 2

*Format to Apply Hegel's Dialectic Approach to Divergent Mixed Method Findings*

---

Topic

Method 1

Conflicting Finding

Method 2

Conflicting Finding

Synthesis

Rationale

Likewise, a pragmatic mixed method study will be developed. To do so, the single method tables of findings for the telephone survey and the focus groups will be compared. The combination of the telephone surveys and the focus group best represent the pragmatic mixing for several reasons. First, the telephone survey and the focus groups contribute both quantitative (i.e., numbers) and qualitative (i.e., words) data to the method mix. Second, the focus groups have several practical advantages over the phenomenological interviews in terms of choosing which qualitative method to include in the pragmatic mix. For example, focus groups benefit from the ability to ask open-ended questions and hear participants' responses in their own words. At the same time, however, they include many more people than the phenomenological interview ( $N = 70$  vs.  $N = 15$ ), so the inferences from focus groups would be more credible than phenomenological interviews to those who value post-positivist criteria for generalizability. Therefore, focus groups can appear a more practical choice

because they would have greater appeal to audiences that value both post-positivist and interpretive methods. In contrast, the appeal of phenomenological interviews is limited only to those who value strongly interpretive methods. Additionally, the focus groups have the practical benefit of having a more straight forward approach and philosophical foundation than the phenomenological interviews. Taken together, the evidence suggests that the combination of the telephone surveys with the focus groups makes an ideal pragmatic pairing.

Like the dialectic mixed method study, the pragmatic study will mix the two single studies primarily at the point of interpretation, however mixing did also occur at the points of sampling and analysis. The weight of each method was equal at all stages of the mixed method research design.

In order to mix the telephone survey and interviews at the point of interpretation, findings will be categorized as being convergent, divergent, divergent by degree or unique. To ensure that the mixing at interpretation is pragmatic in approach, the framework developed by Datta (1997) will be employed. I have operationalized the framework in Table 3 below. After the pragmatic mixed method study is developed, a narrative will be written and findings from the study will be tabled.

Table 3

*Format to Apply Pragmatic Criteria to Pragmatic Single Method Findings*

---

Topic

Method 1 findings

Method 2 findings

Experience-based  
considerationsKnowledge-based  
considerationsContextually responsive  
considerations

Conclusion

Second, each mixed method study will be compared to each single method study in order to identify the unique information, if any, that is produced by each mixed method. This will be done by comparing the tables of findings already produced.

Finally, the two mixed method studies will be compared to each other. The tabled findings from each mixed method study will be merged, and sorted by topic area. The findings will be examined and coded as convergent (“C”), divergent (“D”), unique (“U”) and divergent by degree (“Dd”). The relationship of each finding to the two mixed method studies will be represented in the first column of the table, and the finding will be summarized in the second column. Finally, the table will be examined for patterns of convergence, divergence and uniqueness of findings across mixed

method studies, and conclusions about convergence, divergence and unique findings across studies will be drawn.

*Research Question 3: Examination of Validity of Inferences for Studies*

The third research question is designed to discover the extent to which single and mixed method studies meet key criteria for validity. This section will describe the research methods that will be employed to explore validity with regard to single method findings first and mixed method findings second.

The extent to which single methods meet key criteria of validity will be assessed in three ways. First, methodological experts in each of the three single methods will review the methodology in order to provide evidence of content-related validity of the methods. In the case of this study, the domain in question is sufficiently well constructed follow-up survey, focus group and phenomenological interview methods. Expert comments will be compared to the actual practice of each single method to determine the extent to which the domain of methodological quality for each method is met.

Second, quantitative and qualitative methods will be compared to key standards of validity appropriate to each. Sax and Newton described validity as “the extent to which measurements are useful in making decisions and providing explanations relevant to a given purpose” (p. 304, 1997). For example, the 18 month follow-up telephone survey will be compared to statistical standards of validity as described in Howell (2002), including (1) internal validity, or the extent to which causal statements may be attributed to the intervention; (2) external validity, or the extent to which

findings may be generalized to the population of interest; (3) reliability, or the extent to which scores are consistent, stable and/or equivalent; and (4) objectivity, the extent to which findings are unbiased by the researcher.

Likewise, the focus group and interview methods will be compared to qualitative standards of validity developed by Lincoln and Guba (1985). These standards include (1) credibility based on prolonged engagement, persistent observation, triangulation, etc.; (2) transferability, or the ability to assess the context of the study in order to judge its applicability to other situations; and (3) dependability or confirmability, the extent to which observers may see the process by which the researcher achieved his or her findings, and results in a confirmability judgment. Based on the comparison of the method to the standards, an argument for the validity of methods will be developed.

Next, the single study methods will be compared to standards of evaluation, namely, the Joint Committee's (1994) *Program Evaluation Standards*. An argument for the fidelity of the single methods to evaluation standards will be developed. Similarly, the two mixed method studies will be compared to the same standards of evaluation and an argument for the fidelity of mixed methods to evaluation standards will be made. Finally, the arguments for fidelity to standards for validity across methods will be synthesized and conclusions will be drawn.

#### *Research Question 4: Stakeholder Views of Credibility and Utility*

Research question number four concerns how stakeholders view the credibility and utility of single method findings and mixed method findings (dialectic and

pragmatic). The subjects of the investigation are the primary stakeholders for the evaluation of smoke-free ordinances. They include the Senior Research Program Manager responsible for managing the evaluation grant; the Policy Program Manager who participated in evaluation meetings to provide input on study designs; and the Senior Marketing Manager who requested analyses to inform marketing efforts and attended a presentation of findings.

Data on stakeholder views of credibility and utility of findings was gathered via a five page paper and pencil survey. The survey covers three broad content areas: respondent experience with the three single methods used in the evaluation (telephone surveys, focus groups and phenomenological interviews); respondent opinion on the credibility and utility of single method findings of these methods; and respondent report on the credibility and utility of two mixed method summary of findings: one mixed according to pragmatic principals and including surveys and focus groups, and one mixed according to dialectic principals including surveys and interviews.

Two theoretical frameworks guided the development of the survey. First, Lincoln and Guba's (1985) theory of validity that unifies post-positive and interpretive methods informed how criteria for validity were operationalized. Second, items on utility were developed according to Karen Kirkhart's (2000) notion of the utility of both evaluation processes and findings to stakeholders.

The survey was piloted on three evaluators at PDA who attended ordinance study evaluation grant meetings. Revisions were made based on pilot findings. A copy of the final survey can be found in Appendix A.

*Research Question 5: Cost of Single Methods*

The fifth research question concerns the cost of each single method in terms of billable dollars and subject hours. Therefore, the subjects for the investigation are the nine staff at Professional Data Analysts, Inc. (PDA) that conducted the evaluation and the subjects who completed surveys, interviews or focus groups.

The apparatus for conducting the analysis of billable dollars will be QuickBooks accounting software. Daily each PDA staff person will track the number of hours they worked on the evaluation and what tasks they conducted. The procedure for answering the research question will be to run queries of QuickBooks software on the evaluation project by staff and task. All entries for the ban studies will be categorized by staff, study (survey, interview and focus group) and task. Five main tasks emerged from a review of the detailed tasks descriptions, including: planning, implementation, database management, analysis and reporting.

The total cost for each method will be computed by multiplying each staff person's time by their billing rate and adding together the total billable researcher dollars and expenses. Adding together the cost of each study, the total cost for the evaluation will be computed. Findings will be tabled and two types of proportions will be computed. First, the proportion of the total evaluation cost of each single method will be computed. Second, for each single method, the proportion of costs conducted on each task will be computed.

Conducting the analysis of subject hours expended by study is more straightforward. For each method, the average time subjects take to complete a method



will be multiplied by the number of subjects who completed that method. The average time to complete a survey is tracked by the Computer Assisted Telephone Interview (CATI) software that is used to administer the survey. Interview and focus group time were tracked by the researcher. The total subject hours of the evaluation will be computed by adding together the subject time from each method. Next, the proportion of subject hours of each single method from the total evaluation will be computed.

## CHAPTER FOUR

### RESULTS

Research Question 1: What are the substantive findings of single methods?

What findings converge and diverge? What findings are unique?

Recall that findings are considered information units bounded by substantive content about a specified topic area, and may contain one or more ideas. Findings within topic areas were identified so that similar ideas between methods could be identified and convergence, divergence and uniqueness of findings across methods could be assessed. The purpose of this section is to understand the substantive findings of each method and the extent to which they converge, diverge and are unique to the method. In order to accomplish these goals, this section has two parts. First, findings from the single methods are summarized in narrative and tabular form. Next, findings are examined across methods for convergence, divergence and unique findings.

#### *Summary of Single Method Findings*

This section provides a brief narrative summary of the major findings from each of the three single methods. More detailed findings may be found in Appendixes B and C, in which key topics are identified and major findings within those topics are described.

*Post-Positivist Survey*

The overarching goal of the study was to understand the impact of smoke-free ordinances at the city and county level. The primary research questions this study explored were:

1. Among QUITPLAN participants attempting to quit tobacco, what impact does exposure to smoke-free ordinances have on 7-day abstinence, relapse and new quits at 18 months post-enrollment?
2. To what extent are QUITPLAN participants in ordinance and non-ordinance counties traveling to restaurants or bars in areas with a different ordinance status?
3. What is the impact of the cigarette excise tax on tobacco use according to self-report?

In order to understand potential survey response bias, survey responders were compared to non-responders on several key characteristics. Because having a strong history quitting tobacco is positively associated with subsequent quitting outcomes, the greatest area of concern was that response rates at 18 months would be higher for those abstinent at 6 months, as compared to those still smoking. The analysis showed that the concern was unfounded because response rates were similar at 18 months for those smoking versus abstinent at 6 months. Respondents and non-respondents were also compared on demographic and clinical characteristics, revealing expected patterns shown in previous literature. Respondents were more likely to be stable than non-

respondents (in terms of higher levels of education, employment, insurance coverage and marital status).

In order to determine the impact of exposure to smoke-free ordinances on quit outcomes (7-day point prevalence abstinence, new quits from 6 to 18 months, and relapse from 6 to 18 months), a logistic regression analysis was conducted. The primary independent variable of interest was *exposure to smoke-free ordinances* and several moderating variables were employed<sup>2</sup>. Exposure to a ban was marginally associated with relapse ( $p=.061$ ), and showed no significant relationship to 7-day abstinence or new quits. A bivariate examination of exposure to ban and relapse revealed a powerful relationship: those exposed to smoke-free ordinances were about two times less likely to relapse as compare to those who were not exposed to smoke-free ordinances ( $p= .020$ ). Exposure to a ban was not significantly associated with either 7-day abstinence or having made a new quit.

An important factor moderating the relationship between ordinance conditions and smoking status was frequency of travel between ordinance and non-ordinance areas. Because ordinances are local, it is possible that participants could live in an area covered by an ordinance, but travel regularly to bars or restaurants in non-ordinance communities. Therefore, travel is a major concern of this study. The majority of respondents did not report traveling regularly (74%), but smokers and non-smokers reported similar rates of regular travel. However, those in ordinance communities were

---

<sup>2</sup> Participant demographic and characteristic clusters adapted from previous analyses. Variables from 6-month follow-up include 7-day point prevalence abstinence, prolonged abstinence, utilization of service (high, medium, low), medications used and satisfaction with service. Variables from 18 month follow-up included other stop-smoking services used, QUITPLAN program enrolled in and motivation to quit.

more likely to travel regularly (32.4% versus 21.4%,  $p = .002$ ), possibly reflecting the confounding factor of geographic location. The vast majority of those in ordinance communities lived in Hennepin and Ramsey Counties and lived within close proximity of surrounding suburban counties with different (no) smoke-free regulations.

When asked why they travel, the greatest proportion of respondents reported traveling to bars and restaurants in communities with different smoking restrictions for reasons other than to go to establishments that allow or prohibit smoking (66.9%). A substantial minority did report traveling to bars or restaurants that allow smoking (39.3%). These respondents were more likely to be smokers (not 7-day abstinent at 18 months) and to live in ordinance communities ( $p < .001$ ). On the other hand, about ten percent of respondents reported traveling to establishments that prohibit smoking.

Respondents themselves reported that there was a small impact of smoke-free ordinances on their quitting, but conclusions were difficult to draw because the impact was self-reported. The associations between self-reported impacts and living in a community with a smoke-free ordinance were not consistent. Respondents reported a stronger impact of the cigarette tax. A substantial minority reported that the tax helped them think about quitting. For a comprehensive list of telephone survey findings by topic area, please see Appendix B, Table B1.

### *Focus Groups*

The overarching goal of the study was to understand the impact of smoke-free ordinances at the city and county level on people trying to quit smoking. The primary research questions this study explored were:

1. What impact do smoke-free ordinances have on QUITPLAN<sup>SM</sup> program participants' quit attempts and efforts to maintain their quits?
2. How have ordinances impacted QUITPLAN participants' travel to restaurants or bars in areas with a different ordinance status than where they live?
3. What is the impact of the cigarette excise tax on tobacco use?

Focus group findings revealed that smoke-free ordinances helped make quit attempts easier by removing triggers to smoke such as the smell of smoke and seeing groups of people smoking in bar and restaurant environments. Additionally, ordinances helped respondents smoke less, especially for those who frequently patronize restaurants, coffee shops or bars. For some, ordinances gave them one more reason not to smoke, in part because the ordinance reduces smokers' enjoyment in smoking. They disliked being separated from their friends and some felt like second-class citizens smoking outside the bar or restaurant. Finally, smokers and non-smokers reported that ordinances provided them with a substantially better experience in bars and restaurants. While some were tempted by the smell of smoke, many more found it repulsive even to the point of making some physically sick. Therefore, by at least partially removing

exposure to secondhand smoke, the ordinances improved respondents' bar and restaurant experience.

While many respondents' quit attempts and quality of life were supported by smoke-free ordinances, some reported being unaffected. A weak effect of ordinances is likely one reason for the lack of impact. However, determining the impact of smoke-free ordinances is complicated. The many factors that impede non-smokers' ability to report impacts suggest that impacts of smoke-free ordinances may be under-reported. For example, some smokers already adapted to not smoking near family / friends; others were proud of their willpower and had difficulty seeing the impact of factors outside of themselves. Some smokers' political dislike of government controls appeared to blunt their ability to see positive benefits from ordinances and seeing and articulating the positive impact of an absence of cigarette smoke appeared cognitively difficult for some smokers. Finally, ordinances can also make smokers feel defiant. Some smokers expressed resentment at being told what to do, and were angry that their rights were being impinged upon.

Travel to communities with a different ordinance condition than where one lives is an important consideration in assessing the relationship between ordinances and smoking behaviors. Overall, ordinances appeared about equally likely to have either no effect on travel to bars or restaurants in communities with smoke-free ordinances different than one's own, or to have an effect dependent on smoking status. The reasons for no impact of ordinance on travel were that respondents already adapted to and accommodated their non-smoking friends and family, and that going to a smoking

bar was not worth the inconvenience of the drive. As might be expected, some smokers traveled to smoking bars because it is a relaxing experience, and having to walk outside to smoke is inconvenient and can make them feel like a second-class citizen.

Additional evidence of the effect of smoke-free ordinances is that fact that some non-smokers avoided smoking establishments and traveled to non-smoking ones in order to avoid the smell of smoke.

For a minority of respondents, especially those with little disposable income, the \$.75 excise tax on cigarettes was a “big motivator to quit smoking”. Former smokers described thinking that the taxed price of cigarettes was “ridiculous” and thinking about their other financial needs, and where that extra money could go. The comparison of material needs for everyday life versus addiction to nicotine even helped some quit or stay quit. In response to the tax, many reported now using coupons for cigarettes, buying a cheaper brand, or traveling to neighboring low-tax states to buy them. However, most described that money was not an object for them and the tax had no impact on their smoking.

Two unanticipated topics about ordinances emerged: the impact of drinking and bars on ordinances and quitting, and social norms. Regarding bars, drinking was found to be a trigger to smoke. Bars are often seen as the place where drinking and smoking happens. Not surprisingly, bars are a common site for relapse. For some, the smoking ordinances aided quit attempts because they interrupted the association between drinking, smoking and socializing in a bar.



Regarding social norms, almost all respondents were aware that many others think that it is unacceptable to smoke and expose others to smoke. They experienced these social norms through media messages and the way friends, family or strangers looked down on them or treated them rudely if they smoke in public or at home. The social unacceptability of smoking appeared to be a powerful motivator for smokers to quit and consider quitting. While subtle and not immediate, social norms against smoking appeared to be the most promising mechanism to motivate new quit attempts, prevent relapse, and prevent individuals from starting to smoke in the first place. For a comprehensive list of focus group findings by topic area, please see Appendix B, Table B2.

### *Phenomenological Interviews*

The overarching goal of the study was to understand the impact of smoke-free ordinances. The primary research question this study explored was, *what is the lived experience of smoke-free ordinances for those who have enrolled in QUITPLAN quit smoking programs?* This question was designed to elicit first-person narrative about how Twin Cities residents who have tried to quit have experienced smoke-free ordinances.

For those trying to quit smoking, experiences of smoke-free regulations ranged from being difficult, challenging and rewarding, to having little impact, to being a relief and an aid to tobacco cessation and reduction. Smokers can have a particularly intense experience of smoke-free regulations because being prohibited from smoking forced them to face their cravings and addictions. Many experienced apprehension and

anxiety at being unable to smoke. They may experience an uncomfortable cognitive dissonance as they surmount the barriers of ordinances to smoke. For some, the inconvenience of the ordinances resulted in fresh insight on their addiction and a new motivation to quit. Others continued to struggle with their addiction without change or become angry at being regulated.

Required smoke-free environments relieved smokers and non-smokers alike from the smell of secondhand smoke. All smokers who discussed the smell of tobacco smoke hated it in the strongest terms, and almost all hated the smell of fresh tobacco smoke as well, describing it as “nasty,” “horrible,” “disgusting” and “god awful”. Like air, smoke travels freely and permeates everything around it. The smell of smoke attached to clothes and hair and lingers, an involuntary and disconcertingly intimate signal of being a smoker or spending time in a smoking environment. The smell can “out” a person as a smoker even if he or she would like to keep it a secret. The smell of secondhand smoke was associated with being dirty and with shame of smoking and addiction. Given the powerful associations of the smell of smoke, it is not surprising that smell was both a motivator to quit smoking, and a strategy to maintain a quit. Smoke-free ordinances greatly reduced the smoke from air, reducing the hated smell. Everyone who reacted negatively to the smell of cigarette smoke received relief, regardless if they noticed it or not. Non-smokers were equally adamant in their disgust for the smell of smoke as smokers.

Almost all participants experienced some impact of smoke-free regulations. In terms of ordinances in particular, the most common experience was as an aid to

tobacco cessation or reduction (47% of participants). In addition to helping smokers quit because they face the inconvenience of the ordinances and cognitive dissonance, some smokers liked smoke-free ordinances because they caused them to smoke less. The policies were an external support that helped them change their behaviors in ways they could not seem to accomplish on their own. Other former smokers benefited from the ordinance because their triggers to smoke were reduced. While smoke-free regulations clearly benefited both smokers and non-smokers, about one-third of participants reported that they have no impact, which appeared at least partially related to a lack of exposure to the ordinances themselves.

Many smokers were aware of the harm of secondhand smoke on others. While smokers may not protect themselves from the harm of cigarettes, they often protected those they love, especially children, by instituting personal bans in their homes or cars. Social expectations are powerful, and may contribute to another important experience of smoke-free regulations: shame. Smokers described not wanting family to know they smoke, keeping secrets, and hiding their smoking, all signals that a person may be feeling embarrassment or shame in smoking. Shame is marked by feeling seen “in a powerfully diminished sense” (Kaufman, 1985). The self feels exposed both to itself and others. Shame is a common component of addictions to alcohol, other drugs, gambling, etc. Shame is both a precursor to addiction, as addicts turn towards alcohol, drugs or other behaviors in order to cope with feelings of shame, as well as a consequence of it, when addicts feel shame for their addictive behaviors. Regulations may provide a new set of occasions in which smokers may experience shame for

smoking, and they may heighten the experience of shame by forming social expectations about smoking. However, just as smokers hid their smoking from family and friends before smoke-free regulations, they likely experienced shame in smoking before the regulations as well. Public relations and communications should be sensitive to the issue of shame for smokers regarding smoke-free ordinances.

While the interviews revealed that smoke-free ordinances can assist smokers make quit attempts and former smokers to stay quit, the most encouraging trend for impact on smokers and non-smokers is the possibility of enduring changes via social norms. Smoke-free regulations seemed likely to contribute to a belief that one deserves to breathe clean air and that exposure to secondhand smoke is unacceptable in public spaces. The discrete quit attempts caused by newly instituted smoke-free ordinances may subside. However, a social belief in the right to breathe clean air will likely be powerful in helping smokers to quit and in helping former smokers avoid relapse over the long term. For a comprehensive list of phenomenological interview findings, please see Appendix B, Table B3

#### *Convergence, Divergence and Unique Findings across Methods*

This section describes convergence, divergence and unique findings across methods by substantive topic area. A total of 14 topic areas that include findings from two or more methods were identified. These topic areas include external generalizability, respondent characteristics, exposure to smoke-free ordinances, support for ordinances, travel to restaurants or bars in a community with a smoking restriction different than one's own, self-reported impact of a cigarette tax, social norms, the smell

of smoke / experience of no ordinances, and the impact of exposure to ordinance on six outcomes (7-day abstinence and new quits, relapse, thinking about quitting, smoking less, better bar / restaurant experience, and anger).

A total of 88 substantive findings populate these topics, including three findings on a topic unique to surveys, five findings in two topics unique to focus groups, and nine topic / findings unique to interviews. On average, topic areas contained 5 findings each ( $SD = 3.28$ ). The fewest findings were yielded by the telephone survey ( $N=26$ ). The focus groups and interviews yielded about twice as many findings ( $N=61$  and  $N=59$ , respectively).

A comprehensive table that integrates findings from all three methods by topic area and assesses convergence, divergence and uniqueness may be found in Appendix C. Other relationships appeared and were documented, including when a finding was addressed by three methods, two of which converged and one of which diverged. This was labeled as a finding with convergence and divergence. Additionally, a handful of findings converged overall, but differed in the degree in the strength of the finding. These cases were labeled as having diverged by degree.

In order to understand the convergence, divergence and uniqueness of findings across methods by topic area, this section first provides a brief narrative description of convergence, divergence and uniqueness of findings across methods by topic. Next, it examines the findings for patterns in convergence, divergence and uniqueness.

*Narrative Description of Convergence, Divergence and Uniqueness by Topic*

*External Generalizability.* All studies have the same base population of tobacco users who enroll in an evidence-based QUITPLAN intervention. The population to which focus group and interview findings may be generalized is more similar to each other than to surveys because the standards for generalizability/transferability for interpretive methods differ from post-positivist ones. Additionally, similarities and differences are due to study sampling strategies. Findings from the survey may be generalized to QUITPLAN enrollees statewide, most of whom were still smoking at 18-months post-enrollment and may be considered hard core smokers. The focus group and interview were not designed to be generalized in the same way as surveys. Instead, findings should be transferable to other people and situations; phenomenological interviews achieve this through uncovering invariant meaning (Giorgi, 1997). Regardless, the samples of survey and focus group can be specified. They are QUITPLAN enrollees residing in the 7-county Twin Cities Metropolitan area. Non-smokers were oversampled in these studies so that an equal number of smokers and non-smokers were enrolled; this was not done in the survey because survey sampling was exhaustive to achieve the maximum number of subjects and increase power. Response rates for all methods fell within the range of 47% to 55%. The survey was unique in demonstrating its external validity through a response bias analysis. Results revealed that the key threat of rate of response differing by abstinence status at six

month follow-up was unwarranted. More stable eligible participants responded to the survey, consistent with similar studies.

*Respondent Characteristics.* Gender, age, race employment status, insurance status, highest education level completed and marital status were collected for all respondents of all methods. Clinical characteristics at intake and respondent satisfaction and use of QUITPLAN services at six months were also collected for all respondents in all methods. Most participant characteristics were similar across methods, except that focus groups and interviews included more uninsured participants, and more men participated in interviews. The interview and focus group methods uncovered a similar finding: that a subpopulation of respondents have multiple, severe problems in their lives, such as other addictions, anger management issues, serious health issues, entanglements in unhealthy relationships, mental health issues, etc.

*Exposure to Smoke-free Ordinances.* The proportion of respondents exposed to ordinances in the focus group and interview studies were more similar to each other than to the proportion exposed in the survey study. About one-third of survey respondents were exposed to ordinances as compared to about two-thirds of focus group and interview respondents as assessed via sampling criteria. The survey recruited participants statewide, while the focus groups and interviews recruited from the Twin Cities metropolitan area only, with two-thirds of participants from Hennepin or Ramsey counties (communities with ordinances), and one-third from the 5-county suburban non-ordinance communities. While participants' exposure to ordinances in the focus

group and interviews was similar, it did differ in degree. All interview participants reported frequenting bars or restaurants at least monthly, so had higher levels of exposure than focus group participants in general, some of whom went to bars or restaurants infrequently.

*Smoking Outcomes (Unique to Survey).* This topic comprises three findings that are unique to the survey method. Findings revealed that the 7-day abstinence intention-to-treat rate is statistically similar at 6 and 18 month follow-up. At 18 months, the 7-day abstinence completer rate is 32.7%. The rate of relapse is 26.3%. Interview and focus group methods did not gather information on point prevalence abstinence.

*Support for the Ordinance.* Respondent support for the ordinance differed across all three methods. The survey revealed that respondents were about equally likely to support and not support the ordinance, which differed from other population-based studies showing 70% approval ratings. Focus group participants frequently discussed their support for ordinances, but discussion was not encouraged because the strong political feelings that under lied discussion of support shifted group dynamics, which created divisiveness and distracted from participants' reporting on the impact of the ban. Interviews did not result in the same vociferous opinions, perhaps because of less group posturing, more social desirability, and self-selection. Participants against the ban in interviews were less oppositional, likely because of the one-on-one nature of the method.



*Travel to Bars / Restaurants for Ordinances.* All three methods converged on several key findings. First, most respondents did not travel regularly to bars / restaurants with different smoking regulations. However, some smokers traveled to bars to smoke and some non-smokers traveled to specific bars to avoid smoke. The frequency of this travel ranged from being occasional to regular. The survey uniquely found that those living in ordinance communities were more likely to travel than those living in non-ordinance communities. The survey and focus groups converged on the finding that smokers were more likely to travel than non-smokers, although the evidence was stronger for the survey method. Finally, the interview and focus groups converged to explain why people do and do not travel. Focus group and interview participants also explained why the ordinances had no impact on their travel.

*Impact of Tax on Smoking.* Due to the narrow focus of interviews on ordinances, taxes were not discussed in enough depth to include here, a limitation of this study. The survey and focus group findings converged on several findings. Both the survey and focus group found that the tax caused some smokers to think about quitting, to cut down on their smoking and to quit. At the same time, both methods revealed that the tax had no impact on many smokers. The focus group uniquely explained these findings: those with less disposable income were more likely to have to choose between cigarettes and other necessities, such as shoes for their children. Focus groups also explained two reasons why the tax had no impact (some respondents' addiction was so powerful and they still wanted to smoke, and many had enough disposable income to not be affected by the tax). Finally, the focus group uncovered

some unanticipated results of the cigarette tax: respondents reported using coupons and specials, switching to a cheaper brand, and traveling to neighboring states with lower taxes to buy cigarettes.

*Impact of Exposure on Abstinence and New Quits.* Findings within methods and across methods were mixed regarding exposure to ordinances on 7-day abstinence and new quits. First, the survey and focus group methods converged that exposure to ordinance had no effect on many smokers, although the interview diverged from this finding. At the same time, however, the focus group and interview methods converged on evidence that exposure to the ban did make an impact on quits. Specifically, those who reported no impact tended to not have been exposed to the ordinance. In fact, the interview method provided several unique explanations for an impact of the ordinance on quits that included a discussion of addiction, frustration, and panic; an experience of cognitive dissonance; and the smell of smoke that motivated some respondents to quit

The focus group method produced several unique findings that explain why exposure to ordinances and quits were perceived to be unrelated such as political beliefs and pride in quitting. Taken together, the focus groups findings led to the conclusion that the impact of the ordinances may be under-reported. Interviews converged with focus groups to support additional reasons for this conclusion, such as early adaptation to non-smoking family and friends, cognitive difficulty in assessing the impact of bans, and the strength of addiction.

*Impact of Exposure on Relapse.* All methods supported an inverse relationship between exposure to smoke-free ordinances and relapse. The survey found that exposure to the ban was marginally associated with relapse in a logistic regression ( $p=.061$ ). A bivariate examination showed a stronger, statistically significant relationship ( $p = .020$ ). Interviews and focus groups also found that bans helped smokers to maintain their quit. In interviews, over half of participants reported that the ban aided their cessation, including helping them prevent relapse. Ordinances created conditions where maintaining a quit attempt was easier.

*Impact of Exposure on Thinking about Quitting.* All methods indicated self-report that bans made people think about quitting. The focus groups and interviews converged on explanations for this effect, but the findings differed by degree. For example, focus groups and interviews explained that ordinances motivated some smokers to think about quitting because drinking and smoking in bars is inconvenient. However, interviews delved deeper into inconvenience. Focus groups and interviews also converged on the explanation that ordinances helped people think about quitting because they reduced the enjoyment of smoking because smokers must leave their friends to smoke. This prompted some smokers to feel guilty about their habit. Interviews confirmed patterns of hiding and sneaking and reported findings of shame and guilt in much greater depth. Feelings of guilt and shame were a central experience for many smokers.

*Impact of Exposure on Smoking Less.* Focus group and interviews converged on the finding that ordinances motivated smokers to smoke less, at least partially due to the inconvenience of having to exit an establishment to smoke. The focus group provided unique information on practical details such as a seasonal effect. Interviews provided unique underlying mechanisms for the effect, such as cognitive dissonance and the experience of the ban as a relief because some participants had a desire but not an ability to quit.

*Impact of Ordinances on Bar / Restaurant Experience.* The focus groups and interviews converged on the finding that ordinances gave smokers and non-smokers a substantially better bar and restaurant experience because they preferred an authentically smoke-free environment. However, the methods differed in degree. Interviews gathered more detailed and in depth information about smell and bar restaurant experiences.

*Experience of Anger.* Focus groups and interviews converged on the finding that some smokers were angered and frustrated by bans, because the bans curtailed their autonomy to smoke. Many saw smoking as a right, and felt resentful that this behavior was controlled by governments. Some reacted defiantly in a self-admittedly juvenile manner (i.e., I am going to smoke just because you told me not to). Focus groups contributed some unique findings, such as that people's perceptions of their rights differ; some believed they should be able to smoke everywhere, others in restaurants and bars, and others in bars and bowling alleys. Interview findings

contributed the unique finding that frustrated and angry smokers who were resentful and defiant of the ban defended and treasured their right to smoke in their home, even when doing so worsened their smoking-related illnesses.

*Social Norms.* Focus groups and interviews converged on several findings regarding social norms. All respondents understood that smoking is harmful and felt social pressure to stop smoking. Respondents acknowledged the harm of secondhand smoke. Some took social norms against smoking personally and felt like they were bad people or that they were being told they are bad people because they smoke. Bans and social norms against smoking were shown to be intertwined. While subtle and not immediate, social norms against smoking appeared to be the most promising mechanism to motivate new quit attempts, prevent relapse, and prevent individuals from starting to smoke in the first place. Focus groups uniquely described how respondents experienced social norms. Smokers reported seeing media messages and being glared at, condescended to, judged and treated rudely because they smoked. Finally, interview findings deepened an understanding of social norms by describing one consequence of them: guilt and shame from the social pressure not to smoke.

*Smell of Smoke / Experience of No Ordinance.* Focus groups and interviews revealed that smokers and non-smokers alike hated the smell of cigarettes, and that ordinances benefited all individuals who disliked the smell of cigarette smoke, regardless of their understanding or appreciation of it. Interviews diverged from focus groups in terms of degree, however. Interview subjects focused more strongly on the

negative aspects of cigarette smoke. Unprompted discussion of the smell of tobacco was nearly ubiquitous, almost all interviewees discussed it. Some interviewees described that smoky environments as a temptation, but the vast majority of the discussion was focused on the negative qualities of smoke. On the other hand, in focus groups, there was a greater emphasis on the smell of smoke as a temptation. Still, focus group participants reported better bar and restaurant experiences because of the ban as well.

*Findings Unique to Focus Groups.* Focus groups were the only methods that produced findings about the relationship between drinking, bars, and the effect of ordinances on smoking, and about the role of smoking accommodations in ordinances.

*Findings Unique to Interviews.* Interviews were the only methods that produced findings in the following areas: the impact of the Minnesota Clean Indoor Air Act, the use and impact of personal bans, the meaning of the term regulation, the definition of the function of a smoke-free ordinance, and several key experiences of ordinances, including relief, addiction, hiding / sneaking cigarettes, guilt, shame, and cognitive dissonance.

#### *Patterns in Convergence, Divergence and Uniqueness of Findings*

In order to better understand patterns of convergence, divergence and uniqueness of findings, the frequency of findings by relationship was tabled (see Table 4). The majority of findings, in equal proportions, either converged or were unique (43.2%, respectively). The next most common relationship of findings was to diverge

by degree (8.0%). Few findings both converged and diverged (4.0%) or diverged (1.1%).

Table 4

*Frequency of Findings in Each Mixed Method Relationship Group*

Relationship	N	%
Converge	38	43.2
Diverge	1	1.1
Converge & Diverge	4	4.5
Diverge by degree	7	8.0
Unique	38	43.2
Total	88	100.0

Next, the frequency of relationship of findings was tabled by method (see Table 5). Examining the frequency of findings by methods revealed that the the method combination with the greatest number of findings ( $N=30$ ) was the focus group and interview method combination. The survey and focus group combination also resulted in a high number of common findings ( $N=20$ ), followed by the number of findings unique to focus groups ( $N=17$ ) and unique to surveys ( $N=15$ ). Many common findings were also found in the survey and interview combination ( $N=14$ ) and in the comparison of all three methods ( $N=14$ ). The smallest number of findings was those unique to surveys ( $N=6$ ). Please note that findings may be counted more than once in different

method combinations; therefore, the sum of the subtotals exceeds  $N=88$  as shown in

Table 4.

In all method combinations, the majority of findings were convergent (64.3% to 80%). The method combinations that resulted in the highest proportion of divergent findings were the surveys and interviews and all three methods combined (35.7%, respectively).

Table 5

*Frequency of Findings in each Mixed Method Relationship Group by Method*

Method	N	%
Survey, Focus Group, Interview		
Converge	9	64.3
Diverge	1	7.1
C-FG&I, D-S	3	21.4
C-S&FG, D-I	1	7.1
Subtotal	14	100.0
Survey, Focus Group		
Converge	15	75.0
Diverge	4	20.0
Diverge by degree	1	5.0
Subtotal	20	100.0
Survey, Interviews		



Converge	9	64.3
Diverge	5	35.7
Diverge by degree	0	0.0
Subtotal	14	100.0
Focus Group, Interviews		
Converge	24	80.0
Diverge	0	0.0
Diverge by degree	6	20.0
Subtotal	30	100.0
Unique to Survey	6	100.0
Unique to Focus Group	17	100.0
Unique to Interview	15	100.0

Research Question 2: What are the substantiated findings of pragmatic versus dialectic mixed method studies? How are the two mixed method study findings similar and different from another? What unique information do the mixed method findings produce over and above single methods?

In the analysis of Research Question 1 above, the results from single methods were examined side-by-side. However, no integration of findings occurred, and no method for mixing at the stage of implementation (e.g., pragmatic or dialectic) was applied. Recall that this dissertation investigates pragmatic versus dialectic mixed method approaches. As described in Chapters 2 and 3, a dialectic mixed method study

will be created using findings from a telephone survey and phenomenological interviews, the findings of which will be mixed using a Hegelian-inspired dialectic format. Likewise, the pragmatic mixed method study will combine the findings from a telephone survey with focus groups, which are mixed according to pragmatic criteria identified in the literature. In both the pragmatic and dialectic mixed methods studies, mixing occurs at the stage of interpretation.

The purpose of this section is to develop the two mixed method studies (pragmatic and dialectic) by purposefully integrating findings using the pragmatic and dialectic frameworks, to identify how the two mixed methods studies are similar and different from one another, and to explore the unique information, if any, that the mixed method findings provide over and above the single methods. This section addresses each of these three concerns in turn.

#### *Findings of Pragmatic and Dialectic Mixed Method Studies*

This section discusses the process by which the two mixed method studies were conducted and provides a narrative summary for the substantive findings of each. The pragmatic mixed method study is presented first, followed by the dialectic study.

#### *The Pragmatic Mixed Method Study*

This pragmatic study combined the findings from the telephone survey with the focus groups. First, the process by which the individual studies were combined is discussed. Next, the pragmatic narrative that resulted from the pragmatic mixing is provided.

*The Process of Creating a Pragmatic Mixed Method Study.* The first step in creating the pragmatic mixed method study was to review the findings from the individual methods (telephone survey and focus groups). I compared the findings by topic area using the tables developed and fully presented in Appendix B. Convergence, divergence and unique findings were identified. The telephone survey and focus groups converged in most major topic areas, such as the impact of ordinances on relapse and other outcomes (7-day abstinence and new quits) and on travel. In these cases, the focus groups provided explanations for the effects that were seen in the survey findings and strengthened the findings over and above the single study results. Some divergence was identified between the telephone survey and focus group methods, but only by degree. These small differences were expected because they represent variations in study purpose, external generalizability and respondent characteristics inherent to the methods. The focus groups also provided some unanticipated, important findings that informed and influenced conclusions drawn from both methods. These findings include observed reasons participants may have reported no impact of the ordinance, the impact of social norms on quitting, and ordinances' relationship to those norms.

The second step was to integrate the findings of the two single methods according to the pragmatic criteria using the format identified in Chapter 3. Much of this mixing was so straightforward that the application of the pragmatic criteria was not required. For example, when findings confirmed one another (or were divergent to such a small degree that the impact was negligible), pragmatic considerations were not necessary to conclude that the findings strengthened one another. Also, many findings

from the focus group were unique and not mirrored in the telephone survey findings. These findings often stood alone and were not altered by employing pragmatic considerations.

However, findings on one topic, travel, were sufficiently complex that a pragmatic approach was useful. The telephone survey findings about travel suggested that those in ordinance communities were more likely to travel to a community with a different ordinance status. However, the focus group suggested that ordinances were about as equally likely to influence travel as not, but that any effect depended on smoking status. To resolve the issue of travel, I employed Datta's three criteria for pragmatism: knowledge, experience and contextual responsiveness. Examining findings on travel according to the pragmatic criteria brought forward important considerations that drew me to unique conclusions about travel to ordinance and non-ordinance communities. Table 6 illustrates how I attempted to use Datta's pragmatic approach (1997) to come to a conclusion about the topic. The use of this approach directly contributed to the narrative above and defines the narrative as pragmatic.

Table 6

*Pragmatic Mixed Method Decisions*


---

Topic	Travel to ordinance and non-ordinance communities
Telephone survey method findings	Those in ordinance communities were more likely to travel to a community with a different ordinance status. Those who traveled to communities in order to go to smoking establishments were more likely to be smokers.
Focus group method findings	No strong relationship between ordinance status and travel was found. Instead, ordinances were about as equally likely to influence travel as not, but any effect depended on smoking status. Smokers were more likely to travel to non-ordinance communities, and non-smokers were more likely to travel to ordinance communities.
Experience-based considerations	The association between ordinance status and travel in the telephone survey may be confounded by geographic location. Hennepin and Ramsey counties were the primary ordinance communities and are adjacent to metropolitan non-ordinance communities. Travel between these communities is easy and frequent, therefore influencing reported frequency of travel.
Knowledge-based considerations	The telephone survey and focus group findings converged that the magnitude of travel is small.
Contextually responsive considerations	The telephone surveys and interviews were conducted in the first month after the ordinances were implemented while residents were still adjusting to the new regulations. Local as well as outside tobacco-supported lobbying groups were active in fighting the regulations. Opinions

about the ordinances – especially negative ones – ran hot and high. This milieu likely influenced travel and how it was reported to me. For example, some focus group participants were very transparently politically motivated and appeared to report travel to other communities in a reactionary and ideological manner versus a personal one, suggesting that travel was over-reported.

#### Conclusion

Travel to communities with different smoking regulations appears to be associated with ordinances. However, the impact of the association does not appear to be meaningful because travel is infrequent and may have been over-reported, and possibly confounded with geographic location.

Reflecting on the process of creating a pragmatic mixed method study, I make several observations. First, telephone survey and focus group methods tended to converge with one another. However, the focus group did play an important role in generating important findings that explained how or why the relationships found by the telephone survey occurred. In these cases, the focus group findings explained the telephone survey findings in a simple, one-to-one manner. Second, the relatively convergent findings of the telephone survey and focus group made conclusions simple. The simple convergence was not suited to the pragmatic criteria outlined by Datta (1997). Finally, when findings were complex enough to use the pragmatic criteria, the criteria were effective in generating new thoughts and conclusions. The criteria more systematically introduced important experiential, knowledge-based and context-

oriented information that substantially informed and shaped the final conclusion in one specific topic area.

*Pragmatic Mixed Method Study Results: Narrative Findings.* This section provides the actual pragmatic mixed method narrative that was produced using pragmatic criteria for mixing discussed above. Please note that the narrative may appear redundant to the information presented in Research Question 1. This is due to the strong convergence between the two single methods that did not result in many unique findings. The pragmatic narrative may be found here:

The purpose of the telephone telephone survey<sup>3</sup> and focus group<sup>4</sup> studies was to understand the impact of the ordinance on QUITPLAN participants' quitting tobacco. Subjects from both the telephone telephone survey and focus groups were drawn from the same base population: tobacco users who enrolled in an evidence-based intervention, QUITPLAN. However, the samples from the two methods differed in some important ways. First, the interviews oversampled non-smokers in order to represent their opinions in the small number of focus groups conducted, whereas the telephone survey sample was either exhaustive or random depending on the region. Also, in order to ensure that participants could talk about their experiences of smoke-free ordinances, the focus group sample was drawn from residents in the Twin Cities 7-County metropolitan area, where ordinances were being instituted. The result is that a

---

<sup>3</sup> A follow-up telephone survey was administered to QUITPLAN participants 18 months post-enrollment. Outcomes were compared for those exposed and not exposed to an ordinance condition. A total of 1,169 QUITPLAN participants were telephone surveyed. The response rate was 55.3%.

<sup>4</sup> A total of 13 focus groups with 70 participants were conducted. Groups were two hours in length and smokers and non-smokers were grouped separately. The response rate was 46.7%.

greater proportion of interview respondents were exposed to smoke-free ordinances compared to telephone survey respondents.

Overall, the demographic and clinical characteristics of telephone survey and focus group respondents were approximately similar at intake, as were their satisfaction with and use of the program at 6 month follow-up. For example, a large proportion of both groups of participants were using tobacco at 18 month follow-up, so could be considered hard core smokers. However, a small number of differences in participants did emerge. For example, focus group participants were more likely to be male.

Telephone survey findings revealed that being exposed to an ordinance likely reduced relapse among QUITPLAN participants. Respondents exposed to the ordinance were about two times less likely to relapse than those not exposed. Only 15% of those exposed to the ordinance relapsed, as compared to 32% who were not exposed ( $p = .020$ ). Logistic regression results revealed that the relative risk of relapsing is  $-1.957$  and is marginally significant ( $p = .61$ ).

Focus groups corroborated the finding that smoke-free ordinances reduce relapse. Non-smokers reported that the ordinances created an environment where maintaining a quit attempt was easier because temptations to smoking were removed, such as seeing groups of people smoking and the smell of cigarettes. Many respondents shared that drinking is a powerful trigger to smoke for them and that they strongly associated bars with drinking and smoking. In fact, many participants who relapsed did so in bars. Therefore, focus group data suggested that ordinances helped some former



smokers to maintain their quit because they interrupted the association between drinking, smoking and bars.

The telephone survey examined the impact of ordinances on two outcomes other than relapse, 7-day abstinence and new quits, and found no impact of exposure to ordinances on them. About 70% of telephone survey participants self-reported that the ordinances made no difference. Fewer focus groups participants reported no difference, although it was still commonly reported. Focus group responses suggested that a key reason for no impact is lack of exposure to ordinances. For some, a neutral response reflected that they had already adapted to their family and friends' preference for a smoke-free environment, so the ordinance made no difference. Some non-smoking focus group members were so focused on the self-discipline they used to quit that they seemed as if they could be blind to anything that impacted their quit outside of themselves. Some smokers' addiction was so strong that the ordinance did not dent their smoking. Finally, some focus group members were very ideologically opposed to smoking ordinances and appeared to have a political agenda that would make it unlikely that they would report an impact, even if it existed. Taken together, the evidence from the telephone surveys and focus groups indicated that impacts of the ordinance may be under-reported, but even in this case, only a weak relationship existed, if any, between exposure to the ordinance and the outcomes of 7-day abstinence and new quits.

However, focus groups revealed positive impacts of the ordinance on outcomes other than relapse, especially for smokers. First, smokers reported smoking less due to

the inconvenience of the ordinances. Second, smokers reported disliking being separated from their non-smoking friends, and many said they felt the need to hide or sneak a cigarette out of guilt or shame. One particular participant expressed that ordinances take the enjoyment out of smoking, and that this was another reason to quit. Third, both smokers and non-smokers reported that a key benefit of smoke-free ordinances was a better bar or restaurant experience. For many, exposure to cigarette smoke was very unpleasant and even intolerable because of how smoke irritated the eyes, nose and throat. The thick smoke in bars caused even some smokers to cut their visit short.

Focus group results suggested that ordinances have an additional, indirect impact of motivating smokers to stop smoking. All participants knew the harm of smoking and felt social pressure not to smoke, and many associated ordinances with social norms against smoking and secondhand smoke. Therefore, smoke-free ordinances appeared to contribute to social norms against smoking. The social pressure not to smoke was intensely disturbing to many participants and motivated many participants to want and try to quit.

Considering all positive impacts of ordinances that focus group participants expressed, the most powerful, longest-term impact on smoking behavior may be ordinances' impact on norms. Some impacts of the smoking ordinances may be short-lived, for example, over time smokers may adjust to smoking outside and smoke the same amount regardless of the ordinance. However, the depth of participants' response to engaging in a socially unacceptable behavior suggested that the impact of social

norms may be more meaningful and lasting. While the novelty and new quits associated with recently instituted ordinances may fade, norms have the potential to have a longer term effect on current and potential smokers, as well as those who want to quit.

Finally, some smokers in focus groups reported negative impacts of smoke-free ordinances on their personal liberty. Some expressed defiance and resentment against smoke-free ordinances because the policy threatened their perceived right to smoke in public places. Frequently, these smokers were very angry. A small number of these smokers reported traveling to different cities or counties so they could go to smoking bars and restaurants.

One of the biggest threats to the validity of conclusions about the impact of the ordinance on smoking outcomes was travel between communities with different ordinances. The impact of the ordinance on quitting behaviors may be obscured by smokers in ban communities who frequently travel to bars in non-ban communities, or non-smokers who frequently travel to bars in ban communities. To address this, the telephone survey and interviews asked respondents about their travel to communities with different smoking restrictions and the reasons for that travel. Telephone survey findings revealed that most respondents don't travel to cities and counties with different smoking restrictions than where they live. Only 10% of all respondents and 39% of those that travel regularly reported traveling to bars or restaurants that allow smoking. Many more, 17% of all respondents and 67% of regular travelers, reported traveling for reasons other than smoking restrictions. Therefore,

most travel to bars and restaurants was likely due to work, socializing, tourism, etc., but not smoking restrictions. Focus groups confirmed that travel for the purposes of smoking was infrequent. Additionally, both methods found that a minority of respondents traveled to ordinance communities in order to avoid smoking atmospheres in non-ordinance communities. Focus group findings suggested they did so in order to avoid the smell of smoke they hate so much, or to support their quit attempt.

While the frequency of travel for smoking was low, the telephone survey did find an association between travel and ordinance and smoking status. Respondents in ordinance communities were more likely than those in non-ordinance communities to travel to visit smoking establishments, and those who traveled to smoking establishments were more likely to be smokers versus non-smokers. However, it is important to consider that the association between ordinance status and travel may be confounded by geographic location. Hennepin and Ramsey counties were the primary ordinance communities and were adjacent to metropolitan non-ordinance communities. Travel between these communities was easy and frequent, therefore influencing reported frequency of travel.

Focus group findings on the subject of travel were mixed. Ordinances were about as equally likely to influence travel as not, but any effect depended on smoking status. Smokers were more likely to travel to non-ordinance communities, and non-smokers were more likely to travel to ordinance communities.

In considering the impact of travel, it is important to note that the telephone surveys and interviews were conducted in the first months after the ordinances were

implemented, while residents were still adjusting to the new regulations. Local as well as outside, tobacco-supported lobbying groups were active in fighting the regulations. Opinions about the ordinances – especially negative ones – ran hot and high. This milieu likely influenced travel and how travel was reported to me. For example, some focus group participants were very transparently politically motivated and appeared to report travel to other communities in a reactionary and ideological manner, as opposed to a personal one. These conditions may suggest that travel was over-reported.

Taken together, the evidence from the telephone survey and focus groups suggested that travel to communities with different smoking regulations appeared to be associated with ordinances. However, the impact of the association did not appear to be meaningful because travel was infrequent and may have been over-reported, and possibly confounded with geographic location.

In sum, focus groups provided important explanatory power to precise telephone survey findings. Taken together, findings from the two methods converged on the impact of ordinances in reducing relapse. The two methods also converged on the weak impact of ordinances on the outcomes of 7-day abstinence and new quits, as well as on the conclusion that travel did not appear to meaningfully obscure outcome findings. By themselves, focus groups added unique focus to several unanticipated positive outcomes, including the importance of ordinances in shaping social norms, and the impact of norms on quitting.

*The Dialectic Mixed Method Study*

Mirroring the structure of the pragmatic section above, this portion provides a commentary on the process of producing a dialectic mixed method study, and then provides the actual dialectic narrative that resulted from the mixing.

*The Process of Creating a Dialectic Mixed Method Study.* As with the pragmatic mixed method study, the first step in creating the dialectic mixed method study was to review the findings from the individual methods (telephone survey and phenomenological interviews). The findings were compared and examined for convergence, divergence and uniqueness (see Appendixes B and C). The dialectic mixed method study was dominated by the unique findings generated by the phenomenological interviews in the intersection of participants' experience of the ordinance with addiction, shame, the smell of cigarette smoke and quitting. These elements led to a rich convergence with survey findings, where the power of addiction as revealed through the interviews were illustrative of low levels of self-reported support for and impact of the ordinance. This convergence was richer and more resonant because the interviews illuminated the underlying mechanisms that influenced behaviors as opposed to providing examples that related directly to survey results in a one-to-one manner. The relationship of addiction to support for the ordinance and self-reported impact was more complex and web-like. However, it is important to note that the phenomenological interviews produced simple convergence as well, such as when the survey finding of the relationship between relapse and exposure to ordinance was

supported by the interview finding that ordinances reduced triggers to smoking that supported quit attempts. Similarly, the two methods converged with one another in the area of travel. One substantial and major divergence in findings was identified regarding the impact of ordinances on abstinence.

The second step was to integrate the findings of the two single methods according to the dialectic process identified in Chapter 3. The mixing of convergent findings was so straightforward that the application of the dialectic criteria was unnecessary. Instead, the findings were simply presented as strengthening each other. Resolving divergent findings was more complex as the Hegelian dialectic stance is defined by resolving conflicting findings. When the telephone survey and interview findings were compared, one substantial set of divergent findings was identified: the impact of exposure to ordinances on the outcomes of 7-day abstinence and new quits. While the telephone survey found no impact, the interviews suggested the opposite and even provided detailed information about the underlying mechanisms by which the ordinances impacted outcomes. Table 7 below outlines how I used a Hegelian dialectic approach to resolve the divergence.

Table 7

*Dialectic Mixed Method Decisions*


---

Topic	Impact of Exposure of Ordinances on 7-day Abstinence and New Quits
Survey Conflicting Finding	There was no significant association between exposure to the ban and 7-day abstinence and new quits.
Interviews Conflicting Finding	<ul style="list-style-type: none"> <li>▪ Some interview subjects reported that the ban had no impact, although in most cases, this was because participants had little or no exposure to the ban. This suggests the impact of the ban was associated with exposure.</li> <li>▪ The interviews provided detailed information about the mechanisms by which the ordinances impact outcomes.</li> </ul>
Synthesis	Being exposed to ordinances does not directly increase quit rates. However, being exposed to bans can trigger a complex and influential set of experiences that motivate quitting. The impact of bans on quitting is indirect, and difficult to detect without moderating variables that capture motivating experiences.
Rationale	The lived experiences of QUITPLAN participants' exposure to bans provided convincing evidence of a possible impact that was not accounted for in the statistical model. However, the model does suggest that a clear and direct impact of the bans on quitting across all populations is unlikely.
Evidence Needed to Support the Synthesis	Information about the frequency with which the ban triggers experiences that motivate quitting and about the populations who are most sensitive to being influenced by the ban. Greater depth in understanding experiences that drive quitting.



The dialectic process forced me to integrate the two divergent claims regarding ordinance impact, assuming that both were correct. The quest for a new, more encompassing truth created a puzzle that forced me to be more creative in my approach to the mixing. The dialectic stance forced me to think beyond a simple side-by-side presentation of methods. For example, when I originally examined the divergence, I placed the statistical finding of the telephone survey (no impact) next to the overall finding of the phenomenological interviews (some impact). I could find no resolution to this divergence. More thoughtfulness and creativity was required. Upon further consideration, I determined that the underlying mechanisms for ordinance impact also was pertinent evidence to be considered. Finally, in the spirit of dialectic discovery, I reflected on the two diverse paradigms under consideration: the essential meaning sought by phenomenological inquiry, and the principals and methods of positivist inquiry. Considering how to integrate these two paradigms ignited for me the concept of direct and indirect effects, which resolved the tension between the diverse findings. I realized the telephone survey did not take into account the important factors that triggered an effect of the ordinance as illustrated in the phenomenological interviews. By suggesting an indirect effect, I was able to maintain both the interview findings, and the methodological integrity of the telephone survey. This solution was both consistent with survey findings and generative in the spirit of the dialectic approach. Obviously, more research is necessary to confirm this provisional conclusion. The last step in the dialectic mixing was to determine how to present the copious unique findings of the phenomenological interviews. Because these findings were critical to the dialectic

resolution discussed above and because they related to survey findings in a web-like manner, I chose to foreground the interview findings.

*Dialectic Mixed Method Study Results: Narrative Findings.* This section provides the actual dialectic mixed method study narrative that resulted from the mixing strategies described above. Please note that this narrative contains some redundancies from Research Question 1, where the convergence between telephone surveys and phenomenological interviews are enumerated, and to the divergences discussed above. None-the-less, the presentation of this narrative provides a complete picture of the dialectic analysis process. As such, readers may fully judge the quality of the dialectic mixing conducted; additionally, this narrative provides the transparency necessary for any future replication of this study. The narrative may be found below.

The purpose of the telephone survey<sup>5</sup> and interview<sup>6</sup> studies was to understand QUITPLAN participants' experience of local, smoke-free ordinances and the impact of the ordinance on participants' smoking. Subjects from both the telephone survey and interviews were drawn from the same base population: tobacco users who enrolled in an evidence-based intervention, QUITPLAN. However, the samples from the two methods differed in some important ways. The interviews oversampled non-smokers in order to represent their opinions in the small number of interviews conducted, whereas the telephone survey sample was either exhaustive or random depending on the region.

---

<sup>5</sup> An follow-up telephone survey was administered to QUITPLAN participants 18 months post-enrollment. Outcomes were compared for those exposed and not exposed to an ordinance condition. A total of 1,169 QUITPLAN participants were telephone surveyed. The response rate was 55.3%.

<sup>6</sup> A total of 15 phenomenological interviews were conducted of participants. Interviews were one hour in length and conducted via telephone. The goal was to understand participants' experience with the ordinances, including the experiences with ordinances and quitting. The response rate was 51.7%.

Also, in order to ensure that participants could talk about their experiences of smoke-free ordinances, the interview sample was drawn from residents in the Twin Cities 7-County metropolitan area, where ordinances were being instituted. The sample was further narrowed to those who frequented bars and restaurants monthly or more. In contrast, the telephone survey sample was statewide and did not sample on the frequency of bar and restaurant use. The result is that a greater proportion of interview respondents were exposed to smoke-free ordinances compared to telephone survey respondents.

Overall, the demographic and clinical characteristics of telephone survey and interview respondents were approximately similar at intake, as were their satisfaction with and use of the program at 6 month follow-up. For example, a large proportion of both groups of participants were using tobacco at 18 month follow-up, so could be considered hard core smokers. However, a small number of differences in participants did emerge. First, interview participants were more likely to be male. Second, a substantial proportion of interview respondents (57%) volunteered that they had multiple problems in their lives, such as other addictions, entanglements in unhealthy relationships, mental health issues, etc. This information was not collected from telephone survey respondents.

Interviews revealed that experiences of smoke-free ordinances are often intense. Those still smoking were confronted by their addiction because the ordinances prohibit them from smoking. The impulse to smoke was overpowering; for example, at a party one woman overcame the urge to tackle a friend for his cigarette and instead begged

him to blow smoke on her. Participants also worried about coping without cigarettes. For example, one man wondered if his car would even start without a cigarette in his hand. In the face of an ordinance, participants felt apprehension and became obsessed with planning when and how they could have their next cigarette. One man calculated exactly how soon he could have his next cigarette while chaperoning his child's field trip to the smoke-free Mall of America, and ran the plan through his head in an endless loop until he could smoke. Participants reported spending enormous amounts of energy coping with their addiction.

Ordinances caused smoking participants to experience the extent of their addiction and inability to control their cravings. For example, one man craving a cigarette rooted through his backyard in the spring to find a soggy butt to smoke that he threw there last fall. Confronting addiction was painful for most participants, however, it motivated some to quit. For example, one restaurant server quit because she couldn't handle long shifts smoke-free, and a businessman hated being controlled by nicotine and quit when the ordinance forced him to confront his addiction on business trips. Others simply struggled with their addictions and continued to smoke.

Interviews revealed that shame was a key experience in smoking. Shame is marked by feeling seen "in a powerfully diminished sense. The self feels exposed both to itself and others" (Kaufman, 1992, p. 45). One smoker with a severe smoking-related illness continued to smoke and berated herself as "stupid". Most smokers castigated themselves for smoking because they saw themselves as lacking the discipline to quit when most of their friends already have. Smokers also felt belittled by

others, from strangers glaring at them outside restaurants, to family and friends with whom participants reported hiding or sneaking their smoking. Ordinances created a new set of occasions for shame for participants who were confronted by their addiction. However, ordinances were not the sole generator of shame. Shame is a common component of addictions to alcohol, other drugs, gambling, etc.. It is both a precursor to addiction, as addicts turn towards alcohol or other drugs in order to cope with feelings of shame, as well as a consequence of it, as addicts feel shame for their addictive behaviors (Bradshaw, 2005). Participants reported feeling shame for smoking before the ordinances were instituted.

The smell of cigarette smoke was another key experience illuminated by the interviews. Smokers sometimes referred to cigarettes as “their only friend” and the smell of smoke was a temptation to smoke. At the same time, both former smokers and smokers reported they hated the smell of smoke in the strongest terms, describing it as “nasty” “horrible” “disgusting” and “god awful”. Participants described smoke traveling freely like air and permeating everything around it. They described with disgust how the smell of smoke attached to clothes and hair and lingered, an involuntary and disconcertingly intimate signal of being a smoker or being around smoke. Some smokers described how the smell can “out” a person as a smoker even if he or she would like to keep it a secret. Participants associated the smell of secondhand smoke with being dirty and with the shame of smoking and addiction. Given the powerful associations of the smell of smoke, it is not surprising that smell was both a motivator to quit smoking and a strategy to maintain a quit.

The smell of cigarettes was related to ordinances because smoke-free ordinances greatly reduced the smoke from indoor air, reducing the hated smell. Everyone who reacted negatively to the smell of cigarette smoke received relief. By removing smoke from the air some non-smokers and smokers reported a better bar and restaurant experience.

Given the difficult physical and emotional experience of ordinances, it is not surprising that telephone survey findings revealed that only about half of QUITPLAN participants supported the ordinance at 18-month follow-up, as compared to about 70% of Minnesotans. The difficulty of confronting the ordinance may also explain why most telephone survey respondents, about 70%, self-reported that the ordinance did not make a difference in their quitting.

Similarly, the telephone survey found no significant association between exposure to the ban and the outcomes 7-day abstinence and new quits. However, interview findings diverged on this point. Over half of participants reported that the ban was an aid to their quitting, including motivating them to quit. Interviews provided detailed information on the mechanisms triggering these affects. Additionally, some interview subjects reported that the ban had no impact, although in most cases, this was because participants had little or no exposure to the ban. This suggested the impact of the ban is associated with exposure.

Considering these disparate findings together, I conclude that being exposed to ordinances does not directly increase quit rates. However, being exposed to bans can trigger a complex and influential set of experiences that motivate quitting. The impact

of bans on quitting is indirect, and difficult to detect without mediating variables that capture motivating experiences. The lived experiences of QUITPLAN participants' exposure to bans provided convincing evidence of a possible impact that was not accounted for in the statistical model. However, the model did suggest that a clear and direct impact of the bans on quitting across all populations was unlikely. Information that would further refine this conclusion includes understanding the frequency with which the smoke-free ordinances trigger experiences that motivate quitting and about the populations who are most sensitive to being influenced by the ban. Finally, greater depth in understanding experiences that drive quitting would be helpful.

While the telephone survey and interview findings diverged on the impact of smoke-free ordinances on quitting, they converged on the negative impact of ordinances on relapse. Telephone survey respondents exposed to the ordinance were about two times less likely to relapse than those not exposed, reflecting the desired outcome. Only 15% of those exposed to the ordinance relapsed, as compared to 32% who were not exposed ( $p = .020$ ). Logistic regression results revealed that the relative risk of relapsing is -1.957 and was marginally significant ( $p = 0.61$ ). Some interview respondents reported that ordinances prevented their relapse, but more importantly, they provided detailed information about how ordinances facilitated that effect. For example, interviews reported that bans created conditions where maintaining a quit attempt was easier: they removed the temptation of cigarette smoke from bars and restaurants and mostly removed the temptation of seeing people, especially friends, smoking. It also forced people to leave their friends to drink and smoke, and being

separated from their friends took the enjoyment out of smoking. Finally, having to physically move to smoke was inconvenient and could cause cognitive dissonance where smokers' knowledge of the harm of smoke conflicted with the extra actions they were taking to smoke.

One of the biggest threats to the validity of conclusions about the impact of the ordinance on smoking outcomes was travel between communities with different ordinances. The impact of the ordinance on quitting behaviors may be obscured by smokers in ban communities who frequently traveled to bars in non-ban communities, or non-smokers who traveled frequently travel to bars in ban communities. To address this, the telephone survey and interviews asked respondents about their travel to communities with different smoking restrictions and the reasons for that travel. The two methods converged with one another on a number of findings:

- Most respondents did not travel regularly to bars or restaurants in communities with different smoking regulations
- However, some smokers traveled to non-ban communities and some non-smokers traveled to non-smoking communities
- Travel ranged from very infrequent to more frequent

Telephone survey findings were unique in revealing that those in ban communities were more likely to travel than those in non-ordinance communities, and that those who travel to non-ban communities were more likely to be smokers. Taken together, these findings suggested that traveling may have an impact on outcomes, but the effect was small because the frequency of travel is low.



In sum, interviews were unique in illuminating the experience of smoke-free ordinances. The interviews and telephone surveys diverged in their conclusion about the impact of ordinances on quitting smoking. Taken together, the information suggested that the impact of bans on quitting was indirect and mediated by particular experiences of bans that motivate quitting. However, the telephone survey and interview methods converged to support the conclusion that ordinances negatively impacted relapse, and that travel to communities with different smoking regulations may be a threat to conclusions about the impact of the ban, but was a meaningful one. When findings converged, interviews provided important explanatory power to precise telephone survey findings.

#### *Comparison of Mixed Method Process and Findings*

This section compares the pragmatic mixed method study to the dialectic. First, the two studies are compared in terms of the process used to develop them. Next, the convergence and divergence of results from the two studies are examined.

#### *Comparing the Pragmatic versus the Dialectic Processes*

The pragmatic and dialectic mixing processes were similar in one important regard: unique findings resulted from divergence. In both the pragmatic and dialectic studies, convergence strengthened conclusions but provided no new insights. At the same time, the pragmatic and dialectic mixing differed from each other in several important ways. First, the methods in the pragmatic study tended to converge more than the methods in the dialectic study. Further, the convergence in the pragmatic study

was straight-forward in that the findings mapped to each other in a simple one-to-one manner. In contrast, the convergence in the dialectic study tended to be more complex. This is likely because the phenomenological interviews in the dialectic study illuminated mechanisms that girded behaviors observed in the telephone survey. These mechanisms resonated with several survey findings, so the phenomenological interview findings related to the telephone survey findings in a web-like manner. Finally, the dialectic mixed method study tended to produce more divergent findings than the pragmatic study. Taken together, the dialectic study tended to be more generative than the pragmatic study.

#### *Convergence and Divergence of Mixed Method Findings*

This section compares the substantive findings of the pragmatic and dialectic mixed method studies. One major convergence and one major divergence in findings emerged from a review of the major issues in the two studies. However, on six issues, more subtle differences emerged that could not be clearly classified as either a convergence or divergence. Finally, each mixed method study brought to the table one important and unique issue that was not addressed in the other study. All of these similarities, differences and unique findings will be discussed in turn.

*Convergent Finding.* A strong and clear convergence between the pragmatic and dialectic mixed method studies was found on the issue of relapse. In both pragmatic and dialectic studies, the telephone survey method concluded that being exposed to an ordinance was negatively associated with relapse. In both studies, the

interpretivist method supported the conclusion with participants' experience of ordinances preventing relapse, and by describing how and why the relationship worked. In the pragmatic study, the explanation centered on the relationship between smoking, drinking, bars and relapse, while in the dialectic study, participants' experiences in bars with temptation, social smoking and feeling ostracized explained relapse. Despite these differences, in both studies the conclusion was the same: ordinances impacted relapse downward.

*Divergent Finding.* The pragmatic and dialectic mixed method studies both addressed the impact of ordinances on 7-day abstinence and new quits, but came to different conclusions. The pragmatic study resulted in a greater convergence of evidence that ordinances had no impact on the outcomes of 7-day abstinence and new quits. The telephone survey found no relationship, while the focus groups found a weak relationship, as well as evidence that the impact may be under-reported. In contrast, the dialectic study generated convergence on the impact of ordinances on the outcomes. The telephone survey found no relationship, but the interviews revealed that ordinances did help people quit. They also outlined in depth several mechanisms by which the impact occurred, such as through cognitive dissonance and the guilt and shame of smoking. Based on a dialectic resolution of the conflicting evidence, I conclude that a relationship between exposure to ordinances and quitting does exist, but that the relationship is indirect and heavily mediated by the experiences outlined in the interviews.

*Subtle Differences.* For six topics, the difference between the pragmatic and dialectic mixed method study findings were nuanced, and categorizing the difference as a convergent or divergent was not applicable or helpful. For example, several topics were primarily similar to one another, but diverged only in degree. This was the case for the purpose of the study, where the dialectic study differed from the pragmatic because it explored participants' experiences more deeply. This reflected the differences between the focus group and phenomenological interview methods. Both solicited participant experiences, but the interviews sought more in-depth information that most authentically reflected participants' lived experience. The one-on-one interview format facilitated this discovery.

Another example of difference by degree can be found in the topic of external generalizability. In this case, the dialectic study included participants from a slightly more specific population than the pragmatic study: those who frequented bars or restaurants monthly or more. This resulted in interview participants having greater exposure to smoke-free ordinances than focus group participants.

Finally, the pragmatic and dialectic mixed method studies differed by degree in terms of respondent characteristics. The interviews used in the dialectic mixed method study had a greater proportion of participants who reported having multiple problems in their lives, such as addiction, anger management issues, entanglements in unhealthy relationships, etc. The focus groups included participants who self-disclosed problems such as these, but to a lesser extent.

Overall, the pragmatic and dialectic mixed method studies were quite similar to each other in terms of overall assessments of study purpose, external generalizability and respondent characteristics. Differences between the studies were only by degree, and reflect the differences in the focus group and interview methods. The differences are real and should be seriously considered, but pale in comparison to magnitude of ways in which the studies are similar to one another.

Another way the dialectic and pragmatic studies subtly differed from each other was in how they framed issues. Both the pragmatic and dialectic studies described the role of guilt and shame and the smell of cigarette smoke. However, the two studies framed these topics very differently. For example, in the pragmatic study, the topic of guilt and shame was presented as evidence of unanticipated positive outcomes. For example, the pragmatic study reported that smokers were forced to leave their friends to smoke under ordinances. They reported feeling like they needed to hide in order to smoke, and that feelings of guilt and shame accompanied smoking. These feelings fueled a positive, unanticipated outcome that ordinances provided smokers another reason to quit because they sucked the joy out of smoking.

In contrast, the dialectic study discussed the guilt and shame that smokers felt in much greater depth, and as a core part of their lived experience. Smokers castigated themselves for their weakness in the face of addiction just as they reported being belittled by others when they smoked in public places. The dialectic study described how shame and guilt were an integral component of addiction, and thus both a precursor and antecedent of the ordinances.

A second topic that was framed differently by the two studies was the smell of smoke. Like guilt and shame, the dialectic study described smoker and non-smokers experience of the smell of smoke in great depth. The physical sensation of being in a smoky room was described, as well as the emotions and memories that the smell of smoke conjures. The smell of smoke was a core component of the lived experience of ordinances because ordinances removed the smell of smoke from the environment.

Conversely, the pragmatic study referenced the smell of smoke, but in less detail, and as evidence of another unanticipated positive outcome: a better bar and restaurant experience. The pragmatic study described how many smokers and non-smokers hated the smell of cigarette smoke, so enjoyed going into non-smoking bars much more.

Taken together, the evidence suggested that a pragmatic approach used participant experience consequentially – as evidence of unanticipated positive outcomes, for example. This reflected the tenets underlying pragmatism. In contrast, the dialectic approach gathered more in-depth information about participants' experiences, and held the experience as descriptive of a phenomenon by itself. This reflected the phenomenological theory undergirding the interviews. Therefore, the different ways in which similar topics were framed reflected the diverse paradigms in mixed method studies.

A final, very interesting example of the subtlety and nuance with which the mixed method studies diverged from each other can be found in the topic of travel. The evidence presented on travel in the pragmatic study suggested moderate divergence.

The telephone survey and focus group more strongly suggested that travel is an issue for smokers and non-smokers than the dialectic study. However, several pragmatic considerations were applied to the evidence of travel, including a possible confounding factor and the timing of methods. Once these factors were considered, the pragmatic telephone survey concluded that travel was an issue, but not a meaningful one because it was so infrequent. The dialectic method came to the same conclusion, even though the evidence was less convergent than in the pragmatic study.

*Unique Findings.* The pragmatic study explored one issue unique to itself: the role of social norms. The pragmatic study discussed how smokers felt social pressure not to smoke and how this pressure motivated some to stop smoking. The discussion of social norms stemmed from the focus groups within the pragmatic mixed method study, which concluded that social norms may have the most potential to create lasting change in current and potential smokers, and those who want to quit. The fact that the focus group provided the most detailed information about social norms is interesting, because each focus group was its own laboratory of social interactions.

The dialectic study was unique in reporting on smokers' experiences smoking and quitting. These findings were found in the phenomenological interviews, and discussed in great detail the intensity of addiction and smokers' coping mechanisms for it, and the way that ordinances forced smokers to face their addiction more clearly. The dialectic study reported that this experience motivated some smokers to quit.

*What unique information do the mixed  
method findings produce over and above single methods?*

The purpose of this section is to understand what new information mixed method studies contribute beyond the findings of single method studies. A review of tabled findings from single and mixed method studies revealed that the no new findings were generated. However, the combination of two methods in a mixed method studies generated unique conclusions and interpretations based on the joint examination of findings on a similar topic. Table 8 below highlights the new conclusions and inferences that the dialectic and pragmatic mixed method studies contributed as compared to those of single methods.

Findings on the topic of relapse illustrated how the mixed method studies brought conclusions with new strength forward. For both mixed method studies, the telephone survey provided a somewhat equivocal finding of marginal significance between relapse and exposure to the ban. In contrast, the focus groups and interviews both supported a positive relationship between ordinances and relapse, and explained how the relationship worked. When taken together, the focus group findings strengthened the telephone survey findings in the pragmatic study, and the interview findings strengthened the telephone survey findings in the dialectic study. Both mixed method studies concluded that relapse had a positive association with ordinances more strongly, and with more evidence, than in any of the single studies.



Table 8

*Key Conclusions by Study*

Topic	Mixed Method Conclusions & Interpretations		Single Method Conclusions & Interpretations		
	Pragmatic	Dialectic	Telephone survey	Focus Group	Interview
Impact of ordinances on relapse	Telephone surveys and focus groups converged on the conclusion that ordinances impact relapse.	Telephone surveys and interviews converged on the conclusion that ordinances impact relapse.	Exposure to the ban was marginally associated with reduced relapse.	Bans created conditions where maintain-ing a quit attempt was easier.	Bans created conditions where maintain-ing a quit attempt was easier.
Impact of ordinances on 7-day abstinence and new quits	The impact of the ordinance on 7-day abstinence and new quits may be under-reported. However, even in this case, the relation-ship is weak.	Being exposed to ordinances trigger a complex set of experiences that mediate an indirect relationship between ordinances and 7-day abstinence and new quits.	Exposure to the ban was not significant-ly associated with 7-day abstinence or new quits.	For some, the ban was not helpful and had no impact. However, reported no impact appears to be influenced by several factors.	Over half of respondents indicated the ordinances impacted their quit, including motivating them to quit. The mechan-isms by which this impact was achieved are described.

Travel

- Travel because of ordinances is infrequent

- Travel appears to be associated with ordinances, but the impact is not meaningful

- When considering the impact of travel, it is important to note that telephone surveys and interviews were conducted during the first, controversial months of the ordinance. This may have led to more reactionary and ideological responses.

- Travel because of ordinances is infrequent

-Travel appears to be associated with ordinances, but the impact is not meaningful

- Travel because of ordinances is infrequent

- Most respondents do not travel regularly, but being in an ordinance community is positively associated with regular travel and travel to smoke

- Travel because of ordinances is infrequent

- Bans appear to be equally likely to impact travel or not, but any effect is determined by smoking status.

- Travel because of ordinances is infrequent for several reasons. Some smokers travel to go to smoking establishments, while some non-smokers travel to go to smoke-free establishments.

Findings on the topic of the relationship between outcomes (7-day abstinence and new quits) and ordinances exemplified how the mixed method conclusions were unique in their complexity, over and above any of the single methods. In the pragmatic study, the conclusions gained complexity in two ways. First, the convergence of telephone survey and focus group findings that ordinances and outcomes were unrelated strengthened the conclusion. Second, the focus group findings highlighted several reasons why impact may be under-reported. Considering these factors together, a more nuanced conclusion emerged that the relationship between outcomes and ordinances was determined to be possibly under-reported, but weak at best.

Like the pragmatic study, an examination of the relationship between outcomes and ordinances generated conclusions of greater complexity in the dialectic study. The telephone survey found a lack of relationship, which was softened by the interview findings of a relationship. Perhaps more importantly, the interviews specified the mechanisms by which the relationship functioned. Using a Hegelian dialectic model, the dialectic study concluded that there appeared to be a relationship between outcomes and ordinances, but only an indirect one. Recommendations for future research on this topic were proposed.

The examination of travel in the pragmatic and dialectic studies exemplified how stronger and more nuanced conclusions emerged from the mixed method studies as compared to the single method studies. In the pragmatic study, the finding that travel is infrequent was strengthened by convergence of telephone survey and focus group methods. This conclusion qualified the telephone survey finding that ordinances and

travel are related. A discussion of the politically contentious period during which data collection took place further qualified the finding. Taken together, the pragmatic study concluded with a more complex finding than individual study results: that travel appeared to be associated with ordinances, but not to a meaningful extent. The dialectic study resulted in the same nuanced finding using similar logic.

Research Question 3: To what extent do the inferences drawn from single method findings meet key criteria for validity / trustworthiness? To what extent are conclusions drawn from single and mixed method findings valid / trustworthy according to *The Program Evaluation Standards* (Joint Committee, 1994)?

The purpose of this section is to understand the extent to which conclusions from real-world single and mixed method study findings meet key standards for validity / trustworthiness. Recall that Lincoln and Guba's unified conception of validity / trustworthiness (1985) will be used across post-positivistic and interpretivist methods to assess the validity of conclusions drawn from the telephone survey, focus group and phenomenological interview findings in this dissertation. Table 9 below provides a summary of the relevant concepts. Also recall that the validity / trustworthiness of conclusions drawn from single and mixed method findings will be assessed in light of *The Program Evaluation Standards* (Joint Committee, 1994). Finally, an expert review of single study methodologies will be conducted. It is important to note that the following exploration of validity / trustworthiness reflects my interpretation of the

criteria notes above and my opinion of what it would take to in each instance to meet them. Based on this, I present arguments for my opinion of how well each criteria was met. In short, the assessment of all validity / trustworthiness criteria that follow is based on my interpretations and argumentation.

### *Trustworthiness / Validity*

This section has four parts that correspond to criteria for validity / trustworthiness which are partially outlined in Table 9 above. The first part discusses the “truth value” for each method (internal validity for telephone survey, credibility for focus groups, and credibility for phenomenological interviews). The applicability of each method is examined second (external validity for telephone survey, transferability for focus groups, transferability for phenomenological interviews). Third, the consistency of each method is discussed (reliability for telephone survey, dependability for focus groups, dependability for phenomenological interviews). The fourth part discusses neutrality for each method (objectivity for telephone survey, confirmability for focus groups, confirmability for phenomenological interviews).

#### *“Truth Value”*

According to Lincoln and Guba (1985), “truth value” refers to how to establish confidence in the “truth” of a study’s findings for its subjects, given a particular context in which the study took place. “Truth value” is expressed differently in post-positivist and interpretive studies, and is described below for the three single methods conducted in this dissertation.

Table 9

*Summary of Lincoln and Guba's (1985) Unified Theory of Validity / Trustworthiness*

Trustworthiness Concept	Positivist Analog		Interpretive Analog	
	Concept	Test / Method	Concept	Test / Method
"Truth Value"	Internal validity	Research design – randomized control groups and manipulation of independent variables	Credibility	prolonged engagement or persistent observation, triangulation, building trust, peer debriefing, negative case analysis, referential adequacy, and member checks.
Applicability	External validity	Random sampling and testing for interactions between the treatment and key factors	Transfer- ability	Thick description of context and dataset of minimum elements needed to assess transferability
Consistency	Reliability	Coefficient alpha via test-retest, equivalent forms or split-half reliability	Depend- ability	Overlap of methods, stepwise replication, and inquiry audit
Neutrality	Objectivity	Distance between observer and observed, value-free measurement	Confirm- ability	Audit trail and audit process

*Internal Validity of Telephone Survey.* In post-positivist methods such as the telephone survey conducted for this dissertation, “truth value” is expressed as internal validity and indicates the extent to which the outcome may be attributed to the intervention. This is chiefly realized through the use of control groups and manipulation of the independent variables. I was unable to design control groups for the intervention, which was the institution of smoke-free policies. Because the study was observational in nature, nor was I able to manipulate independent variables. Therefore, the internal validity of the telephone study was low.

*Credibility of Focus Groups.* The focus group conducted for this dissertation was guided by an interpretive paradigm in which “truth value” is expressed through credibility. According to Lincoln and Guba (1985), credibility is judged by prolonged engagement / persistent observation, triangulation of investigators, methods, data sources and theories; participant-investigator trust; peer debriefing; negative case analysis; referential adequacy; and member checks. Areas of high, moderate and low credibility will be discussed below. Credibility criteria were most strongly met in the areas of peer debriefing, negative case analysis and referential adequacy. Each of these is discussed in more detail below.

- Peer debriefing. Because focus groups were conducted by two to three people (a facilitator, a note taker and an assistant when possible), peer debriefing was instituted in the focus group protocol. After the close of the focus group, the team would debrief for approximately 10 minutes on tape

about the content and process of the focus group. As more focus groups were conducted, the team debriefed about emerging themes across groups as well. The peer debriefings were transcribed and used as a data source in analysis. This criterion was strongly met.

- Negative case analysis. I conducted negative case analysis when analyzing the focus group data. The area in which this was most productive was analyzing the impact of the ban on participants. Theories of positive ban impact were challenged by testimony from focus group members who claimed no impact or negative impact, and vice versa. This strongly met criterion of credibility resulted in a more balanced description of ban impact in the focus group results.
- Referential adequacy. This refers to the availability of source data to the investigator for use and to other researchers for confirmation. Focus groups were taped and transcribed. I read the transcript of every focus group dozens of times, and based my analysis on this. Although the data exists, it was not provided to other researchers. None-the-less, I consider the referential adequacy to be strong.

Credibility criteria were met with more moderate strength in the area of participant-investigator trust. Through the use of carefully crafted prenotification letters, telephone solicitations for participation and follow-up communications, I prioritized building trust with respondents. During the focus groups, I was mindful to respect all participants, encourage participation, and referee discussions with fairness



and equanimity. While not intensive, this communication likely enhanced the credibility of conclusions drawn from focus group findings.

The areas where validity criteria were most weakly met were triangulation, persistent observation and member checks. Triangulation did not substantively occur at the level of methods, investigators or theories. However, triangulation did occur at the level of data sources, when smokers and non-smokers were purposively sampled. The credibility criteria of persistent observation was also weakly met. The observation was for two hours with an average of six participants. This is a relatively short period of time for interpretive research. Also, member checks were not conducted due to the brief interaction.

*Credibility of Phenomenological Interviews.* The same criteria of credibility applied to focus groups above are used to judge interpretive phenomenological interviews in this section. Credibility criteria which were met at strong, moderate, and weak levels are discussed below. In the interviews, the most strongly met criteria were negative case analysis, respondent-investigator trust, and referential adequacy. Each of these areas is discussed in more detail below.

- Negative case analysis. Similar to the focus groups, negative case analysis was used to analyze the impact of the ban. It was a more effective technique in the interviews as compared to the focus groups because the interviews contained thicker description and more vivid detail upon which to draw.

- Respondent-investigator trust. Similar to the focus groups, I attempted to gain trust through carefully crafted prenotification letters and initial telephone communications. Conducting the phenomenological interviews with respect and fairness was a baseline expectation. The one-on-one interviews elicited intimate content about participants' smoking, addiction, and experiences with the ban as compared to the focus groups. Therefore, the credibility criteria of trust was met more strongly for the interviews than with the focus groups.
- Referential adequacy. The referential adequacy of the phenomenological interviews was also strong because the interviews were taped and transcribed. As with the focus groups I read the transcripts dozens of times. However, they have not been analyzed by other researchers. Despite this, the referential adequacy was strong.

In the phenomenological interviews, a total of three credibility criteria were met with moderate strength. They are prolonged engagement, triangulation of theories and data sources, and member checks, and are discussed more substantively below.

- Prolonged engagement. Interviews only lasted approximately one hour, which is brief for interpretive research. However, the one-on-one nature of the interview allowed me to explore smoke-free regulations in depth. By focusing on participants' experiences, I was able to capture richer description and vivid detail. While the criteria of prolonged engagement

was met at only a moderate level, I believe the intensity of the encounter lead to greater credibility as compared to the focus group.

- Triangulation. Moderate levels of triangulation occurred among theories and data sources. Because of the phenomenological approach used for the interviews, the use of multiple data sources was encouraged. Therefore, an etymological analysis was conducted, and literature on psychological theories on topics of addiction, shame and cognitive dissonance were explored. This triangulation resulted in richer conclusions than were reached in focus groups.
- Member checks. Member checks were conducted when I reviewed tapes and/or transcripts and did not sufficiently understand participants' recorded comments. This occurred four times (one-third of the interviews). The criterion of member checks is only moderately met because I did not provide my findings to respondents for their assurance of accuracy.

The credibility criterion of peer debriefing was not met within the phenomenological interviews. As the sole investigator, I conducted all interviews without assistance from other team members. I wrote notes about my reactions and impressions of the interview for myself, but did not institute a peer debriefing process for the interviews.

*Summary of "Truth Value" Across Single Methods.* Overall, the "truth value" of the telephone survey is low. Both the focus groups and interviews well meet criteria for truth value due to negative case analysis and referential adequacy. Lower levels of

credibility were achieved for the five other criteria for credibility. When the focus groups and phenomenological interviews are compared, conclusions from the phenomenological interviews overall demonstrated higher credibility than the conclusions from the focus groups in terms of negative case analysis, prolonged engagement, triangulation and member checks. Focus groups exceeded phenomenological interviews in meeting credibility criteria only for peer debriefing.

### *Applicability*

Applicability refers to the extent to which findings are relevant to other contexts or subjects. Post-positivist methods like the telephone survey conducted for this dissertation express applicability as external validity, and is described below. Next, the interpretive focus groups and phenomenological interviews describe how applicability is expressed through the criteria of transferability.

*External Validity of Telephone Survey.* External validity is concerned with the ability of a study to generalize to other populations. This is achieved through random sampling and examining interactions between the outcome and population characteristics, setting characteristics, etc. Considering sampling in the telephone survey study, both exhaustive and random sampling techniques were employed. For the 7-county metropolitan area, sampling was exhaustive, and random sampling was employed for counties in outstate Minnesota. Overall, the sampling strategies used support claims of generalizability to QUITPLAN participants.

A more precise argument for generalizability may be developed when response bias is examined. The telephone survey achieved a 55.3% response rate, and a response – non-response analysis was conducted to understand the extent to which survey responders represented the sample from which they were drawn. Because the tobacco cessation literature found that previous abstinence predicts future abstinence (Hooten, Wolter, Ames, Hurt, Vickers, & Hayes, 2005), the key concern for bias was that participants' abstinent at 6 months would be more likely to respond at 18 months. A response bias analysis revealed that the key concern for bias was unwarranted. Response to the 18 month survey was similar regardless of whether a person was abstinent at 6 months. Therefore, concern that only the most successful QUITPLAN participants responded to the 18 month survey is unwarranted.

However, response bias analyses on demographic characteristics revealed that responders tended to be more stable than non-responders, mirroring findings in previous quit studies. Older, more educated, employed, married, and insured people were more likely to respond than their less stable counterparts. Also, those with an older age of onset of tobacco use were more likely to respond, as were white participants compared to participants of other race / ethnicities.

Another concern for external generalizability is the extent to which the treatment interacts with subject attributes and different settings, as well as the interference of multiple treatments. To the best of my knowledge, subjects did not substantially differ in the setting in which they completed the surveys (via telephone). However, treatment-attribute effects and multiple treatment interference may have

posed substantial threats to external validity. To help control for this, participant demographic and clinical characteristics were controlled for in the logistic regression, along with use of other treatments since enrollment. These independent variables were found to have no significant effect on the outcomes of 7-day abstinence, new quits and relapse.

Taken together, the external validity of the telephone survey appears moderately strong if one seeks to generalize to QUITPLAN participants. Generalizing to other populations requires further study.

*Transferability of Focus Groups.* As an interpretive method, the applicability of focus groups would be judged via criteria for transferability (Lincoln and Guba, 1985). The authors described two key criteria to establish transferability: a thick description of the context for the study conducted, and a thick description of the “receiving” context to which findings would be transferred. The extent to which the focus groups met these criteria is discussed in this section.

Lincoln and Guba (1985) indicated that a criterion of transferability may be description of the study communicated in a database. Indeed, the focus group study included a database enumerating the demographic and clinical characteristics of participants. It is also important to note that the study occurred in the 7-county metropolitan area. However, a thicker description of the context of the study is provided in the body of the focus group report. The report describes that the study occurred two months after the stop-smoking regulations were instituted. The issue of smoking regulations was very politicized, and emotions ran high among supporters and

detractors. As described in the final focus group report, this may have influenced people to respond more vigorously about their support for or disfavor of the regulations. The respondents with the most vocal and strongest opinions about the ban were those who did not support it. The focus group report described how the political opinions of some individuals may have influenced their ability to detect any impact of the regulations that did occur. Additionally, highly political rants distracted focus group members from discussing the subject of the focus groups: the impact of the ban on quitting.

Taken together, the evidence for transferability suggests a moderately strong ability to apply findings to other, similar metropolitan areas where ordinances are hotly contested.

*Transferability of Phenomenological Interviews.* This section discusses the extent to which the interpretive, phenomenological interviews meet criteria for transferability, which in turn indicates the concept of applicability. The criteria are a thick description of both the context in which the study was conducted, and the context of the “receiving” context.

Like the focus groups, the phenomenological interviews were conducted in the 7-county metro area. A database exists that enumerates participants’ demographic and clinical characteristics. However, as compared to the focus groups, the interviews provided much thicker description of the context for each participants. Participants’ specific lived experiences of smoking, bans and quitting are provided. They shared the sights, sounds, smells and tactile sensations. The thick description of the smell of

cigarette smoke was especially vivid, as was the emotional resonance of shame in smoking. Within the phenomenological framework, the specificity and vividness of each individual's story expands the transferability of the study findings because the essential nature of smoke-free ordinances and quitting smoking is better revealed. This thick description indicates a moderately strong ability to transfer findings to similar contexts. Because of the thick description, this ability is greater for interviews as opposed to focus groups.

*Summary of Applicability Across Single Methods.* Within their own frameworks, both telephone surveys and phenomenological interviews provided evidence supporting criteria of applicability. The nature of the evidence was unique for each, but relatively strong. Because it lacked the same thickness of description, the evidence provided by the focus group was somewhat less convincing than for the phenomenological interviews.

### *Consistency*

For Lincoln and Guba (1985), consistency determines the extent to which findings from a study would be repeated if the study were replicated with the same or similar subjects, or in the same or similar context. In the context of a post-positivist study, like the telephone survey conducted for this dissertation, consistency is expressed as reliability. However, for the interpretive focus groups and phenomenological interviews, consistency is expressed as dependability. The extent to which each of these methods met these criteria is discussed in detail below.



*Reliability of Telephone Survey.* Reliability is expressed through coefficient alpha, the proportion of true score variance divided by observed score variance. In the telephone survey, the constructs of greatest concern for reliability are smoking outcomes: 7-day abstinence, new quit and relapse. A large body of literature in tobacco cessation assesses the criterion-related and construct-related validity of these outcomes; most commonly, self-reported tobacco use is compared to the level of cotinine (a by-product of nicotine) in subjects' saliva (SRNT Subcommittee, 2002). A recent meta-analysis concluded that self-reported tobacco use items produce valid inferences about tobacco use for most populations (Patrick, Cheadle, Thompson, Diehr, Koepsell, & Kinne, 1994). Additionally, a large literature examines smoking cessation outcomes at various follow-up periods such as 7-days, 30-days, 1 month, 3 months, 6 months, 12 months and 18 months post-enrollment (Velicer, Prochaska, Rossi, & Snow, 1992). Velicer and Prochaska (2004) recently examined the relationship between 7-day and 30-day abstinence (the two measures used in this study) and found a correlation of  $r = .92$ . Many studies also examine outcomes over shorter periods of time such as 2 days, 4 days, 7 days 10 days and 14 days (Hughes & Gulliver, 1992; Hunt & Bospalec, 1975). However, the purpose of these studies is not to assess reliability, but to track changes in smoking behaviors over short periods of time in order to define the phenomena of slipping, lapsing, and relapse.

The apparent lack of literature that demonstrates test-retest reliability, equivalent forms reliability, or split-half reliability of self-report tobacco use measures is likely due to the fact that outcomes are measured by single items, so split-half

reliability is not a suitable method (Pedhazur & Schmelkin, 1991). Given the malleability of smoking behaviors, assessing reliability through any multiple administrations over time (such as test-retest and equivalent forms reliability) would likely capture changes in smoking patterns and not the reliability of the measure itself. These limitations apply equally to the examination of the reliability of self-reported tobacco use measures in this study. Therefore, test-retest, equivalent forms, or split-half reliability were not conducted for this dissertation. However, several key studies conclude that the inferences drawn from self-report tobacco cessation measures have high levels of criterion- and content-related validity (Velicer & Prochaska, 2004; SRNT Subcommittee, 2002; Velicer, Patrick, Cheadle, Thompson, et al., 1994; Prochaska, Rossi, et al., 1992), suggesting that reliability is at least equally as strong. Therefore, I conclude that the self-reported measures used in this study have high levels of reliability as well.

*Dependability of Focus Groups.* For the interpretive methods, Lincoln and Guba (1985) prescribed that dependability would be assessed first by overlap of methods (triangulation). However, the focus group did not overlap any methods; they comprise one single method. A second strategy for assessing dependability is stepwise replication where teams of researchers analyze the same data. This was not conducted in this study, although the all data are available for multiple analyses. Finally, audits that examine both the inquiry process and accuracy of findings may be used to assess dependability. A rich package of data is available for other researchers to audit, but this has not been done. In short, data is available to conduct dependability analyses, but

none have been conducted. Therefore, the dependability of the focus groups is unknown.

The lack of information about the dependability of focus groups reflects the fact that the data collected for this dissertation was obtained in the context of a real-world evaluation funded by a not-for-profit organization and reflect a real world mixed method evaluation. With constrained funding, the organization was unable to fund multiple external evaluators to conduct stepwise replication and dependability audits. Indeed, in her discussion of generalizability in interpretive methods, Janet Schofield described audits as often “impractical” and replication as “so arduous a task that it is unlikely that high quality researchers could be located” to engage in it (1990, p. 203). Qualitative researcher Ian Dey (1993) acknowledged the difficulty of audits when he writes, “if we cannot expect others to replicate our account, the best we can do is explain how we arrived at our results”. Creswell (1998) acknowledged the labor intensive nature of establishing trustworthiness in qualitative research, and recommends that of the menu of strategies possible, at least two be employed. For the focus groups conducted for this dissertation, I selected the methods of peer debriefing and negative case analysis. For the phenomenological interviews, I selected thick description, fostering respondent-interviewer trust, and negative case analysis. To specifically address dependability, I sought to design both studies in such a way as to allow replication by clearly describing my procedures and processes and providing a very precise and clear audit trail for future researchers that might seek to conduct a dependability audit of it in the future. Had dependability strategies such as stepwise

replication and dependability audits had been conducted, this study would not reflect a real-world mixed method evaluation endeavor.

*Dependability of Phenomenological Interviews.* Unlike focus groups, the phenomenological interviews triangulated to enhance dependability by using multiple sources such as an etymological analysis and literature on addiction and shame to inform conclusions. However, for the reasons specified above, data is available for stepwise replication and a process and accuracy audit of the phenomenological interviews.

*Summary of Consistency Across Single Methods.* Evidence suggests the outcomes measures in the telephone survey are reliable. Unfortunately, the dependability of focus groups and phenomenological interviews is not known due to both the labor intensive and costly task of finding researchers to replicate analyses (Schofield, 1990; Dey, 1993), as well as the financial constraints of the funding agency in supporting reliability analyses. The lack of these analyses reflects real-world mixed method evaluations.

### *Neutrality*

Lincoln and Guba described neutrality as “the degree to which findings of an inquiry are determined by the subjects (respondents) and conditions of the inquiry and not the biases, motivations, interests or perspectives of the inquirer” (p. 290, 1985). In post-positivism, neutrality is expressed as objectivity and is described for the telephone survey below. In contrast, neutrality is described as confirmability in the interpretive focus groups and phenomenological interviews discussed below.

*Objectivity of Telephone Survey.* Evidence of objectivity in the telephone survey is strong. First, the study was conducted by an external telephone survey firm with no vested interest in the outcome of the study. The survey was brief and conducted via telephone six and 18 months post-enrollment, so the relationship of the interviewers with the respondents was quite distant. Second, the survey instrument and procedures were created to assess all components of the study in as value-free manner as possible. In other words, I designed the survey and procedures with the intent of not communicating any of my personal beliefs about the study at any level. The instrument and procedures consist primarily of standard items and protocols used within the field of tobacco cessation. Regulation-related items were adapted from other population-based telephone surveys.

As the Principal Investigator of the telephone survey, my role was as an external evaluator to the funding agency. My expressed role was to assess the impact of the regulations in as objective manner as possible. However, this required acknowledging that the client organization was eager to prove an impact of the regulations. To guard against any impulse to bias findings I adhered to my role as an external evaluator and standardized instruments and procedures. Interestingly, this was accomplished by borrowing from phenomenological methodology and placing myself in a state of reduction, the purpose of which is to observe information (data, results) precisely as they are given and to set aside preconceptions and bias to the best of my ability.

It is possible that knowing this work would contribute to my dissertation would bias the development or administration of the telephone survey study in one direction or another. However, because the dissertation is methodological and does not rely on the substantive findings of the telephone survey itself, I do not believe that this was a threat. Overall, the objectivity of the telephone survey appears moderately strong.

*Confirmability of Focus Groups.* For interpretive methods, such as the focus group examined here, Lincoln and Guba (1985) described confirmability as reflecting the quality of the data in terms of the extent to which it is confirmable. This is tested by providing an audit trail that includes raw data; data reduction and analysis products; data reconstruction and synthesis products, such as categories, findings and a final report; process notes; materials relating to intentions and dispositions; and instrument development information. All but two of these materials are available for the focus groups. The focus group study has available raw data tapes, data reduction and analysis products in NVivo, data reconstruction and synthesis products in NVivo and the final report and draft instruments and instrument guides. However, process notes and materials relating to intentions were not developed for the focus group study.

Once the audit trail is collected, Lincoln and Guba (1985) advocated a five stage process for conducting the audit that includes pre-entry, determination of auditability, formal agreement, determination of trustworthiness and closure. This process has not been instituted for the focus group study. Taken together, I conclude that the confirmability of findings from the focus group data is unknown, but all data are in place to conduct a confirmability audit. As described in the dependability section

for focus groups above, the lack of audits is due to the real-world funding constraints of conducting an evaluation for a not-for-profit agency.

*Confirmability of Phenomenological Interviews.* As described in the focus group section above, the confirmability of findings for the phenomenological interviews depends on the availability of an audit trail and the conduct of a confirmability audit. As with the focus group study, the phenomenological interview study has available raw data tapes, data reduction and analysis products in NVivo, data reconstruction and synthesis products in NVivo and the final report and draft instruments and instrument guides. While the focus group study did not include process notes or materials relating to intentions, the phenomenological interview study does include these items. This is because the phenomenological process demands greater attention to the investigator as the instrument and the iterative nature of analysis.

While the phenomenological study includes all the data necessary for a confirmability audit, no audit has been conducted. Therefore, I conclude that the confirmability of findings from the interview study is unknown. However, one's confidence in the confirmability findings would be stronger for the phenomenological interviews as compared to the focus groups because a more complete audit trail is provided for the interviews.

*Summary of Neutrality Across Survey Methods.* The neutrality of the focus groups appears quite strong. However, the neutrality of both the focus groups and phenomenological interviews is unknown because a confirmability audit has not been

conducted. Given that the audit trail for the phenomenological study is more complete than for the focus groups, it appears as if the conclusions drawn from the phenomenological interview confirmability audit will be more trustworthy than for the focus group confirmability audit.

### *The Joint Committee's Program Evaluation Standards*

*The Program Evaluation Standards* (Joint Committee, 1994) named four major standards by which the quality of evaluations may be judged. They include utility, feasibility, propriety and accuracy, and comprise a total of 30 sub-standards that explicate the content areas of each major standard in greater detail. The standards reflect the applied nature of the evaluation field and are flexible enough to be applied to the multitude of methods evaluators have at their disposal.

The purpose of this section is to briefly assess each evaluation standard enumerated by the Joint Committee for the three single methods and two mixed method studies that were conducted for this dissertation. The extent to which each single method met each standard is briefly summarized below. Four tables (D1, D2, D3, and D4) describing these findings in detail may be found in Appendix D. Finally, the ability of the mixed methods to address these standards is discussed last.

### *Utility Standards*

The utility standards are intended to ensure that the evaluation will serve the information needs of the intended users (Joint Committee, 1994). They comprise seven sub-standards, which include Stakeholder Identification, Evaluator Credibility,



Information Scope and Selection, Values Identification, Report Clarity, Report Timeliness and Dissemination, and Evaluation Impact.

On the first criteria, Stakeholder Identification (U1), all methods explored and met different needs of stakeholders. As a quantitative measure, the telephone survey met the needs of the client's Board of Director and staff for quantitative evaluation findings. The focus group identified and met the need of the client's staff to hear the opinions of a large handful of clients. Finally, the phenomenological interviews met the clients' need to hear in-depth the motivations and experiences of a small number of clients.

The telephone survey, focus group and phenomenological interview methods performed similarly on two criteria, Report Timeliness and Dissemination and Evaluator Impact. For all three methods, interim findings were reported orally to the client before the final report was ready in order to enhance the utility of the report. Also, the client stressed the importance of assessing travel in the study. All three studies were planned and conducted in order to report on travel so that the findings would be of greater use to the client.

On two criteria the survey and focus group were similar to each other and the phenomenological interviews differed from the two. The first instance is on evaluator credibility. I have demonstrated credibility with the tobacco cessation content and the telephone survey and focus group methods. However, I have less experience with phenomenological interviews, which may decrease my credibility for that study. Second, the telephone survey and focus group methods were similar in terms of value

identification. Both clearly explicated the process for interpreting findings, but provided less information on the perspective and rationale that under girded the process. In contrast, the phenomenological interviews clearly explicated these features. Similarly, all methods described the intervention and the purposes, procedures and findings of the evaluation. However, the survey provided little information on the context. The focus group provided some context, but the interviews provided extensive and detailed information in this area.

Taken together, these findings suggest that on a small number of items, utility is similar for the three single methods. It is more common, however, that the interviews differ the most in areas such as providing more context and more perspective and rationale for interpretation, but less evaluator credibility, by a small degree.

### *Feasibility Standards*

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic and frugal. The three sub-standards comprising feasibility include Practical Procedures, Political Viability and Cost-Effectiveness. Each will be discussed below.

All three methods performed equally on two of the three feasibility measures. First, in terms of Practical Procedures for each method, the evaluation measures were practical and resulted in a negligible amount of disruption to the client organization. In terms of Political Viability, the evaluation took into consideration the differing needs of different divisions within the client organization. For example, the Research and Evaluation division was most interested in method quality, the Intervention division

was most interested in gathering information to report on quit rates at 18 months post enrollment, and the Policy division was most interested in accounting for the impact of travel between ban and non-ban communities on outcomes. Each method was planned and implemented with these interests in mind. However, one stakeholder group, activists backed by tobacco lobbyists, was not considered in this evaluation.

Finally, the cost-effectiveness of the three methods differed from one another. Please see page 191 for a detailed treatment of method costs. For the purposes of this research question, however, it is sufficient to report that the telephone survey was by far the most expensive method. However, it was also the kind of information most valued by the client's Board of Directors and staff. In contrast, the focus groups and phenomenological interviews were much more modestly priced, and still contributed very useful information to the client.

### *Propriety Standards*

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved by the evaluation and those affected by its results. Eight sub-standards explicate propriety including Service Organization, Formal Agreements, Rights of Human Subjects, Human Interactions, Complete and Fair Assessment, Disclosure of Findings, Conflict of Interest, and Fiscal Responsibility.

With one exception, the propriety standards were met equally across all three studies. The status of the three studies on each of these standards is described here:

- Formal agreements. In all three methods, a formal contract was signed stipulating responsibilities of each party.
- Rights of human subjects. All methods adhered very closely to the spirit of the IRB. Permission to use secondary data was obtained from the IRB for this dissertation.
- Human interactions. Persons associated with the evaluation were treated with respect and dignity to my maximum capacity in all three methods.
- Complete and fair assessment. To my maximum capacity, I assessed the strengths and weaknesses of the intervention in each study with completeness and fairness.
- Disclosure of findings. Findings of each study have been made public within the client organization through presentations, and to the public through conference presentations. Reports have been provided to all interested parties.
- Conflict of interest. The three studies presented no conflict of interest.
- Fiscal responsibility. Expenditures were accounted for and appropriate in all three methods.

The one propriety standard that the methods differed from each other on was the first, Service Organization (P1). In this standard, evaluations are encouraged to help organizations effectively serve the full range of their participants. Indeed, this was the goal of all three methods. However, because the sample of the telephone survey was so large, that method was able to analyze differences in outcomes between important

subgroups, such as by race, insurance status, geographic location, etc. The focus groups and interviews could not report with such detail on these groups. However, the number of respondents who participated in the focus groups and interviews provided sufficient information to meet the claim that the evaluation is designed to help the client improve service for all its participants.

### *Accuracy Standards*

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. Twelve sub-standards comprise the accuracy standard. They include Program Documentation, Context Analysis, Described Purposes and Procedures, Defensible Information Sources, Valid Information, Reliable Information, Systematic Information, Analysis of Quantitative Information, Analysis of Qualitative Information, Justified Conclusions, Impartial Reporting and Metaevaluation.

Three-quarters of the sub-standards described above are similarly met for all three single methods. All three methods describe the program clearly, and describe the information sources with sufficient detail so that the adequacy of the sources may be assessed. For all three methods, the study was planned and implemented to produce reliable inferences, but the actual reliability of inferences for the three studies is unknown. Additionally, quality assurance processes were used to correct errors in all three methods, although the processes differed according to method. Quantitative data was appropriately and systematically analyzed using SPSS; qualitative data was

analyzed appropriately and systematically using NVivo. In all three studies, conclusions were explicitly justified, although the information used to justify conclusions differed according to method. To ensure impartial reporting in all three methods, I adhered to my role as an external evaluator and to the standards of each specific methodology to the best of my ability. Finally, a metaanalysis of all three methods using the Standards for Program Evaluation is currently being conducted in this section.

The three methods differed on propriety standards in three areas. The first is context analysis. The telephone survey provided basic information about the context only, while the focus groups commented on the politics of the situation. The phenomenological interviews provided the richest data on context for each individual interviewed. The second area of difference is the description of purposes and procedures. The telephone survey described purposes and procedures with sufficient clarity so that they could be assessed. The focus groups and phenomenological interviews did this as well. However, because the analysis of qualitative data is so complex, a full description of study procedures is best revealed with NVivo files, which are available, but not reported in the final report. Finally, the validity of findings differed between the three methods. Truth value was low for the telephone survey, moderately strong for the focus groups, and incrementally stronger for the interviews. Applicability was moderately strong for the telephone survey and phenomenological interviews, and incrementally weaker for the focus groups. Finally, neutrality was

moderately strong for the telephone survey and unknown for the focus groups and phenomenological interviews.

Taken together, this analysis suggests that the single methods are more similar to each other than different in terms of accuracy criteria. The areas where accuracy differs for the three methods are in reporting context, the complexity of describing qualitative research procedures and validity.

### *Mixed Method Studies*

In sum, the examination suggests that the three studies perform quite similarly to each other in terms of propriety, feasibility and most areas of accuracy. The biggest differences between studies are found in terms of reporting context, providing perspective and rationale for interpretations, the complexity of describing qualitative research procedures, validity, and cost-effectiveness. In general, the interpretive methods are more similar to each other than to the post-positivist survey, and the phenomenological interviews differ from the telephone surveys more strongly than the focus groups.

The extent to which the mixed method studies adhere to the Standards of Program evaluation mirror the summary above. This is because the pragmatic mixed method study combines the qualities of the telephone survey and focus group methods, while the dialectic mixed method study combines the qualities of the telephone survey and phenomenological interview methods. Because the telephone survey stays constant in both mixed method studies, the difference between the pragmatic and dialectic

mixed method studies on the Program Evaluation Standards is effectively the difference between the focus group and phenomenological interviews.

To reconsider the findings, the dialectic mixed method study may have less utility than the pragmatic mixed method study because I have less experience with phenomenological interviews. However, the dialectic study may have higher credibility because it contains more information about the perspective and rationale for interpreting findings, and because it provides richer context. Providing richer context may also lead to higher levels of accuracy in the dialectic study as compared to the pragmatic study. Additionally, the dialectic study could have greater accuracy because validity appears to be incrementally stronger for the dialectic versus the pragmatic study in terms of truth value, applicability and possibly neutrality.

#### *Expert Review of Single Methods*

The purpose of this section is to provide further evidence of the validity of the single methods employed in this dissertation. In order to assess the methodological integrity of each method, I developed a 12 to 13 page summary of each methodology, including a description of the problem, the clients' needs, the research question, the rationale for the method choice, my understanding and perspective on the method, and a detailed description of the methodology, which included information about study participants, the sampling plan, logistics and scheduling, recruitment methods, incentives provided, instrumentation, facilitator training, data analysis strategies, and study limitations.



I selected one expert in the field to assess each methodology summary. To assess the focus groups, I selected Dr. Richard Krueger at the University of Minnesota, noted author and expert on focus groups (Krueger, 2000; Krueger & Casey, 2000). For the phenomenological interviews, I recruited Dr. Noriko Ishihara at the recommendation of Dr. Gerri McClland, who teaches the course in Phenomenology in the Department of Curriculum and Instruction at the University of Minnesota. Dr. Ishihara specializes in phenomenological approaches as applied to Japanese language acquisition within the Department of Language and Foreign Studies at American University. To review the telephone survey methodology, I recruited Dr. Michael Luxenberg, President of Professional Data Analysts, Inc., an independent evaluation and statistical consulting firm that specializes in evaluating tobacco cessation interventions (North American Quitline Consortium, 2008).

I requested that the experts review the description of the methodology and comment on strengths and weaknesses as a part of a validity study for my dissertation. The experts wrote one to two page critiques of the methodologies that may be found in Appendix E, although I will excerpt key findings for each method here. Following a description of each expert's comments, I will compare their assessments to my own presented previously in this chapter. This additional source of information fills some important gaps in understanding the validity or trustworthiness of the single methods.

### *Phenomenology*

Dr. Ishihara described the following strengths of my phenomenological methodology as presented to her as follows (personal communication, 2007):

- Description of direct and concrete lived experience of the phenomenon under investigation
- Comparative view gained through the stratified sampling (i.e., full, partial and no ordinance conditions, and current and former smokers)
- Anonymity enhanced through the nature of the phone interview and potentially less inhibition on the participants' part in describing their experiences (this point can greatly enhance the credibility of the data.)
- The sense of trust, warmth, and safety established through the researcher's conscious effort (e.g., verbalizing non-verbal cues)
- Enhanced credibility of research through the field-testing and refining the selection of interview questions
- Enhanced credibility of the research through employment of *reduction* (i.e., suspending the researcher bias and preconceptions)

Dr. Ishihara also commented on potential methodological weaknesses, as

follows (personal communication, 2007):

- Potential of not capturing the full range of the essence of the phenomenon due to stratified sampling (two participants in each category in the interview sampling may not be sufficient to demonstrate variant and invariant essences.)
- Less direct means of communication through the telephone (However, benefits of the interviews discussed above appear to outweigh drawbacks.)
- Limitation of human consciousness, participants' limitations in ability to articulate only through verbal channel in a concrete detailed manner (The assumptions listed on p. 6 seem optimistic, considering the demand of the phenomenological research methods.)
- Another point that I am unable to assess from the given information is the effectiveness of the qualitative software program (NVivo) employed for the 'identification and transformation of meaning units' (p. 6). The way in which the program transformed the meaning units remain unclear and warrants further information.

The strengths and weaknesses of the methodology described above provide evidence to confirm or challenge my opinions on the credibility, applicability, dependability and neutrality of the phenomenological interview method. They also provide evidence to confirm or challenge my assessments of the phenomenological

interviews in meeting the *Standards for Program Evaluation*. I will discuss the intersections of my assessments with those of Dr. Ishihara below.

First, I described that the negative case analysis conducted strengthened the credibility of inferences made from the phenomenological interviews. Dr. Ishihara echoed this point by identifying the strength of developing a comparative view between those living in different ordinance conditions and with different smoking statuses. At the same time, Dr. Ishihara issued two broad cautions as to the credibility of inferences drawn from my study design. She pointed out that two interview subjects in each condition (ordinance condition by smoking status) may be insufficient to capture the full range of the essence of ordinances. She also points out that limitations in human consciousness and limitations in participants ability to articulate their experience in a concrete and detailed manner may threaten the credibility of findings.

Several of Dr. Ishihara's comments reflect on the transferability of findings. She describes four key strengths of the study that include (1) the description of direct and concrete lived experience, (2) the anonymity of telephone interviews to enhance credibility, (3) the sense of trust, warmth and safety established between the researcher and participant, and (4) the employment of reduction to suspend researcher bias and preconception. These four qualities combine to produce thick description, rich detail and critical context that support the transferability of findings to other, similar contexts. Dr. Ishihara's assessment concurs with mine on this point. Additionally, these four qualities also support three other of my assessments. These include neutrality, where the thick description and rich detail produced contribute to findings that reflect

participant experience and not researcher bias. The four qualities also support the program evaluation standard of Context Analysis, where phenomenological interview excel at providing detailed and expressive information. Finally, the four qualities support the program evaluation standard of Evaluator Credibility. Although I report that my experience with phenomenological interviews is less than with focus groups or telephone surveys, Dr. Ishihara assesses my understanding of phenomenological methods in these four areas as strengths of the methodology.

In terms of dependability, Dr. Ishihara posed no critical concerns with the conduct of the interviews. This suggests that a researcher who conducted the focus groups similarly would achieve similar results. Dr. Ishihara's review of the interview protocol also suggests that it was not biased or influenced by preconceived notions, and would be robust to considerations of neutrality.

An important point for clarification for Dr. Ishihara is with the use of the NVivo program in transforming meaning units, one of the analysis strategies I employed. She requests further clarification. This comment reflects on the program evaluation standard Analysis of Qualitative Information. I report that I analyzed qualitative data appropriately, while Dr. Ishihara requests additional information to make this judgment. Dr. Ishihara's comments suggest that she is skeptical that a software program could conduct the complex analysis tasks sufficiently for the demands of phenomenological inquiry. In this concern she is correct. The software program was used as method for *the researcher* to create, manipulate and examine categorizations, or meaning units. Had this been clarified in the methodology summary

I provided to Dr. Ishihara, it is possible that she would not have expressed a concern in this area.

Dr. Ishihara confirms my assessment that the standard of Report Clarity was met. I asserted that the intervention purposes and procedures were clearly enough described to make an assessment, and (with the exception of my description of NVivo discussed above) Dr. Ishihara's forthright assessment of those purposes and procedures confirms this claim. Additionally, Dr. Ishihara also tacitly confirmed that the program processes and procedures were documented sufficiently to be assessed (excepting the description of NVivo above, standard A3), and Dr. Ishihara's discussion of sampling confirmed that the sources of information were sufficiently described to assess their adequacy (standard A4).

### *Focus Groups*

Dr. Krueger reviewed the description of focus group methodology and endorsed it overall with the statement, "The process you describe in your proposal makes sense .... The study looks reasonable and should yield helpful insights" (personal communication, 2007). He goes on to highlight four areas that he sees in need of improvement.

The first area of concern is the sampling plan. He recommended reconsidering the sampling plan from a crossed design of ordinance condition and smoking status because travel between ordinance and non-ordinance counties muddies the geographic distinction. He recommended instead separating them by how frequently they go to restaurants, characteristics of their smoking, or other factors. Second, Krueger

recommended having participants complete registration forms to collect demographic and other relevant data. Third, Krueger recommended tracking comments by person for coding purposes. Finally, Krueger made several recommendations regarding the ordering of focus group questions. Specifically, he recommended beginning by asking about their current quit status and intentions to quit now and in the past, and asking what factors contributed to that change. He also suggested asking about the key ingredients of the intervention program, and not the program by name. Further, he suggested asking for a numeric score to indicate the strength of various factors in influencing changes in their quitting / smoking status.

Several of Dr. Krueger's suggestions were implemented. For example, he recommended the use of registration forms, however, demographic and clinical information was available from participants based on the intake, 6 month follow-up and 18 month follow-up survey responses. Second, Dr. Krueger recommended against sampling by geography, but this was done in part to ease the burden of driving to the focus group on participants. We did devote a substantial portion of focus group questions to travel, which was not a common phenomenon among participants. Finally, Dr. Krueger's recommendations for the introduction were implemented based on experience with the first focus group.

Dr. Krueger's assessment of the focus group methodology provides evidence to confirm or challenge my opinions on the credibility, applicability, dependability and neutrality of the phenomenological interview method. They also provide evidence to confirm or challenge my assessments of the phenomenological interviews in meeting

the *Standards for Program Evaluation*. I will discuss the intersections of my assessments with those of Dr. Krueger below.

First, in this dissertation, I conclude that the credibility of the focus groups is moderate because of the use of peer debriefing, negative case analysis and referential adequacy. Dr. Krueger provided no recommendations for improvements in these areas. However, he provided several suggestions that may be seen to improve credibility of findings broadly. In recommending an alternate sampling on bar and restaurant use, he provided guidance so that focus group findings may be more meaningful.

In terms of dependability, Dr. Krueger's approval of the focus group processes and protocols suggest that a researcher who conducted the focus groups similarly would achieve similar results. His review of the protocol also suggests that it was not biased or influenced by preconceived notions, and would be robust to considerations of neutrality.

In the *Program Evaluation Standards* accuracy standard called Analysis of Qualitative Information, I assess my analysis to be appropriate. However, Dr. Krueger recommended identifying the comments of individuals. This is an excellent suggestion that was not implemented due to resource concerns, and because sufficiently detailed notes were not available. This would have indeed improved the analysis of data.

Dr. Krueger's assessment overall supports my assertion that I have established credibility conducting focus groups (Evaluator Credibility). Dr. Krueger also confirms my assessment that the standard of Report Clarity was met. I asserted that the intervention purposes and procedures were clearly enough described to make an

assessment, and Dr. Krueger's forthright assessment of those purposes and procedures confirms this claim. Additionally, Dr. Krueger also tacitly confirmed that the program processes and procedures were documented sufficiently to be assessed (standard A3), and Dr. Krueger's discussion of sampling confirmed that the sources of information were sufficiently described to assess their adequacy (standard A4).

#### *Telephone Interview*

Dr. Luxenberg reviewed the telephone interview methodology statement and commented on several strengths and limitations of the study. He also disclosed that he is President of Professional Data Analysts, Inc. who employs me and received the contract for the evaluation. He disclosed that he had some input in the design, but that I was the principal investigator for this study.

#### Strengths

- Large sample sizes
- Well thought out sampling plan
- Thoroughly designed survey instrument
- Carefully crafted recruiting procedures
- Strong consent procedures and training to implement them
- Reasonably high response rates
- Development of the variable "exposure to ordinances"



## Limitations

- Sample includes self-selected population of smokers seeking to quit using evidence-based intervention
- Subjects were not randomized to a particular ban condition
- Some response bias on demographic and clinical characteristics
- Confounding of ban type (full, partial, none) with geography (Hennepin and Ramsey Counties, 5 county suburban metropolitan area, outstate Minnesota)
- Unanticipated changes in the ordinance conditions in Hennepin County
- Diffusion of intervention for border communities with different ordinance conditions within and between states

The strengths and limitations as described above confirm my assessment of weak truth value (internal validity) but relatively strong applicability (external validity). Dr. Luxenberg did not comment on the consistency (reliability) of measures, other than to state that the instrument was thoroughly designed, which includes using measures from validated instruments in the field of tobacco cessation. Factors influencing the neutrality of findings were not discussed.

In terms of the *Standards for Program Evaluation*, Dr. Luxenberg's assessment overall supports my assertion that I have established credibility conducting telephone surveys (Evaluator Credibility). My assessment of the Rights of Human Subjects is

also confirmed as Dr. Luxenberg affirmed the strong consent procedures that were employed.

Dr. Luxenberg confirms my assessment that the standard of Report Clarity was met. I asserted that the intervention purposes and procedures were clearly enough described to make an assessment, and Dr. Luxenberg's assessment of those purposes and procedures confirms this claim. Additionally, Dr. Luxenberg also tacitly confirmed that the program processes and procedures were documented sufficiently to be assessed (standard A3), and Dr. Luxenberg's discussion of sampling confirmed that the sources of information were sufficiently described to assess their adequacy (standard A4).

Research Question 4: How do stakeholders view the credibility and utility of single method findings and mixed method findings?

The purpose of this substudy is to understand stakeholders' perceptions of single and mixed method findings in three areas: respondent experience with the three methods used in the evaluation, respondent opinion on the credibility of single method findings, and respondent report on the credibility and utility of two mixed method summary of findings.

Three stakeholders who were critical to the development and use of the evaluation were recruited to participate in the paper and pencil study. Two completed the survey, the Senior Research Program Manager who managed the evaluation grant, and the Senior Marketing Manager, who requested additional analyses from the studies to inform marketing efforts. The Public Policy Manager did not complete the survey

because her position was eliminated and she left the organization. She was not able to be contacted to complete the survey. Findings from the two respondents are summarized in the tables and narrative below.

### *Respondent Experience with Single Methods*

Both stakeholders reported the most experience with focus groups (see Table 10 below). One reported using focus groups in her job, and the other reported a wide range of experience: understanding the basics of the method, using them in her job, designing or administering them, and overseeing and providing technical assistance for research projects involving focus groups. The stakeholders differed in their experience of surveys. One reported understanding the basics of the method only, while the other had all levels experience, including managing telephone survey projects. In sharp contrast, both stakeholders reported that they had little or no experience with phenomenological interviewing. In summary, respondents have a moderate amount of experience with focus groups and telephone surveys, but little or no experience with interviews.

Table 10

*Stakeholders' Previous Experience with Methods*

Experience Level	Survey	Focus Group	Interview
Little / no experience	0	0	2
Understand basic method	2	1	0
Used in my job	1	2	0
Designed or administered	1	1	0
Managed a project	1	0	0
Other	0	1	1

Table 11

*"Other" Comments on Experience with Methods by Method*

Method	Comment
Focus Groups	Overseen and provided TA for research projects involving focus groups
Interviews	I did interviews for my dissertation research but not using a phenomenological approach

*Perceptions of Credibility of Single Methods*

Respondents were asked to consider each single method and rate how rigorous the method is overall; the meaningfulness and trustworthiness of findings; and the credibility of findings to oneself and their own stakeholders. The rating scale ranged from 1 to 10, and each rating number was labeled. Ratings and labels can be found in Table 12 below.

Table 12

*Rating Points and Labels for Stakeholder Survey*

Rating Number(s)	Rating Label
1	Not at all
2, 3, 4	Not very
5, 6, 7	Somewhat
8, 9	Very
10	Completely

*Perceptions of Methodological Rigor*

Stakeholders rated both phenomenological interviewing and surveys to be somewhat rigorous. Each method received a rating of a 6 and 7 (see Table 13). Focus groups had slightly higher ratings because one stakeholder rated focus groups as being very rigorous (8).

Table 13

*Stakeholder Ratings of the Validity of Findings by Method and Respondent*

	Telephone		Focus Group		Interview	
	Resp 1	Resp 2	Resp 1	Resp 2	Resp 1	Resp 2
Rigorousness	7	6	8	7	7	6
Meaningfulness	8	8	8	7	8	7
Trustworthiness	7	8	8	7	8	6
Credibility to you	7	8	8	7	8	8
Credibility to	7	9	7	8	8	9

stakeholders

### *Perceptions of the Meaningfulness and Trustworthiness of Findings*

Both respondents rated the phenomenological interviews as being very meaningful, the highest rating of the three methods by a very small margin (ratings of 8 and 8, see Table 13). Ratings for both focus groups and surveys were very slightly lower (8 and 7 for each).

The two respondents were somewhat mixed in their opinions of the trustworthiness of single methods. One respondent rated interviews as very trustworthy (8), followed by focus groups (7) and surveys (6). The other respondent rated focus groups and phone surveys as very trustworthy (8), while interviews were rated as somewhat trustworthy (7). Overall, the respondents rated all three methods as a mix of being very and somewhat trustworthy.

In summary, respondents appear to have more confidence in the meaningfulness and trustworthiness of methods as compared to their rigor. This is because at least one respondent endorsed that each method was both very meaningful and very trustworthy. However, perceptions remain mixed, because in most cases one respondent reported each method to only be somewhat meaningful or trustworthy.

### *Credibility of Findings to Oneself and One's Stakeholders*

Respondents rated telephone surveys as being very credible to themselves (a rating of 8 for each). This was higher by a small margin than the ratings for focus

groups and interviews. Both focus groups and interviews had a rating of 7 (somewhat credible) and 8 (very credible) each.

Credibility to stakeholders yielded the only scores higher than 8, and telephone surveys were rated the highest (8 and 9). Findings about phenomenological interviews were mixed. For one respondent, interviews are very credible to stakeholders (9), but for the other they are only somewhat credible (7). This is similar to focus groups, which were rated as somewhat (7) and very (8) credible to stakeholders.

Taken together, the findings on credibility reveal that respondents rate telephone surveys to be very credible to themselves to their stakeholders. Credibility for oneself and stakeholders is mixed for focus groups and interviews (one rating of very and somewhat for each respondent to each item). Interesting, one respondent rates phenomenological interviews as very credible (9) to stakeholders.

#### *Perceptions of Credibility of Mixed Methods*

Summaries of findings for survey and focus group methods (method mix 1) and survey and interview methods (method mix 2) were created and distributed with the survey. Respondents were asked to read the mixes and rate them on several dimensions. The rating scale mirrors that used to rate single methods (see Table 12). The items and their ratings are enumerated in Table 14.

Table 14

*Stakeholder Ratings of Method Mixes*

	Method Mix 1: Survey & Focus Group		Method Mix 2: Survey & Interview	
	Resp 1	Resp 2	Resp 1	Resp 2
Reflected participants' perspectives	7	8	8	8
Describe and account for changes in the research setting	6	3	6	3
Reflect a neutral point of view on the part of the evaluator	6	7	7	7
Could be confirmed or corroborated by others	5	8	5	5
Could be generalized to groups of people beyond the study sample	4	8	4	5
Would be the same if subjects were reassessed, excluding changes due to the passage of time	DK	7	DK	7
Reflect the underlying truth of participants' experience	6	5	6	7
Are the result of the ordinance and not other factors	7	5	7	5



Looking across the two respondents, most ratings hovered in the range of 5, 6 and 7. Therefore, respondents seem to be somewhat confident in the findings of method mixes overall. A small number of results diverge from this pattern. First, the only time both respondents were very confident was that findings from Method Mix 2 reflected participants' experience (8, 8). This was the slightly higher than the ratings for Method Mix 1 (7 and 8). Second, respondents reported being not very confident in findings from the method mixes in two areas. One respondent was not very confident that findings from either method mix described and accounted for changes in the research setting (ratings of 6 and 3 for each method mix). Also respondents seemed less confident that findings from Method Mix 2 could be generalized to groups of people beyond the study sample (4, 5). The rating for Method Mix 1 was higher, but still mixed (4, 8).

Respondents reported being somewhat confident in findings from both method mixes on the remaining items. The content areas include that findings reflect a neutral point of view on the part of the evaluator, could be confirmed or corroborated by others, would be the same if subjects were reassessed, reflect the underlying truth of participants' experiences, and are the result of the ordinance and not other factors.

#### *Reported Utility of Mixed Methods*

Respondents were asked if the method mixes would influence their organization's approach or activities. For Method Mix 1, both respondents reported yes. Both responded yes for Method Mix 2 as well. For each method mix, respondents were asked what changes did or might occur, and the findings or evaluation processes

that did affect those changes (see Table 15 for verbatim comments). Finally respondents were asked to share any additional comments.

Table 15

*Comments on the Impact and Potential Impact of the Studies and the Findings that Lead to them, by Method Mix*

Method Mix	Changes that did or might occur	Findings that lead to those changes
1	This information could be used to create new messaging in a marketing campaign for QUITPLAN services	Information from the findings would be used in a create brief or platform that is used by ad agencies to develop new ads and media decisions
1	Advertising to become more sensitive to how smokers experience the ordinance, wanting to fit in, showing how ordinances can help with quitting	Focus on norms, highlighting the importance of impact of social norms
2	This information could be used to create new messaging in a marketing campaign for QUITPLAN services	Information from the findings would be used in a create brief or platform that is used by ad agencies to develop new ads and media decisions
2	Shift in advertising messages to hone in on smokers <u>experience</u> with ordinances	Focus on shame and smell

Table 16

*General Comments on Mixed Methods' Use*

<i>N</i>	<i>Comment</i>
1	From a marketing perspective both methods provide the same outcomes and learnings. Either mix is helpful in learning from the experiences of smokers and non-smokers and what kinds of messages may help motivate them to quit smoking or stay quit.

All comments pertained to the marketing efforts of the organization. One respondent reported that Method Mix 1 caused “advertising to become more sensitive to how smokers experience the ordinance,” such as smokers’ desire to fit in. The information that affected this change was the discussion of the relationship of ordinances to social norms and their impact on social norms. For Method Mix 2, the same respondent reported that the study would “shift advertising messages to hone in on smokers’ experience with ordinances”. In Method Mix 2, the discussion of shame and smell lead to this change.

The other respondent reported that from her perspective in marketing, “both methods provide the same outcome and learnings” and that both mixes are helpful in revealing smokers’ and non-smokers’ experiences. Therefore, her comments on the utility of Method Mix 1 and Method Mix 2 are identical. For both she reported that the mix could be used to “create new messaging in a marketing campaign”. Specifically, the method mix findings would be used “in a creative brief or platform that is used by ad agencies to develop new ads and media decisions”. The respondent did not share any specific findings that would be used in this way.

Research Question 5: What are the cost of the  
single methods in terms of billable dollars and subject hours?

The purpose of this section is to understand the costs of the single methods and how they compare to each other. In order to accomplish these goals, this section has two parts. First, costs of each method in billable researcher dollars will be computed. Next, costs are examined in terms of subject hours expended.

*Cost in Billable Researcher Dollars*

Billable research dollars were computed based on an arbitrary rate of \$100 per hour for lead staff and \$50 per hour for support staff. Table 17 illustrates the total cost for all three single methods was \$178,756. The 18 month follow-up survey was the most cost intensive, totaling \$99,814, over half (55.8%) of the total budget for all three single methods. The study with the next highest cost is the focus groups (\$47,475), which comprised about one quarter of the total budget (26.6%). The least expensive method was the phenomenological interviews. They totaled just \$31,467 and 17.6% of the total budget. Additionally, cost per respondent in billable researcher dollars was calculated by dividing budget by number of respondents. The average cost per respondent across single methods was \$244.54. Surveys yield the lowest cost per respondent at \$154.51, followed by focus groups (\$678.21). Phenomenological interviews yield the greatest cost per respondent (\$2,097.81).

Table 17

*Total Billable Researcher Hours Expended by Method*

Method	\$	%	# Respondents	Billable \$ per Respondent
Focus Groups	47,474.88	26.6	70	\$678.21
Interviews	31,467.28	17.6	15	\$2,097.81
Surveys	99,814.15	55.8	646	\$154.51
Total	178,756.31	100.0	731	\$244.54

Table 18 itemizes each single method budget by primary task. An analysis of billing records revealed six major tasks: planning, implementation, database management, analysis and reporting. Expenses are a final category. When the three single methods are combined together, the greatest proportion of billable dollars, over three in ten (30.7%), falls within the category of implementation. Approximately similar amounts of billable researcher dollars are distributed between the remaining five tasks and expenses.

When the costs of tasks are compared across single methods, differences may be seen. For example, surveys generate the greatest proportion of billable dollars for database management (22.6%), which is negligible for focus groups and interviews (1.5% and 2.2%, respectively). Surveys also have the lowest proportion of billable dollars for allocated for reporting (8.6%), as compared to focus groups (21.1%) and interviews (29.2%). Costs for interviews show a different pattern. Interviews required the least planning (7.8%) and expenses (6.3%), but the highest combined analyses and

reporting budget (57.0%). In contrast, focus groups spent the greatest proportion of their costs on implementation (39.9%).

Table 18

*Dollars Worked and Percent of Budget for Study Tasks by Method*

	\$	%
<b>Planning</b>		
Focus Groups	6,825.00	14.4
Interviews	2,450.00	7.8
Surveys	14,912.50	14.9
Subtotal	24,187.50	13.5
<b>Implementation</b>		
Focus Groups	23,450	39.9
Interviews	8,425	12.5
Surveys	23,013.00	23.1
Subtotal	54,888	30.7
<b>Database</b>		
Focus Groups	700.00	1.5
Interviews	700.00	2.2
Surveys	22,587.50	22.6
Subtotal	23,987.50	13.4
<b>Analyses</b>		
Focus Groups	1,800.00	3.8
Interviews	8,750.00	27.8
Surveys	19,350.00	19.4
Subtotal	29,900.00	16.7
<b>Reporting</b>		
Focus Groups	10,037.50	21.1
Interviews	9,175.00	29.2

Surveys	8,612.50	8.6
Subtotal	27,825.00	15.6
Expenses		
Focus Groups	4,662.38	9.8
Interviews	1,967.28	6.3
Surveys	11,338.63	11.4
Subtotal	17,968.31	10.0
Total		
Focus Groups	47,474.88	100.0
Interviews	31,467.28	100.0
Surveys	99,814.15	100.0
Grand Total	178,756.31	100.0

#### *Cost in Subject Hours*

This section examines the cost of the three individual studies in terms of subject hours. The surveys used the greatest number of respondents ( $N=646$ ), but had the shortest time with each (6.5 minutes, see Table 19). Conversely, interviews had the fewest number of subjects ( $N=15$ ), and spoke with each for an average of 70 minutes. In focus groups, 70 subjects spent a total of 120 minutes each. The comparison of focus groups to interviews is somewhat misleading, however. Each subject in the interviews conversed one-on-one for 70 minutes, while each participant in the focus groups did not converse directly for 120 minutes because the groups consisted of 5.4 people on average. Assuming that focus group subjects spoke one at a time and participated equally in a 120 minute focus group, each participant would have spoken for about 22.2 minutes each. This translates to 25.9 hours of direct subject time.

Table 19

*Cost in Subject Hours for Recruitment and Completion*

Method	N of Subjects	Minutes per Subject	Total Minutes	Total Hours
Survey	646	6.5	4,199	70.0
Focus Groups Total Time	70	120.0	8,400	140.0
Focus Groups Direct Time <sup>1</sup>	70	22.2	1,554	25.9
Interviews	15	70.0	1,050	17.5

<sup>1</sup> An average of 5.4 subjects participated in each focus group.

Because focus group participants were asked to listen, reflect and speak during the entire 120 minutes of the group, the total time for focus groups will be used. It is instructive to compare the total subject hours expended across methods. Focus groups were most intensive in subject hours by far. They used 140.0 hours, twice the number as surveys (70.0 hours) and four times the number as interviews (17.5 hours). All three studies expended a total of 227.5 hours of subjects' time.



## CHAPTER FIVE

## CONCLUSIONS AND DISCUSSION

The intentional use of mixed methods and research on the methodology has only recently coalesced into a field of study. Use of mixed methods has been most prominent in applied fields such as evaluation (Greene & Caracelli, 1997; Greene, 2007), health sciences (O’Cathain, 2009), and more recently, education (Day, Sammons, & Gu, 2008), as researchers seek to meet the information needs of a stakeholders with diverse opinions about the credibility of different evidence types. Over the past three decades, a greater acceptance of qualitative methods has emerged and resulted in a recognition of the value of using diverse research strategies to answer research questions and guide decision making. Mixed methods has become a valuable tool in this context.

Explication of the methodology of mixed methods has primarily occurred in academic journals concerned with evaluation, especially *New Directions in Evaluation* and *The American Journal of Evaluation*. However, the last five years has seen an explosion of scholarship on the topic. Tashakorri and Teddlie published their handbook on mixed methods in 2003, followed by Creswell and Plano Clark (2007) and Greene (2007). Also in 2007, the first issue of the *Journal for Mixed Methods Research* was published.

At the same, time, the field is struggling to identify a philosophic framework to guide it. While post-positivism and interpretivism draw from rich traditions of theory

and practice, the practice of mixed methods has yet to develop a coherent theory driving it, despite claims that mixed methods is the third research paradigm (Johnson & Onwuegbuzie, 2004). The primary theoretical approach to mixed methods has been the dialectic approach (Greene & Caracelli, 1997), which seeks to gain insight by juxtaposing methods conducted using clearly defined and diverse research paradigms. However, the empirical literature on the use of a dialectic approach to mixed methods is sparse (Riggin, 1997; Creswell, Trout, & Barbuto, 2002; Patton, 1985) and the approach appears alternately poorly understood or misunderstood in the literature (Stufflebeam, 2001; Mark, Feller, & Button, 1997) and infrequently used (Riggin, 1997). An evolving and growing body of work is focusing on a pragmatic approach to mixed methods which challenges the primacy of Greene and Caracelli's dialectic stance. The lack of coherent philosophical framework for mixed methods has left the field straddling the domains of quantitative research methodology, qualitative research methodology, evaluation, and the content of fields such as health sciences and education. A coherent theory for mixed methods research is a necessary step in the growth of the field.

The value of this study is that it explores the two competing theories of mixed methods currently in the field (pragmatic and dialectic), in an effort to guide further theoretical development of mixed methods methodology. It is hoped that an empirical comparison of the two theories based on a real-world evaluation would assist researchers and evaluators to optimize their use of mixed methods. In this quest, this study also probes underlying assumptions and rationales of mixed methods: that mixed

method studies yield findings over and above single methods presented side-by-side, that the paradigmatic divergence of methods is a critical factor in mixed method studies, that mixed method studies can better meet the demands of multiple stakeholders, and to examine the increased financial demands of mixed method studies. In the following paragraphs each research question is stated and answered; conclusions about conducting mixed method research studies are presented and limitations of the study are discussed. Finally, I suggest recommendations based on this research and implications for evaluation and mixed method inquiry.

#### Research Questions Answered

Five research questions were developed and results were presented in detail in Chapter 4. This section will restate each question and briefly answer it.

*Research Question 1: What are the substantive findings of single methods?*

*What findings converge and diverge? What findings are unique?*

#### *Substantive Findings of the Telephone Survey*

Telephone survey analyses showed a strong bivariate relationship between exposure to smoke-free ordinances and relapse. Controlling for key factors, logistic regression results suggest that exposure to the ordinance was marginally associated with relapse ( $p = .061$ ). However, exposure to the ban was not significantly associated with either 7-day abstinence or having made a new quit. Findings show that travel

between communities with and without smoke-free ordinances was infrequent and the most common reason for traveling was for reasons other than to go to establishments that allow or prohibit smoking. Travel did not impact outcomes in the logistic regression, but bivariate comparisons indicated that those in ordinance communities were more likely to travel to bars and restaurants that allow smoking.

### *Substantive Findings of the Focus Groups*

Focus group findings suggest that smoke-free ordinances helped make quit attempts easier, helped respondents reduce their smoking, and assisted former smokers to maintain their quits. At the same time, some smokers and former smokers reported being unaffected by bans, sometimes because of a lack of exposure to them. Other reasons for lack of impact were identified. Finally, ordinances also made a small number of smokers feel defiant and angry that their rights were being impinged upon. The focus groups provided valuable examples of how each of these effects occurred that substantiated findings. Focus group findings also suggest that ordinances appeared about equally likely to either have no effect on travel, or to have an effect dependent on smoking status. An unanticipated finding about social norms emerged. Almost all respondents felt that it was socially unacceptable to smoke and expose others to smoke, which appeared to be a powerful motivator for smokers to quit and consider quitting. While subtle and not immediate, social norms against smoking appeared to be the most promising mechanism to motivate quitting and prevent the initiation of smoking in the first place.

*Substantive Findings of the Phenomenological Interviews*

Almost all participants experienced some impact of smoke-free regulations. The most common experience was as an aid to tobacco cessation or reduction. About one-third of participants reported that ordinances have no impact, which appears at least partially related to a lack of exposure to the ordinances. Finally, ordinances also made a small number of smokers feel defiant and angry that their rights were being impinged upon. The phenomenological interviewed provided thick description and rich detail of respondents' lived experience that uncovered fundamental mechanisms indicating a relationship between ordinances and smoking that stem from being forced them to face their cravings and addictions. For some, experiencing ordinances resulted in fresh insight on their addiction and a new motivation to quit. Others continued to struggle with their addiction without change or become angry at being regulated.

Almost all smokers and non-smokers initiated a discussion of the smell of smoke and hated it in the strongest terms. Respondents associated it with being dirty and with shame of smoking and addiction. Public relations and communications should be sensitive to the issue of shame for smokers regarding smoke-free ordinances. The smell was both a motivator to quit smoking, and a strategy to maintain a quit. Everyone who reacted negatively to the smell of cigarette smoke received relief from it through the ordinances, regardless if they acknowledged the benefit or not.

*Convergence, Divergence and Uniqueness of Findings*

Recall that findings from all three methods were combined, and 88 substantive findings emerged. Findings were categorized so that similar ideas between methods could be identified and convergence, divergence and uniqueness of findings across methods could be assessed. The results indicate that findings most frequently either converged between two or three findings (43%), or were unique to a single finding (43%). It was relatively uncommon for findings to diverge by degree (8%) and even more uncommon for them to diverge outright (1%). A small number of findings both converged and diverged (4%; in this case two methods converged and one diverged or vice versa).

Findings were categorized into 17 topic areas ranging from external generalizability to the impact of exposure to the ordinance on relapse to travel to social norms. Do certain topic areas result in more convergence or divergence? The answer is yes. On the primary question of interest, the impact of ordinances on quitting outcomes, a mixed pattern of divergence was found among all three single methods. On the topic of travel, the three methods primarily converged, although some divergence emerged. However, the greatest convergence was seen between focus groups and interviews on the four topics not addressed by the telephone survey (the impact of exposure to ordinances on smoking less and the bar and restaurant experience, respondent anger over ordinances, the relationship between social norms and the ordinance, and the smell of smoke)..

How do combinations of methods compare to one another in terms of convergence, divergence and uniqueness of findings? To answer this question, the convergence status for findings was tallied for each method combination. The focus group and interview combination had the greatest number of findings in relationship with each other ( $N=30$ ), most of which converged ( $N=26$ ). The next greatest number of common findings were found in the survey and focus group combination ( $N=20$ , where most findings (75%) were convergent and one quarter were divergent or divergent by degree. The survey and interview methods had slightly fewer findings ( $N=14$ ), but a greater proportion of them were divergent (36%) as compared to the survey-focus group pairing. Findings unique to focus groups ( $N=17$ ) and interviews ( $N=15$ ) also garnered a high number of findings. Surveys had only 6 unique findings.

*Research Question 2: What are the substantive findings of pragmatic versus dialectic mixed method studies? How are the two mixed method study findings similar and different from another? What unique information do the mixed method findings produce over and above single methods?*

Pragmatic and dialectic frameworks were utilized to integrate the single methods into two mixed methods studies. The integration of the telephone survey and focus group methods in the pragmatic mixed method study was largely straightforward because most findings were convergent. However, findings on the topic of travel diverged and pragmatic criteria were used to integrate the findings. Examining divergent findings based on experience, knowledge and context brought important

information to the table that helped to resolve the divergence. The resolution was a unique conclusion because the information the pragmatic criteria elicited were brought to bear. Based on the integration of findings, a pragmatic narrative was produced.

The integration of the telephone survey with the phenomenological interviews in the dialectic approach entailed greater complexity. While several simple convergences were found where examples from interviews related in a one-to-one manner with survey findings, more complex convergences also emerged. The interviews illuminated underlying mechanisms for observed relationships that related to the telephone survey findings in a resonant and web-like manner. The mechanisms were unique findings produced from the phenomenological interviews. One substantial and major divergence in telephone survey and phenomenological interview findings was identified regarding the impact of ordinances on abstinence, and was resolved using an Hegelian-inspired dialectic approach. The quest for an overarching dialectic truth created a puzzle that forced me to be more creative in my approach to resolve both the content and methods being combined. The solution produced was consistent with both survey and interview findings and was generative in the spirit of the dialectic approach.

When compared, the dialectic mixed method study contained more divergent findings and more complex convergence. The pragmatic and dialectic mixed method findings did not produce any unique findings over and above the single method studies. However, they did produce unique conclusions and interpretations based on the single methods and the pragmatic or dialectic framework that was employed to mix them. As



described above, both the pragmatic and dialectic studies resulted in converged that provided compelling strength to conclusions. Additionally, both mixed method studies also produced conclusions of greater complexity and nuance than single methods. This was due to at least a degree of divergence between methods in both mixed method studies.

*Research Question 3: To what extent do inferences drawn from single method findings meet key criteria for validity / trustworthiness? To what extent are conclusions drawn from single method and mixed method findings valid / trustworthy according to The Program Evaluation Standards (Joint Committee, 1994)?*

*The Extent to which Inferences Drawn from Single Methods Meet Key Criteria for Validity / Trustworthiness*

Recall that four criteria for validity / trustworthiness were selected and I argued for my opinion of the extent to which the inferences drawn from the telephone survey, focus groups and phenomenological interviews met the criteria. Overall, each method demonstrated moderate to strong “truth value” and applicability, with one exception. The “truth value” of the telephone survey was determined to be low due to the lack of random assignment. It is also important to note that the focus groups were determined to have more moderate levels of “truth value” and applicability as compared to interviews, which provided stronger evidence. In terms of neutrality, the telephone interviews provided evidence that inferences were strongly objective. However, neutrality was unknown for the focus groups and interviews due to both the labor

intensive and costly task of finding researchers to replicate analyses (Schofield, 1990; Dey, 1993) and the financial constraints of the not-for-profit agency funding this research. Had these dependability tasks been conducted, the study would not reflect a real-world evaluation. A final criterion for validity / trustworthiness was consistency. As with the concept of neutrality, it was unknown for focus groups and phenomenological interviews for the reasons described above.

In terms of the *Program Evaluation Standards*, the three single studies performed quite similarly to each other in terms of propriety and feasibility standards, and in most areas of accuracy. The biggest differences between studies are found in reporting context and providing perspective and rationale for interpretations. In these areas, the interpretive methods were more similar to each other than the post-positivist telephone survey, and the phenomenological interviews differed from the telephone survey more strongly than focus groups.

The extent to which the dialectic and pragmatic mixed method studies in this dissertation meet the *Program Evaluation Standards* was considered. The dialectic study may have less utility than the pragmatic mixed method study because I have less experience with phenomenological interviews. However, the dialectic study may have higher credibility because it contains more information about the perspective and rationale for interpreting findings, and because it provides richer context. Providing richer context may also lead to higher levels of accuracy than the pragmatic study. The dialectic study's higher levels of validity would also lead to greater accuracy than the pragmatic study.

An expert review of single methods was conducted to provide further evidence of the validity of the single methods and to warrant or challenge my assessments of validity / trustworthiness criteria and the extent to which methods met the *Program Evaluation Standards*. Dr. Noriko Ishihara, who specializes in phenomenological approaches as applied to Japanese language acquisition within the Department of Language and Foreign Studies at American University, was selected to review the phenomenological interview methodology. Her review both supported my claims of strong credibility and provided two cautions to credibility: too few subjects in each sampling strata and the limitations of human consciousness and participants' ability to articulate their experience. Dr. Ishihara named four strengths of the study that confirmed my assessment of strong transferability and neutrality. In terms of dependability, Dr. Ishihara posed no critical concerns with the conduct of the interview. In terms of *The Program Evaluation Standards*, the four strengths cited by Dr. Ishihara also support my claims of strong Context Analysis and Evaluator Credibility.

Noted focus group scholar Dr. Richard Krueger (Krueger, 1998; Krueger & Casey, 2000) at the University of Minnesota, was selected to review and critique the focus group methodology. Overall, his comments supported many of my validity claims, especially in the area of credibility and in *The Program Evaluation Standards* of Evaluator Credibility, Report Clarity, Described Processes and Procedures, and Defensible Information Sources. In terms of dependability and neutrality, Dr. Krueger's approval of the focus group processes and protocols suggest that they were

not biased or unduly influenced by preconceived notions, and that a researcher who conducted the focus groups similarly would achieve similar results. However, Dr. Krueger did recommend using an alternative sampling criterion to increase the meaningfulness of findings. He also made one suggestion that challenges my claim of having conducted appropriate Analysis of Qualitative Information. He recommended identifying the comments of individuals. This is an excellent suggestion that would have strengthened my analysis. Unfortunately, due to resource concerns and insufficiently detailed notes, this suggestion was not implemented.

Finally, the President of an independent evaluation and statistical consulting firm that specializes in survey research on tobacco cessation interventions (North American Quitline Consortium, 2008), Dr. Michael Luxenberg of Professional Data Analysts, Inc. was selected to review and critique the methodology for the telephone survey. He confirmed my assessment of weak internal validity but relatively strong external validity. Dr. Luxenberg did not comment on the reliability of measures, except to state the instrument was thoroughly developed. In terms of *The Program Evaluation Standards*, Dr. Luxenberg's assessment supports my claims of Evaluator Credibility, strong Rights of Human Subjects, Report Clarity, Described Processes and Procedures, and Defensible Information Sources.

*Research Question 4. How do stakeholders view the credibility and utility of single method findings and mixed method findings?*

Overall, respondents found both mixed method studies to be somewhat or very credible on a variety of criteria, and both mixed method studies were rated similar to each other overall. The small number of differences that were found appears to reflect the relative strength of the interpretive approach used. For example, the method mix incorporating the phenomenological interviews was judged to better reflect participant perspectives and the underlying truth, but to be less confirmable and less generalizable than the method mix incorporating the focus groups. This reflects a common understanding of the strengths and weaknesses of phenomenological methods. Interestingly, when asked about the single methods directly, respondents perceived focus groups to be slightly more rigorous, but interviews to be slightly more credible to themselves and to their stakeholders. Telephone surveys were also rated as being highly credible to external stakeholders. Respondent experience with the single methods did not appear to affect their assessments of the single methods and mixed method studies.

Respondents also commented on the utility of the mixed method findings. Both respondents reported that the studies would inform media messaging, perhaps in the form of a creative brief or platform that is used by ad agencies to develop new ads and media decisions. One respondent differentiated that the method mix with the phenomenological interviews would “shift advertising messages to hone in on smokers’ experience with the ordinance,” while the method mix with the focus group

would help advertising to become “more sensitive”. These were unintended uses for study findings.

*Research Question 5. What are the cost of the single methods in terms of billable dollars and subject hours?*

The total cost for the suite of three single studies was about \$178,750. The surveys were the most costly method by far (almost \$100,000), followed by focus groups (about \$47,500) and interviews (about \$31,500). Proportionally more billable researcher dollars were used on telephone survey database management, focus group implementation costs, and interview analyses and reporting. In terms of subject hours, focus groups were most intensive (140 hours), followed by the survey (70 hours) and interviews (17.5 hours).

### Conclusions

Researchers and evaluators frequently use mixed methods, but mixed methods' use has been criticized for being reflexive, without thoughtful consideration (Twinn, 2003), simply because it is popular (Stufflebeam, 2001). However, the results presented earlier suggest that mixed methods have important advantages over single methods. Both the pragmatic and dialectic mixed method studies produced unique conclusions on important research questions, above and beyond single methods presented side-by-side. Additionally, mixed methods appear better able to more fully elaborate complex phenomenon than single methods. This can be seen in the complex, web-like divergence between methods. For example, the focus group might contain

contradictory evidence that converged with both the telephone survey and interview, which diverged from one another. These findings reflected the complexity of the phenomenon of study. Single method findings (both post-positivist and interpretive) failed to produce patterns of findings with the same complexity as mixed methods. Taken together, this study illustrates the advantages mixed methods may have over single method studies.

Results presented in the previous sections confirm commonly understood properties of single methods. It is not surprising that post-positivistic surveys produce more narrow and precise findings, while interpretive methods produce more findings on a greater variety of topics and illuminate how and why relationships occur. It is also widely known that strongly interpretive phenomenological interviews provide thicker description and richer information from which the fundamental mechanisms motivating impacts might be revealed, as compared to more weakly interpretive methods. In contrast, focus groups can provide vivid examples of how relationships did or did not occur. Additionally, focus groups are very useful in describing group consensus and reflecting social phenomenon and interaction, while phenomenological methods illuminate sensory lived experience. In the context of mixed methods, these findings confirm the import of thoughtfully choosing individual methods within a mixed method design.

This study also affirms that the choice of interpretive methods matter. In the mixed method literature, interpretive methods can sometimes be lumped together as “qualitative” methods (O’Cathain, Murphy, & Nicholl, 2007; Johnson &

Onwuegbuzie, 2004), with insufficient attention paid to the paradigms that undergird them. In fact, when all method combinations were compared to examine patterns of convergence, the method combination with the greatest number of findings in common was the focus groups and phenomenological interviews, where most of the findings were convergent (80%). This analysis might suggest that the focus group and phenomenological interview methods produced findings that stood in for one another, and that the choice of either focus groups or interviews would result in the same conclusions when mixed with the telephone survey. Perhaps the method and paradigm difference between focus groups and phenomenological interviews did not matter. However, this did not prove to be true. The focus group method produced 17 findings that were substantively unique to the focus groups and not found in any other method (including the phenomenological interviews). Likewise, the phenomenological interview method produced 15 unique findings. The strongly interpretive phenomenological interviews produced different findings than the weakly interpretive focus groups.

Not only do interpretive methods provide different kinds of information such as group consensus versus rich experience, but they produced different finding content. Focus groups provided greater insight into social norms, while phenomenological interviews illuminated smell and respondents' hidden hatred of smoking. Additionally, because the strongly interpretive phenomenological interviews illuminated direct lived experience, which tends to be less mediated by respondent preconceptions and rationalizations, the findings from the interviews were more powerful and trustworthy



than from the focus groups. The examination of trustworthiness revealed that overall, phenomenological interviews tended to have higher levels of credibility and perhaps transferability as compared to focus groups. Additionally, both phenomenological interviews and telephone surveys were rated as being very credible sources of information to the stakeholders of the client organization on the stakeholder survey. When the credibility of the method mix with phenomenological interviews was rated by stakeholders, they found it to better reflect participant perspectives and the underlying truth as compared to the method mix with focus groups. The examination of single methods discussed above illuminates a researcher's choice of what single methods to combine in a mixed method design. This dissertation provides evidence that more strongly interpretive methods may be more powerful and generative when combined with post-positivistic methods as compared to weaker interpretive methods in a mixed method design.

The choice of single methods is especially important when a dialectic approach is used, because the dialectic method is optimized through the investigation of paradox and divergence. Results described above indicate that when combined with telephone surveys, phenomenological interviews produced fewer common findings than focus groups ( $N=14$  vs.  $N=20$ ), and a higher proportion of divergent findings (36% vs. 25%). Therefore, it is possible that a dialectic approach may be more productive with single methods whose paradigms diverge substantially from one another.

Several mixed method theorists have argued that integrating single methods in a mixed method study generates fresh insights above and beyond the side-by-side

presentation of findings from two or more single methods. (Greene, 2007; Tashakkori & Teddlie, 2003; Barbour, 1999; Sandelowski, 1995). However, such claims have not been empirically substantiated and Alicia O’Cathain, Elizabeth Murphy, and Jon Nicholl (2007) demurred that actually specifying unique insights attributable to mixed methods would be difficult due to the level of transparency needed in the analysis stage. This study elaborated with full transparency the analysis stage and found that pragmatic and dialectic approaches to mixed methods produced new conclusions and insights. This finding provides unique empirical support for a key rationale for mixed methods research.

One difficulty in conducting both dialectic and pragmatic mixed method studies is a lack of guidance on how those approaches would be operationalized in practice. The principals of both methods have been articulated to varying degrees in the literature, but the actual processes of integrating methods, especially at interpretation, have garnered much less attention. Operationalizing mixed methods at the point of interpretation is especially important because it is one of the most common points at which methods are integrated (O’Cathain, Murphy, & Nicholl, 2007). This dissertation represents the first explicit operationalization of the dialectic and pragmatic mixing at the point of interpretation, which was based on detailed attention given to their respective philosophical traditions. The dialectic approach was in particular need of articulation because it embodies lofty goals and is supported by eloquent theory, but the practical procedures for its conduct have remained open for interpretation.

My efforts to develop processes and protocols for mixing at the stage of interpretation are first steps. However insufficient, they may begin to actualize the promise of the dialectic or pragmatic approach and provide a stepping stone for other researchers to further advance this field of study. However, further study on a Hegelian-inspired dialectic approach is warranted. One concern is that the Hegelian synthesis requires that two conflicting findings be united by an overarching truth. It is critical that this overarching truth not be used reflexively and ignore the possibility of legitimately conflicting findings due to external factors such as different samples. All synthesis judgments should broadly consider reasons for conflicting findings and incorporate them into synthesis statements. More research on appropriate protocols for dialectic mixing would strengthen the approach.

Pragmatic approaches to mixed methods have gained popularity in the last five years, forcing researchers and evaluators to choose an approach absent empirical evidence. Results presented above provide one side-by-side example of the two approaches that tentatively offers new guidance to the field. Results described above suggest that the advantages of the pragmatic approach are that it is straight-forward. In contrast, the dialectic study had more divergence and more complex convergence than the pragmatic study. Across both studies, divergence was required in order to generate fresh and unique insights. The dialectic mixed method approach was the most generative, supporting a more challenging mixing process that required greater creativity. It is important to note that the conclusion drawn from this study reflects the findings of one, real-world evaluation. Much more research is necessary in order to

draw conclusions about the dialectic approach in other contexts. However, I speculate that this finding would be replicated because the dialectic approach favors more strongly interpretive methods as compared to the pragmatic approach. The difference in kinds of information produced in the two types of studies seems likely to result in similar findings across diverse research and evaluation contexts.

It is also important to note that the dialectic process does not exclude using pragmatic criteria to guide mixing. In fact, the spirit of the dialectic approach would encourage that extant knowledge, practical experience and contextual responsiveness be brought to bear in mixing. In spirit, the dialectic approach to mixed methods subsumes the pragmatic approach and moves beyond it. However, to make this claim in actuality, pragmatic criteria would need to be included in the dialectic mixing protocol, which I did not do. More research in this area would be illuminating. Considering the comparison of the dialectic and pragmatic mixed method studies, the dialectic mixed method study appeared better able to reflect the complexity of the phenomenon of study and to generate the fresh insights unique to mixed methods. It is possible that a dialectic approach is better suited for evaluations seeking to understand more complex phenomenon, while the pragmatic approach may be more appropriate for simpler and more bounded interventions.

A critical threat to my conclusions about the pragmatic and dialectic approaches is that the effect I am attributing to the approaches may be due instead to the single methods (phenomenological interviews for the dialectic approach, focus groups for the pragmatic). Indeed, the single methods play a critical role in the final outcome of the

mixed methods. However, it is important to recall that the dialectic approach prioritizes consideration of paradigms in all stages of mixed method inquiry, from design to interpretation of findings. Because the phenomenological interviews are strongly interpretive, those using a dialectic approach would prefer strongly interpretive methods over weakly interpretive ones in most cases. Likewise, focus groups have many practical advantages such as engaging a larger number of respondents and having a more straightforward approach and philosophy as compared to phenomenological interviews. Therefore, those using the pragmatic approach may tend to use focus groups as opposed to phenomenological methods in a mixed method study. Taken together, the single methods do contribute substantially to the findings about the pragmatic and dialectic approaches. However, the single methods were chosen based on the unique demands of each mixed method approach.

The diverse information needs of multiple stakeholders has been an important rationale for mixed methods (Patton, 1997; Chelimsky, 1997; Benofske, 1995). However, the extent to which mixed methods meets these needs has received less attention. This study lends support to the usefulness of mixed methods for multiple stakeholders. The stakeholder survey suggests that the mixed methods studies were able to meet the needs of my immediate stakeholders, who rated the rigor, meaningfulness and trustworthiness of single methods differently. Additionally, stakeholder survey respondents reported that their stakeholders viewed both the telephone surveys and phenomenological interviews as being most credible. The combination of these two methods in the dialectic mixed method provided both types

of information. In fact, the method mix with the phenomenological interviews was rated to both better reflect participant perspectives and the underlying truth.

The conduct of mixed methods faces several limitations. One is exemplified in stakeholder feedback on single and mixed methods. Interestingly, when asked to rate the credibility of methods on a variety of criteria, respondents rated the single methods higher on average than the mixed method options (7.5 for single methods and 6.1 for mixed methods, on a scale of 10). I hypothesize that the difference in ratings is due to both familiarity with single versus mixed methods and the challenges in judging the quality of mixed method studies. Respondents reported on average a moderate level of experience with the single methods, which presumably enabled them to judge them. Further, stakeholders hold inherit preconceptions of methods that support their decisions making (Green, 2007; Patton, 1997). However, the integration of two or more methods in a mixed method study is much less common (O’Cathain, Murphy, & Nicholl, 1997), and integration with the intent of precisely articulating dialectic or pragmatic approaches is less common still (Riggin, 1997; Green, 2007). Therefore, I hypothesize that stakeholders’ lower ratings for mixed method versus single approaches reflects an unfamiliarity with mixed methods and a general confusion of how to judge the quality of mixed method studies. This is borne out in the literature, where scholars commonly agree that the discussion of validity of mixed method approaches requires further study (Greene, 2007; Dellinger & Leech, 2007; Creswell & Plano Clark, 2007; Teddlie & Tashakkori, 2003).

Another potential limitation of mixed methods is the difficulty of developing expertise in a variety of methodologies. This is especially true for the dialectic approach, which appears to flourish when the single methods employed differ substantially on the paradigm continuum. Candidly, this was a struggle I faced because I have less experience with strongly interpretive methods and more experience with post-positivist and weakly interpretive methods. To prepare for the phenomenological interviews in this dissertation I took a semester-long class in phenomenological methods and conducted a pilot study as a part of that class. In all honesty, this experience was insufficient to prepare me to fully appreciate phenomenological methods and conduct them with some level of mastery. My experience, however, was competent as expressed in the method validation conducted by content expert Dr. Noriko Ishihara. The experience of stretching my boundaries has resulted in a deeper and more sincere appreciation for strongly interpretive methods; I also have gained important skills in conducting tactful and sensitive interviews. Further, the dialectic mixed method approach enhanced the quality of my phenomenological interviews because the approach prioritizes paradigms. Therefore, the dialectic mixed method approach forced me to closely examine and attempt to adhere to the strongly interpretive phenomenological paradigm, which required substantial attention and discipline. I observe that the dialectic approach has an added benefit of helping to maintain the integrity of single methods in a mixed method study, as compared to methodological laziness, in which single methods in a mixed method study are conducted more similarly to one another than differently.

The benefits of having engaged in a dialectic mixed method study are broader than an expanded methodological toolkit. While I may not be an expert in phenomenological methods, my conduct of the methodology resulted in important findings that were of critical use to my client organization. The phenomenological interview results sensitized my client to the experience of smokers in smoke-free ordinances which has impacted the approach and tone of media designed to motivate smokers to quit smoking. The results from my experience and this study should be encouraging to other evaluators who have some capacity for both post-positivist and interpretive methods, as well as the zeal to practice them and experience the benefits of their diversity.

The additional cost of mixed methods above single method designs is another potential limitation to mixed method methodologies. The results described in the previous section shed some helpful light on the dilemma. By far, the telephone survey method was most costly; however, it was rated as highly credible to stakeholders of the client organization and is also the only evidence of effectiveness accepted within the field of tobacco cessation. Therefore, the telephone survey method is necessary for certain purposes. The interpretive methods were much less costly, and the phenomenological method was more than half the cost of the focus group. Therefore, considering a dialectic approach that includes telephone surveys and phenomenological interviews appears to be a useful and cost effective choice.

To what extent does the variety of evidence provided about single and mixed methods in this study support the most cost-effective choice of dialectic methods?



First, when compared to the focus groups, the inferences from phenomenological interviews were found to have stronger trustworthiness and were richer, revealing fundamental mechanisms explaining observed behaviors. Therefore, the choice of the phenomenological interviews had important methodological and substantive benefits. Second, the phenomenological method was found to confirm and diverge from the telephone survey on critical topic areas, and also contributed substantial information over and above the telephone survey. Third, when the dialectic approach was examined, it was found to produce unique insights over and above the single methods presented side-by-side. Therefore, the dialectic mix was more generative and appeared to more fully reveal the complex phenomenon of smoke-free ordinances than the pragmatic approach. Finally, feedback from stakeholders suggests that the dialectic approach is perceived as the most credible in terms of reflecting participant experience and the underlying truth of a phenomenon.

Limitations of the dialectic method include that it was rated by stakeholders as having some weaknesses, such as being somewhat less generalizable and confirmable than the method mix with the focus groups. Also, both method mixes were proven to provide unique insights over and above the single methods presented side-by-side, so the dialectic approach has no inherent advantage over pragmatic methods on this point.

Considering cost together with conclusions from the examination of single and mixed methods, I draw the following conclusions. The pragmatic approach has several disadvantages in comparison to the dialectic approach: the method is less generative and the inferences drawn from focus group method are less trustworthy than those from

the phenomenological interviews. However, if the purpose of mixed method is to more fully uncover a phenomenon, the pragmatic approach is suitable. It confirms and diverges from the telephone survey and provides important unique findings. However, the dialectic approach achieves the goal of more fully uncovering a phenomenon with greater trustworthiness and with more useful and fundamental description. If the purpose of the evaluation is to more fully uncover a phenomenon, the dialectic approach achieves this more effectively than the pragmatic approach and at a lower cost. Additionally, the dialectic approach is superior to the pragmatic approach in yielding more fresh insights and perspectives than the single methods set side-by-side.

#### Limitations

Several limitations in this study circumscribe the conclusions that may be drawn. First, it is important to note that this study represents one simple investigation using one real-world evaluation. Further, the context of this study is a relatively young field with little empirical literature to inform it. As such, this study is most valuable in that it raises important insights that require much additional scholarship for illumination and confirmation. For example, the evaluation used in this case study embodied small treatment effects of the ordinances on smoking outcomes. In this single-case example, one might argue that the difference between pragmatic and dialectic approaches were sensitive to these small differences, but that a case study with a large treatment effect would find no difference between two approaches. However, I speculate that this concern is unfounded. The pragmatic focus group

provided examples of how and why behaviors were seen that focused on group experience and consensus. In contrast, the phenomenological interviews provided richer lived experience that illuminated more fundamental mechanisms of observed behavior. It is, in part, the differences in these methods, chosen to meet the demands of the dialectic and pragmatic approaches, that produced the difference in pragmatic and dialectic approaches. A large treatment effect would not necessarily impact this effect of “kind” of information.

Second, this study focused on mixing methods at the point of implementation, although mixing did occur at both sampling and analysis. Therefore, conclusions are best drawn about mixing at the point of interpretation. Advances are being made in mixing at the point of analysis (Day, Simmons, & Gu, 2008). Results may have been different had I mixed more at that level.

Third, the validity of single and mixed method studies was one component of this study. Unfortunately, a coherent conception of validity in mixed methods requires further scholarship. The strength of my validity claims would have been stronger if a mixed methods validity framework were in place. Further, the assessment of the validity of inferences is largely based on my opinions of what the standards mean, what is required to meet the criteria and the extent to which the criteria are met. No information is available for consistency for any method and the confirmability of focus groups and phenomenological interviews. Additionally, external reviews and stakeholder opinions of single and mixed methods could have been strengthened by having experts complete a form rating the study on all validity / trustworthiness

criteria, and the *Program Evaluation Standards*. These factors limit my ability to make claims about the validity of inferences. Finally, the dialectic and pragmatic method mixes that were submitted to stakeholders to review were drafts of the final pragmatic and dialectic summaries that were developed for this dissertation. The content and format was very similar, but rating may have been different if the final version was used.

Fourth, three practical considerations require attention. I conclude that the dialectic approach is the most productive approach for mixed methods in most cases. However, the cognitive demands of the dialectic approach are considerable, especially as compared to the pragmatic approach. Much more research is necessary to refine the dialectic approach, and training in the dialectic approach would be necessary in research methodology programs. Next, if mixed methods methodology is taught in research programs, it is critical to recognize that funding for research using mixed methods may be limited. Not only is the methodology currently not well understood among funders, but the necessary inclusion of qualitative data collection in a mixed methods study is not supported by many federal research programs. Moreover, mixed method approaches are more expensive than single methods. More study is necessary to understand the research questions and contexts in which the additional costs of mixed methods are outweighed by gains in knowledge and guidance for decision-making.

Finally, it is important to note that the telephone survey and focus groups fully explored the issue of cigarette excise tax. However, due to time constraints and the in-

depth nature of the method, the phenomenological interviews did not discuss the issue in sufficient depth to report. This limited the comparison of convergence, divergence and uniqueness on this subject.

### Recommendations

Using mixed methods is recommended when convergence of findings is desired for triangulation, when diverse perspectives are desired to expand one's understanding of a phenomenon, if a study seeks to initiate fresh insight through an examination of divergent findings, and/or if multiple stakeholders have differing information needs. The dialectic method is best positioned to meet all of these evaluation purposes, although a pragmatic approach may also be suitable if the phenomenon of study is less complex and the development of fresh insights through the examination of divergent findings is not a priority. If a dialectic approach is used, it is optimized by using two single studies that differ substantially from one another on the paradigm continuum. A dialectic method is non-optimal when no divergence of findings is expected between methods.

Mixed method design for the purpose of convergence and triangulation may also be optimized by considering the timing of single study implementation. If mixed method designs are conducted to examine convergence, a sequential design is most effective, where the interpretive methods may develop the content areas that would be explored in a post-positivist survey. Therefore, the amount of unique information in broad topic areas that is contributed by interpretive methods would be somewhat

diminished. Conversely, conducting the post-positivist method first and the interpretive method second would likely result in many unique findings in the interpretive method. This approach can be ideal approach if stakeholders are interested in how different methods produce different findings.

When conducting mixed method research, single methods must be thoughtfully selected. I recommend using a more intensely interpretive approach as compared to a weaker one in order to more fully exploit the benefits of interpretive inquiry. Using a dialectic approach may be helpful in maintaining method integrity. However, it is important to note that more intensely interpretive approaches require substantially more time than less interpretive methods in order to explore the same breadth of issues due to the nature of the inquiry.

A common point to mix methods is at interpretation. It is critical that mixed methods researchers take care to fully integrate findings at the point of interpretation and move beyond presenting method findings in a side-by-side manner. I recommend using an Hegelian framework to mix at the point of interpretation for the dialectic approach. Use of a framework is also recommended for the pragmatic approach. Further research on pragmatic criteria would be helpful to refine the framework based on criteria first proposed by Datta (1997). Additional research is also necessary to develop a framework and criteria for judging the validity of inferences based on mixed method studies.

Finally, practioners of mixed methods optimize the conduct of their studies when they have a sincere interest in and capacity for diverse methods, as well as the

time and budget necessary to carefully attend to each. The value of mixed method studies are diminished by methodological laziness due to lack of budget or lack of interest or capacity on the part of the researcher. Successfully conducted dialectic mixed method studies require the researcher to carefully attend to the conduct and implications of paradigms within the study. While pragmatic studies demand less attention to paradigms, they too require conscientious attention to the details of each single method.

## REFERENCE LIST

- Barbour, R. S. (1999). The case for combining qualitative and quantitative approaches in health services research. *Journal of Health Services Research and Policy*, 4(1), 39-43.
- Barcus, F. E. (1959). *Communications content: Analysis of the research, 1900-1950*. Unpublished doctoral dissertation, University of Illinois.
- Bednarz, D. (1985). Quality and quantity in evaluation research: A divergent view. *Evaluation and Program Planning*, 8, 289-306.
- Beiser, F. (Ed). (1993). *The Cambridge companion to Hegel*. Cambridge, England: Cambridge University Press.
- Benkofske, M. T. (January 1, 1996). An examination of how evaluators and evaluation clients choose data collection methods: The factors and decision-making process. *ETD collection for University of Nebraska - Lincoln*. Paper AAI9623619.
- Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: Free Press.
- Blackburn, S. (Ed.). (1996). *The Oxford dictionary of philosophy* (2<sup>nd</sup> ed.). Oxford, England: Oxford University Press.
- Bliss, L. B. (2008). Media review: Greene, J. C. (2007). Mixed methods in social inquiry. San Francisco: Jossey-Bass. *Journal of mixed methods research*, 2(2), 190-198.
- Bracht, G. H. & Glass G. V. (1968). The external validity of experiments. *American Educational Research Journal*, 5, 437-474.
- Bradshaw, J. (2005). *Healing the shame that binds you*. Deerfield Beach, FL: HCI.
- Brent, J. (2001). Charles Sanders Pierce. In P. S. Boyer & M. Dubsky (Eds.), *The Oxford companion to united states history*. Oxford, England: Oxford University Press.
- Calhoun, C. (Ed.). (2002). Dialectics. *Dictionary of the Social Sciences*. Oxford, England: Oxford University Press.



- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Caracelli, V., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. *New Directions for Evaluation*, 74, 19-30.
- Chelimsky, E. (1997). The political environment of evaluation and what it means for the development of the field. In E. Chelimsky & W. R. Shaddish (Eds.), *Evaluation for the 21<sup>st</sup> century* (pp 53-68). Thousand Oaks, CA: SAGE Publications.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, pp. 37- 46.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: SAGE Publications.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed method approaches*. Thousand Oaks, CA: SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE Publications.
- Creswell, J., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research*. Thousand Oaks: SAGE Publications.
- Creswell, J. W., Trout, S., & Barbuto, J. E. (2002). A Decade of Mixed Methods Writings: A Retrospective. *Research Methods Forum*. pp. 1-30. Published by the Academy of Management. <http://aom.pace.edu/rmd/2002forum.retrospect.html>. ARD # 14355.
- Datta, L. (1994). Paradigm wars: A basis for peaceful coexistence and beyond. *New Directions for Evaluation*, 61, 53-70.
- Datta, L. (1997). A pragmatic basis for mixed method designs. *New Directions for Evaluation*, 74, 33-46.

- Day, C., Sammons, P., & Gu, Q. (2008). Combining qualitative and quantitative methodologies in research on teachers' lives, work, and effectiveness: From integration to synergy. *Educational Researcher*, 37(6), 330-342.
- Dellinger, A. B., & Leech, N. L. (2007). Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, 1(4), 309-332.
- Denzin, N. K. (1978). *The Research Act* (2nd ed.). New York: McGraw-Hill.
- Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.
- Giorgi, A. (1997). The theory, practice and evaluation of the phenomenological method as a qualitative research procedure. *Journal of Phenomenological Psychology*, 28(2), 235-260.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco: Jossey-Bass.
- Greene, J. C., & Caracelli, V. (1997). Defining and describing the paradigm issue in mixed method evaluation. *New Directions for Evaluation*, 74, 5-17.
- Greene, J. C., Caracelli, V., & Graham, W. F. (1989). Toward a conceptual framework for mixed method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.
- Greene, J. C., & McClintock, C. (1985). Triangulation in evaluation: design and analysis issues. *Evaluation Review*, 9(5), 523-545.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Hammersley, M. (1992). *What's wrong with ethnography? Methodological exploration*. London: Routledge.
- Hanson, K. (2005). John Dewey. In T. Honderich (Ed.), *Oxford companion to philosophy*. Oxford, England: Oxford University Press.
- Hare, R. M. (1983). *Plato*. Oxford, England: Oxford University Press.
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.
- Hookway, C. J. (2005). Charles Sanders Pierce. In T. Honderich (Ed.), *Oxford companion to philosophy*. Oxford, England: Oxford University Press.

- Hooten, W. M., Wolter, T. D., Ames, S. C., Hurt, R. D., Vickers, K. S., & Offord, K. P. (2005). Personality correlates related to tobacco abstinence following treatment. *The Journal of International Psychiatry in Medicine*, 35(1), 59-74.
- Howatson, M.C., & Chilvers, I. (Eds.). 2003. Enchelus. *The concise Oxford companion to classical literature*. Oxford, England: Oxford University Press.
- Howell, D. C. (2002). *Statistical methods for psychology* (5<sup>th</sup> ed.). Pacific Grove, CA: Duxbury Thompson Learning.
- Hughes, J. R., Gulliver, S. B., & Fenwick, J. W. (1992). Smoking cessation among self-quitters. *Health Psychology*, 11, 331-334.
- Hunt, W. A., & Bospalec, D. A. (1974). An evaluation of current methods of modifying smoking behavior. *Journal of Clinical Psychology*, 30, 431-438.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602-611.
- Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Boston, MA: Allyn and Bacon.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Joint Committee for Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. McGraw Hill Book Company.
- Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards* (2<sup>nd</sup> ed). Thousand Oaks, CA: SAGE Publications.
- Jolivet, J. (2001). Dialectic. In A. Vachez (Ed.), *Encyclopedia of the middle ages*. Oxford, England: Oxford University Press.
- Kahn, D. L. (1993). Ways of discussing validity in qualitative nursing research. *Western Journal of Nursing Research*, 15, 122-126.

- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Kaufman, G. (1992). *Shame: The power of caring*. Rochester, VT: Schenkman Books.
- Kidder, L. H., & Fine, M. (1987). Qualitative and quantitative methods: When stories converge. *New Directions in Program Evaluation*, 35, 57-76.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, (88), 5-23.
- Krueger, R. A. (1998). *Moderating focus groups*. Thousand Oaks, CA: SAGE Publications.
- Krueger, R.A., & Casey, M.A. (2000). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: SAGE Publications.
- Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.
- Leininger, M. (1994). Evaluation criteria and critique of qualitative research studies. In J. M. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 95-115). Thousand Oaks, CA: SAGE Publications.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: SAGE Publications..
- Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation*, 30, 73-84.
- Lynch, K. B. (1983). Qualitative and quantitative evaluation: Two terms in search of a meaning. *Educational Evaluation and Policy Analysis*, 5(4), 461-464.
- Mark, M. M., Feller, I., & Button, S. B. (1997). Integrating qualitative methods in a predominantly quantitative evaluation: A case study and some reflections. *New Directions for Evaluation*, 74, 47-60.
- Mark, M. M., & Shotland, R. L. (1987). Alternative models for the use of multiple methods. *New Directions in Program Evaluation*, 35, 95-100.

- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(3), 13-17.
- Maxcy, S. J. (2003). Pragmatic threads in mixed method research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in the social and behavioral sciences*. Thousand Oaks: SAGE Publications.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.
- Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (1995, September), 741-749.
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48-76.
- Morgan, G. (1983). *Beyond method: Strategies for social research*. Beverly Hills, CA: Sage Publications.
- North American Quitline Consortium. (2008, November). *Calculating quit rates*. (White Paper). Phoenix, AZ: An, L., Luxenberg, M. G., Betzner, A., Rainey, J. T., & Capesius, T. C.
- O'Cathain, A. (2009). Editorial: Mixed methods research in the health sciences: A quiet revolution. *Journal of Mixed Methods Research*, 3(1), 3-6.
- O'Cathain, A., Murphey, E., & Nicholl, J. (2007). Integration and publications as indicators of "yield" from mixed methods studies. *Journal of Mixed Methods Research*, 1(2), 147-163.
- Palmer, R. (1969). *Hermeneutics: Interpretation theory in Schleiermacher, Dilthey, Heidegger, and Gadamer*. Evanston, IL: Northwestern University Press.
- Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health*, 84(7), 1086-1093.
- Patton, M. Q. (1985). Logical incompatibilities and pragmatism. *Evaluation and Program Planning*, 8, 307-308.

- Patton, M. Q. (1997). *Utilization-focused evaluation*. Thousand Oaks, CA: SAGE Publications.
- Patton, M. Q. (2002). *Handbook of qualitative research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: SAGE Publications.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4<sup>th</sup> ed.). Thousand Oaks, CA: SAGE Publications.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Polkinghorne, D. E. (1989). Phenomenological research methods. In R. S. Valle & S. Halling (Eds.), *Existential-phenomenological perspectives in psychology: Exploring the breadth of human experience* (pp. 41-60). New York: Plenum Press.
- Potts, S. A. K. (1998). Impact of mixed method designs on knowledge gain, credibility, and utility of program evaluation findings. *Dissertation Abstracts International*, 59(06), 1942A.
- Reichardt, C. S., & Rallis, S. F. (1994a). The relationship between the qualitative and quantitative research traditions. *New Directions for Evaluation*, 61, 5-12.
- Reichardt, C. S., & Rallis, S. F. (1994b). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. *New Directions for Evaluation*, 61, 85-92.
- Rescher, N. (2001). *Cognitive pragmatism: The theory of knowledge in pragmatic perspective*. Pittsburgh, PA: University of Pittsburgh Press.
- Riggin, L. J. C. (1997). Advances in mixed-method evaluation: A synthesis and comment. *New Directions for Evaluation*, 74, 87-95.
- Roberts, J. (Ed.). (2007). Plato. *Oxford dictionary of the classical world*. Oxford, England: Oxford University Press. *Oxford Reference Online*. Retrieved 13 June 2008 from <<http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t180.e1744>>.
- Robinson, R. (1953). *Plato's Earlier Dialectic*. Oxford, England: Clarendon Press.
- Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large scale research study. *Evaluation Review*, 9(5), 627-643.

- Sandelowski, M. Triangles and crystals: On the geometry of qualitative research. *Research in Nursing and Health*, 18, 569-574.
- Sax, G., & Newton, J. W. (1997). *Principals of educational and psychological measurement and evaluation* (4<sup>th</sup> ed.). New York: Wadsworth Publishing Company.
- Schofield, J. W. (1990). Increasing the generalizability of qualitative research. In E. W. Eisner & A. Peshkin (Eds) *Qualitative inquiry in education: The continuing debate* (pp. 201-232). New York: Teachers College Press.
- Sedley, D. N. (2004). *The midwife of Platonism : Text and subtext in Plato's Theatetus*. Oxford, England: Clarendon Press.
- Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. *New Directions for Evaluation*, 60, 13-58.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Boston: Houghton Mifflin.
- Shaffer, D. W., & Serlin, R. S. (2004). What good are statistics that don't generalize? *Educational Researcher*, 33(9), 14-25.
- Simon, T. L., & Burstein, P. (1985). *Basic research methods in the social sciences*. New York: McGraw Hill.
- Singer, P. (1983). *Hegel*. Oxford, England: Oxford University Press.
- Singer, P. (2001). *Hegel: A very short introduction*. Oxford, England: Oxford University Press.
- Shotland, R. L., & Mark, M. M. (1987). Improving inferences from multiple methods. *New Directions in Program Evaluation*, 35, 77-94.
- Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the qualitative-quantitative debate among educational researchers. *Educational Researcher*, 15, 4-12.
- Smith, M. L. (1994). Qualitative plus/versus quantitative: The last word. *New Directions for Evaluation*, 61, 37-44.

- Smith, M. L. (1997). Mixing and matching: Methods and models. *New Directions for Evaluation, 74*, 73-86.
- Sprigge, T. L. S. (2005). William James. In T. Honderich (Ed.), *Oxford companion to philosophy*. Oxford, England: Oxford University Press.
- SRNT Subcommittee on Biochemical Verification. (2002). Biochemical verification of tobacco use and cessation. *Nicotine and Tobacco Research, 4*, 149-159.
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation, 89*, 7-98.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks: SAGE Publications.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in the social and behavioral sciences*. Thousand Oaks: SAGE Publications.
- Thayer, H. S. (1968). *Meaning and action: A critical history of pragmatism*. Indianapolis, IN: Bobbs-Merrill.
- Twinn, S. (2003). Status of mixed method research in nursing. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 3-50). Thousand Oaks: SAGE Publications.
- Van Manen, M. (1990). *Research lived experience: Human science for an action sensitive pedagogy*. New York: State University of New York Press.
- Velicer, W. F., & Prochaska, J. O. (2004). A comparison of four self-report smoking cessation outcome measures. *Addictive Behaviors, 29*(1), 51-60.
- Velicer, W. F., Prochaska, J. O., Rossi, J. S., Snow, M. G., (1992). An evaluation of current methods of modifying smoking behavior. *Journal of Clinical Psychology, 30*, 431-438.
- Yu, C. H. (2006). *Philosophical foundations of quantitative research methodology*. Lanham, MD: University Press of America.
- Walker, T. J. E. (2008). Pragmatism. In P. N. Stearns (Ed.), *The Oxford encyclopedia of the modern world*. Oxford, England: Oxford University Press.



- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell, A. Kubisch, L. Schorr, & C. Weiss, (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts*, Washington, DC: Aspen Institute.
- Whittenmore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11(4), 522-537.
- Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (Eds). (2004). *Handbook of practical program evaluation* (2<sup>nd</sup> ed.). San Francisco, CA: John Wiley & Sons.

## APPENDIX A. STAKEHOLDER SURVEY

## Stakeholder Feedback Survey on Mixed Methods in the Ordinance Impact Study

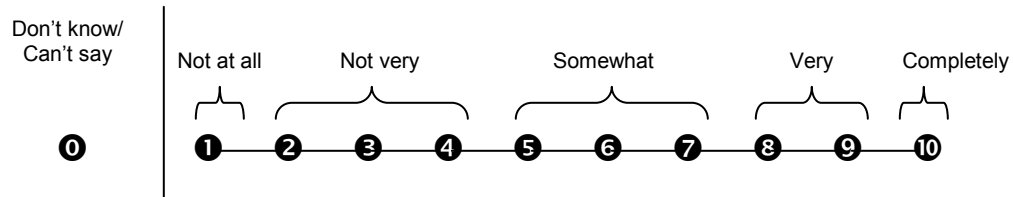
*Thank you for agreeing to participate in this survey about your feedback on which method mixes are most relevant and useful to you in the Ordinance Impact Study. The survey should take about 15 minutes to complete. Please return your survey via email by June 20, 2007. Your participation is voluntary and your responses will be kept confidential. If you have any questions about the survey, please telephone Anne Betzner at 612/623-9110 or email to [abetzner@pdastats.com](mailto:abetzner@pdastats.com). Thank you for your time.*

1. What is your title? \_\_\_\_\_

*This section asks about your experience with a range of evaluation methods.*

2. Before working on the Ordinance Impact Study, what was your experience with telephone surveys? Would you say .... (check all that apply):
- Little/no experience
  - I understand the basic method of telephone surveying
  - I have used telephone survey findings in my job
  - I have designed or administered a telephone survey individually or as a part of a team
  - I have managed telephone survey projects
  - Other (please specify): \_\_\_\_\_
3. Before working on the Ordinance Impact Study, what was your experience with focus groups? Would you say .... (check all that apply):
- Little/no experience
  - I understand the basic method of focus groups
  - I have used focus group findings in my job
  - I have designed or administered focus groups individually or as a part of a team
  - I have managed focus group projects
  - Other (please specify): \_\_\_\_\_
4. Before working on the Ordinance Impact Study, what was your experience with phenomenological interviewing? Would you say .... (check all that apply):
- Little/no experience
  - I understand the basic method of phenomenological interviewing
  - I have used phenomenological interview findings in my job
  - I have designed or administered phenomenological interviews individually or as a part of a team
  - I have managed phenomenological interview projects
  - Other (please specify): \_\_\_\_\_

This section asks about your general opinions about the same three evaluation methods and their findings. For each question below, please use this rating system and write in the number that best matches your opinion:



5. Thinking about what you know about the phenomenological interviewing method right now, how rigorous a method do you think it is? Please write the rating number that best fits your opinion:

**Rating Number:** \_\_\_\_\_

6. Thinking about findings of phenomenological interviews, please answer the following questions:

6a. In general, how meaningful do you think the findings are?      **Rating Number:** \_\_\_\_\_

6b. In general, how trustworthy do you think the findings are?      **Rating Number:** \_\_\_\_\_

6c. In general, how credible are the findings to you?      **Rating Number:** \_\_\_\_\_

6d. In general, how credible are the findings to your stakeholders?      **Rating Number:** \_\_\_\_\_

7. Thinking about what you know about the focus group method right now, how rigorous a method do you think it is? Please write the rating number that best fits your opinion:

**Rating Number:** \_\_\_\_\_

8. Thinking about findings of focus groups, please answer the following questions:

8a. In general, how meaningful do you think the findings are?      **Rating Number:** \_\_\_\_\_

8b. In general, how trustworthy do you think the findings are?      **Rating Number:** \_\_\_\_\_

8c. In general, how credible are the findings to you?      **Rating Number:** \_\_\_\_\_

8d. In general, how credible are the findings to your stakeholders?      **Rating Number:** \_\_\_\_\_

9. Thinking about what you know about the telephone survey method right now, how rigorous a method do you think it is? Please write the rating number that best fits your opinion:

**Rating Number:** \_\_\_\_\_

10. Thinking about findings of telephone surveys, please answer the following questions:

10a. In general, how meaningful do you think the findings are?      **Rating Number:** \_\_\_\_\_

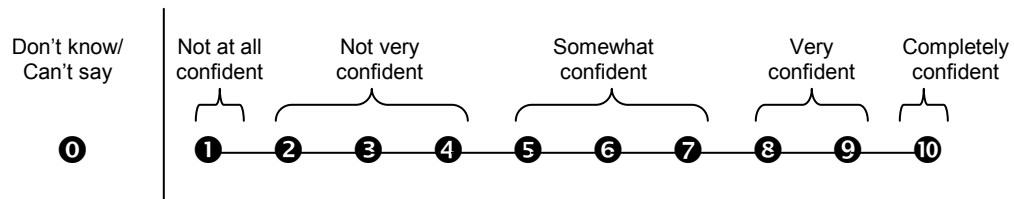
10b. In general, how trustworthy do you think the findings are?      **Rating Number:** \_\_\_\_\_

10c. In general, how credible are the findings to you?      **Rating Number:** \_\_\_\_\_

10d. In general, how credible are the findings to your stakeholders?      **Rating Number:** \_\_\_\_\_

## Feedback on Mixed Method Studies #1 and #2

Please read the Method Mix 1 and Method Mix 2 documents provided with this survey. Mixed Mix 1 summarizes and combines the findings of the telephone survey and focus group methods. In Method Mix 2, the telephone survey and interview findings are summarized and combined. After reviewing the two documents, please use the rating system below and write in the number that best matches your opinion for each question.



11. Please rate each method mix using the scale above. Thinking about Method Mix 1 and 2, how confident are you that the findings of each study ...

	<u>Ratings for Each Method Mix</u>	
	Method Mix 1	Method Mix 2
20a. Reflected participants' perspectives?	_____	_____
20b. Describe and account for changes in the research setting?	_____	_____
20c. Reflect a neutral point of view on the part of the evaluator?	_____	_____
20d. Could be confirmed or corroborated by others?	_____	_____
20e. Could be generalized to groups of people beyond the study sample?	_____	_____
20f. Would be the same if subjects were reassessed, excluding changes due to the passage of time?	_____	_____
20g. Reflect the underlying truth of participants' experiences?	_____	_____
20h. Are the result of the ordinance and not other factors?	_____	_____

12. Do you think that Method Mix 1 has or would influence ClearWay Minnesota's approach or activities?

- Yes                       No    →    IF NO, Continue to Item 13



*IF YES: Please briefly describe:*

12a. What changes did or might occur?	12b. What are the findings or evaluation processes that would influence those changes?

13. Do you think that Method Mix 2 has or would influence ClearWay Minnesota's approach or activities?

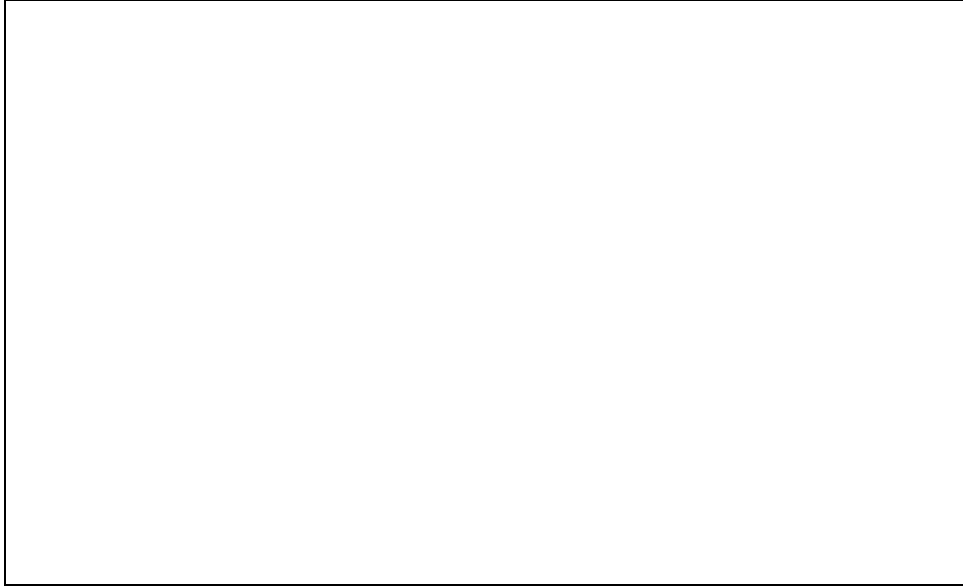
- Yes                       No    →    IF NO, Continue to Item 13



*IF YES: Please briefly describe:*

13a. What changes did or might occur?	13b. What are the findings or evaluation processes that would influence those changes?

14. Please share any additional comments you have on Method Mixes 1 and 2 in the space provided below:



**Thank you very much for your time!**

*Please email the survey back to Anne Betzner at [abetzner@pdastats.com](mailto:abetzner@pdastats.com) by 7-1-07.*

APPENDIX B: SINGLE METHOD FINDINGS BY TOPIC AREA

Table B1

*Survey Study Findings by Topic Area*

Topic Area	Finding
External generalizability	<ul style="list-style-type: none"> <li>• Findings may be generalized to Minnesota smokers who have been motivated enough to quit that they enrolled in an evidence based intervention. Despite this effort, most have failed in their quit attempt and may be considered hard core smokers.</li> <li>• 55.3% response rate</li> <li>• Rates of participant response to the 18 month survey were similar regardless of whether a person was abstinent at 6 months, a key threat to validity.</li> <li>• Survey responders tend to be more stable than non-responders.</li> <li>• Overall, the non-response analysis shows that external validity is moderately strong. The key threat was unwarranted, and more stable responders is expected and found in similar studies.</li> </ul>
Respondent characteristics	<ul style="list-style-type: none"> <li>• About one-half of respondents live in outstate Minnesota (45.7%), about one third in Ramsey or Hennepin Counties (34.4%), and one-fifth in surrounding 5-county suburban area (19.9%).</li> <li>• Respondents were most likely to be female (59.9%), 41 years of age or older (61.3%), white (94.0%), employed (77.4%), insured (88.2%), and completed some college or more (74.3%). About half were married (55.4%).</li> <li>• Respondents were most likely to be daily smokers (81.5%) who smoked within 30 minutes of waking (72.3%). About half were moderate smokers (50.1%) and made a quit attempt in year prior to enrollment (55.6%). More than one fifth quit for a year or more at some point during their lives (23.1%).</li> </ul>

Comparison of outcomes at 6 and 18 months post enrollment

- Respondents were most likely to be somewhat or very satisfied with QUITPLAN services 6 months after enrollment (78.1%) and program utilization at that time ranged from low (21%) to moderate (43%) to high (36%).
- The 7-day abstinence ITT rate is statistically similar at 6 and 18 months.
- The 7-day abstinence completer rate is 32.7% (95% CI = 29.1%, 36.3%). This is due to the 17.7% ( $N=112$ ) of completers who stayed quit from 6 to 18 months, plus 14.9% ( $N=94$ ) who achieved a new quit from 6 to 18 months. The 6.4% who relapsed from 6 to 18 months ( $N=40$ ) is subtracted from this group.
- The relapse rate is 26.3% (95% CI = 19.3%, 33.3%).
- 32.5% of respondents reported being exposed to the ban.

Exposure to ban (Independent variable of interest)

Key moderating variables: Support for the ordinance

- Respondents are about equally likely to support (46.1%) and not support the ordinance (42.6%). This differs from other population-based studies that show 70% approval ratings.
- The difference in support is likely due to the survey population which is mostly comprised of hard core smokers.

Key moderating variables:

Travel to communities with different smoking regulations

- Most respondents do not travel regularly to bars / restaurants with different regulations (74.2%). Those in ordinance communities were more likely to travel regularly than those in non-ordinance communities ( $p = .002$ ).
- The difference in regular travel between ordinance and non-ordinance communities may reflect confounding factor of geographic location (ordinance communities are largely in Hennepin-Ramsey Counties proximal to suburban non-ordinance counties).
- The most respondents report traveling for reasons other than smoking regulations (66.9%).
- Those who travel to go to smoking communities are more likely to be smokers ( $p<.001$ ) and live in



ordinance communities ( $p = .002$ ).

Key moderating variables:

- A substantial minority of respondents (19.2%) report that the cigarette tax helped them think about quitting.

Self report of cigarette tax impact

Impact of ordinance on quitting

- A bivariate examination of exposure to ban and relapse shows that those exposed to smoke free ordinances are two times less likely to relapse as compared to those not exposed ( $p=.004$ ). Exposure to ban was marginally associated with relapse in a logistic regression ( $p=.061$ ).
- Exposure to ban was not significantly associated with either 7-day abstinence or having made a new quit.
- Participant self-report of impact of bans is mixed and not consistent.

Table B2

*Focus Group Study Findings by Topic Area*

Topic Area	Finding
External generalizability	<ul style="list-style-type: none"> <li>• Focus groups are not designed to be judged by the same standards of generalizability as telephone surveys. Instead, they are designed to be transferable to similar contexts. Findings may be generalized to smokers in the Twin Cities metropolitan area who have been motivated enough to quit that they enrolled in an evidence based intervention. Despite this effort, most have failed in their quit attempt and may be considered hard core smokers.</li> </ul>
Respondent characteristics	<ul style="list-style-type: none"> <li>• One third of respondents each lived in Hennepin County, Ramsey County and the 5 County suburban area. About 60% were female. About 60% were smokers.</li> </ul>
Travel to communities with different smoking regulations	<ul style="list-style-type: none"> <li>• For many smokers and non-smokers, bans have no impact on their travel either because they already don't smoke when they go out with friends, or because the inconvenience of traveling outweighs the benefit.</li> <li>• As might be expected, some smokers avoided non-smoking bars or traveled to smoking ones, if the bans had an effect. They did this because smoking is relaxing to them and being in non-smoking establishments makes them feel like second-class citizens.</li> <li>• Some non-smokers reported avoiding smoking bars or travel to non-smoking bars, if the ban had an effect because they hate being in a smoky atmosphere or were actively trying to support their quit and avoid trigger circumstances. Some smokers also avoid smoking bars during their quit attempt. Most can return to smoky bars, but for some non-smokers, it is less frequent.</li> <li>• Some travel just for special occasions, others travel more regularly. Some smokers report going out less because of they can't smoke in bars or restaurants, but some report going out more. This latter group prefers non-smoking environments because they hate the smell in</li> </ul>

public places, which is absent under the ordinance.

Key moderating variables:  
Self report of policy impact

- Decisions to travel often have to do mood or with friends and whether they smoke and want to travel.
- Bans appear about equally likely to have either no effect on traveling or to have an effect based on smoking status.
- Money was a big motivator for some non-smokers to quit because they had to choose between spending money on cigarettes or other things they wanted or needed.
- The tax caused some smokers to use coupons, buy a cheaper brand, or travel to another state to buy cigarettes.
- The tax caused some smokers to think about quitting or even motivated quit attempts due to personal financial constraints.
- The tax helped some smokers cut down and buy and smoker fewer cigarettes.
- The tax had no meaningful effect on many smokers. They said it made them think about quitting, but they never did.
- Bans created conditions where maintaining a quit attempt was easier: they removed the temptation of cigarette smoke from bars and restaurants and at least partially removed the temptation of seeing groups of people enjoying cigarettes.
- Bans gave non-smokers a substantially better experience in bars and restaurants because the lack of cigarette smoke and its smell.
- The ban was not helpful and had no impact for some non-smokers.
- No impact appears caused by several factors: they already adapted to not smoking due to family / friends, they were proud of their willpower and had difficulty seeing the impact of

Impact of ordinance on  
non-smokers – Positive  
impacts

Impact of ordinance on  
non-smokers – Neutral  
impact

factors outside of themselves, political dislike of government controls blunted their ability to see positive benefits from bans, non-smokers take bans for granted, seeing and articulating the positive impact of an absence of cigarette smoke is cognitively difficult.

#### Impact of ordinance on smokers – Positive impact

- The many barriers to identifying the impact of bans suggests that the impact may be underreported.
- Smoke free ordinances motivated some smokers to think about quitting. This is because they make drinking and smoking in bars inconvenient, and they take some of the enjoyment out of smoking by separating friends and prompting smokers for feeling guilty for smoking.
- Smoke-free ordinances motivated some smokers to smoke less due to inconvenience of having to get up and smoke. Smoking less is mitigated by seasonal effect, for example, it is easy to smoke outside in the summer and less comfortable in the winter.
- An unexpected finding was that some smokers have a better bar and restaurant experience because they hate the smell of smoke (contrarily, even as they are smokers themselves). The bans create smoke-free environment that these smokers prefer to bars where tobacco smoke is everywhere.

#### Impact of ordinances on smokers – Neutral impact

- Some smokers report that the ban had no impact on them.
- Reports of no impact may be due to lack of exposure to ordinances, the strength of their addiction was more powerful than the impact of a ban, they were acclimated to not smoking in public due to family, political beliefs that make acknowledgement of impact unlikely, and cognitive difficulty expressing positive impact of lack of cigarette smoke.

#### Impact of ordinances on smokers – Negative impact

- Some smokers feel resentment at being told what to do; they desire autonomy in their decisions to smoke, often because they believe it is their right to do so. However, people's perceptions of their rights differ; some believe they should be able to smoke everywhere, others feel in restaurants and bars, and still others in bars and bowling alleys only.

- Smokers can be angry at being told what to do, and this can result in feeling defiant. The defiance may be righteous or self-admittedly juvenile.
  - Smokers' defiance may be a reaction to their frustration of their addiction that is highlighted by the ban and which conflicts with their independent nature
- Impact of ordinances –  
Role of smoking accommodations
- Smoking accommodations help smokers feel less like second class citizens. For some, they continue to smoke less. Others smoke just as much, like when drinks are served on a patio, for example.
  - Some smokers are motivated to avoid smoking by enclosed smoking rooms. The rooms make the impact of smoking on your body more apparent by concentrating smell, smoke, particulate, etc.
- Impact of ordinance – Role of drinking and bars
- Drinking is a trigger to smoke, and drinking and smoking are activities associated with bars, especially because bars are places where friends gather and socialize, which is also associated with drinking and smoking.
  - Bars are a common site for relapse.
  - Impact of bans is due at least in part to relationship between drinking and smoking and bars. Smoking ordinances aid quit attempts and prevent relapse because they at least partially interrupt the association of drinking and smoking in bars.
- Experience of social norms
- Everyone understands that smoking is harmful and feels social pressure to stop smoking. Some take the message personally and feel they are bad people because they smoke.
  - Messages not to smoke come from the media. They can cause guilt and anxiety and motivate people to quit. They can also trigger people to feel they are bad that they are smoking.
  - Messages come from people's reactions to seeing a smoker smoke a cigarette. They are

glared at, condescended to, judged and treated rudely. Smokers were pained by these experiences.

Social norms and quitting

- Smokers are developing new social norms about smoking that is more considerate of non-smokers, like not smoking near others, not leaving ashes and butts, etc.
- Smokers hear anti-smoking messages from their kids who get tobacco education.
- The social unacceptability of smoking appears to be a powerful motivator for smokers to quit and consider quitting.

Impact of bans – role of ordinances

- Bans are an external enforcement of social norms and are highly visible. This is a reason some people quit.
- Younger smokers appear more amenable to social norm messages than older smokers. More research in this area is necessary.
- While subtle and not immediate, social norms against smoking appear to be the most promising mechanism to motivate new quit attempts, prevent relapse, and prevent individuals from starting to smoke in the first place.

Table B3

*Interview Study Findings by Topic Area*

Topic Area	Finding
External generalizability	<ul style="list-style-type: none"> <li>Phenomenology is not designed to meet the standards of generalizability, but it is designed to uncover the invariant meaning of a phenomenon (Giorgi, 1997) that may be transferred to similar contexts. Therefore, it is important to understand that interview participants includes smokers in the Twin Cities metropolitan area who have been motivated enough to quit that they enrolled in an evidence based intervention. Despite this effort, most have failed in their quit attempt and may be considered hard core smokers. All participants report frequenting bars or restaurants at least monthly.</li> </ul>
Respondent characteristics	<ul style="list-style-type: none"> <li>One third of respondents each lived in Hennepin County, Ramsey County and the 5 County suburban area. About 50% were female. About 50% were smokers. Many interviewees (N=8 of 14) volunteered that they had multiple problems, such as other addictions, anger management issues, serious health issues, very unhealthy relationship situations, etc.</li> </ul>
Travel to communities with different ordinances	<ul style="list-style-type: none"> <li>Most participants did not report traveling. However, three smokers traveled to smoking establishments. For two, travel was infrequent.</li> </ul>
Impact of ordinance – Positive impacts	<ul style="list-style-type: none"> <li>Over half of participants reported that the ordinances motivated them to quit, supported them in their quit, or helped them smoke less.</li> <li>Some participants reported having better bar and restaurant experiences because they avoided exposure to secondhand smoke.</li> </ul>
Impact of ordinance – Neutral impact	<ul style="list-style-type: none"> <li>Fewer participants reported no impact on their quit versus some impact, largely because they had little exposure to the ordinances or little desire to quit.</li> </ul>
Impact of ordinances –	<ul style="list-style-type: none"> <li>Some smokers were angered and frustrated by not being able to maintain autonomy over</li> </ul>

Negative impact

their smoking. One result is feeling self-admittedly defiant in a juvenile way.

Impact of Minnesota Clean Indoor Air Act

- Almost all participants were exposed to the Act and reported that it reduced their smoking.

Impact of personal bans

- About half of participants reported encountering personal bans, most commonly home and car bans, often because of other family members' insistence. These taught participants that they could live without a cigarette, and most adapted successfully to the change. People who instituted personal bans to help them quit had mixed success. Almost all current smokers hide or sneak cigarettes occasionally in response.

Social norms

- Participants acknowledge the harm of secondhand smoke. Non smokers were most protective of their right to breathe clean air, but smoker also acted to protect their loved ones from secondhand smoke.
- Smokers also show consideration for nonsmokers in their smoking activity and understand some people don't like to be around smoke. This includes family members and strangers.
- The factors seem to coalesce into a growing set of social norms of where smoking is acceptable and not acceptable.
- Smoking behavior seems to be sensitive to these norms. Bans create physical environments where smoking is acceptable and not acceptable. Bans came into being because of norms about clean air, and continue to create them.

Meaning of the term regulation

- The term regulation comprises several meanings that contrast with and confirm one another. Regulation suggests a rule, law or system that when initially instituted causes abrupt change, as well as what is customary or ordinary. Regulation can be understood as the process by which an agent seeks to adapt to a change: either by staying the same or moving towards development, growth and integration. Bans also seek to accomplish a variety of goals such as to equally protect all citizens from secondhand smoke in the workplace and in public places.



Functions of the ban

- To enforce specified smoking behaviors and reduce or eliminate secondhand smoke in specified areas. The bans make smoking inconvenient, especially where drinks are not allowed outdoors, because people must leave friends and their drink to smoke.

Emotional experience –  
Relief; Impact – cut down

- Some smokers experience relief at a policy that is an external support to smoke less in ways they wish they could change on their own, but cannot. Smokers and non-smokers experience relief at not being exposed to smoky environments.

Emotional experience –  
Struggle with addiction /  
panic; Impact – none,  
motivation to quit

- Smokers struggle with their cravings and addiction when they have cravings but are prohibited from smoking. This can lead to apprehension, frustration and even panic. Some weather these situations, while others gain fresh insight into the control nicotine has over them and are motivated to quit.

Emotional experience –  
Anger; Impact – Smoke in  
home

- Some smokers are angry at being restricted from smoking, and resentful of having their autonomy to smoke taken away from them. These smokers defend and treasure smoking in their home.

Emotional experience –  
Hiding / sneaking, shame

- Hiding and sneaking cigarettes is a central experience for many smokers who cannot smoke in certain places and/or do not want to be seen smoking by others. This can both stem from and lead to feelings of guilt, shame and embarrassment for smoking.

Emotional experience –  
Hiding / sneaking, shame

- Feelings of guilt and shame are a central experience for many smokers. Shame is marked by feeling seen “in a powerfully diminished sense where the self feels exposed both to itself and others” (Kaufman, 1992, p. 45). It is a common antecedent and consequence of addiction. Regulations may provide new opportunities to experience existing shame.

Cognitive experience –  
Cognitive dissonance.  
Impact – Reduction in  
smoking, motivation to  
quit, no change.

- Some smokers experience cognitive dissonance due to the ban because existing knowledge of the harms of tobacco are confronted by a person’s active efforts to overcome the inconvenience to smoking that bans cause. Cognitive dissonance can result in reduction in smoking and efforts to quit.

Experience of no ban –  
Gross physical experience

- A smoky environment is a palpable physical experience for many. Smoke is absorbed through the eyes, abrades the throat and assaults the senses. It has a weight and abrasion that lasts beyond the time that one is exposed to it. Being in a smoky environment can give you a smoking hangover.

Experience of ban – Smell,  
mostly bad

- Almost all interviewees discussed the smell of tobacco smoke unprompted. Some stated that the smell was appealing and a trigger to them to smoke, but more described it in strongly emotional, guttural and unequivocally negative terms.

Experience of ban – Smell,  
why bad

- Current and former smokers describe the smell as nasty and disgusting. They hate the smoke permeates everything around it and attaches to clothes and hair and lingers, an involuntary and disconcertingly intimate reminder of being in a smoking environment. The smell of smoke could “out” a person as a smoker involuntarily. Participants associated the smell of smoke as being dirty.

Experience of ban – Smell,  
bad because of shame

- Being seen by others as being a smoker that smells dirty echoes smokers’ experience of shame in smoking. Smell elicited memories and retrospective shame in smokers about themselves as children smelling like cigarette smoke.

Impact of smell –  
Motivation to quit, prevent  
relapse

- The smell of smoke, especially in enclosed areas like a car, is a powerful motivator for some smokers to quit and stay quit.

Reflection on method

- Phenomenology seeks to understand participants’ lived experience unfettered by intellectualized reflections, preconceptions, rationalizations and political beliefs. Therefore, the description of a participant’s senses (sight, smell, sound, touch and taste) is critical; and not surprising that the phenomenological interviews focused so strongly on the smell of tobacco smoke. Given the framework of phenomenology, it is possible that the smell of tobacco – and participants’ disgust of it, even among smokers – may reflect how they experience the harm of tobacco, as it difficult for them to express this cognitively.

Impact of bans - Positive

- Bans benefit all individuals who dislike the smell of cigarette smoke, regardless of their understanding or appreciation of it.

APPENDIX C: CONVERGENCE, DIVERGENCE AND  
UNIQUENESS OF ALL SINGLE METHOD FINDINGS BY TOPIC AREA

Table C

*Convergence, Divergence, and Uniqueness of All Single Method Findings by Topic Area*

Relation-ship <sup>1</sup>	Substantive Findings
	<i>Topic: External Generalizability</i>
C	<ul style="list-style-type: none"> <li>All studies have the same base population: tobacco users who enrolled in an evidence-based intervention, QUITPLAN.</li> </ul>
C-FG&I, D-S	<ul style="list-style-type: none"> <li>For surveys, the population of respondents was statewide and most were still smoking at 18 months post-enrollment so can be considered hard core smokers. The findings of the survey may be generalized to this group only. Non-smokers at 18 months were not oversampled because sampling was exhaustive in all but the outstate / no ban group.</li> </ul>
Dd-FG&I	<ul style="list-style-type: none"> <li>The respondent population for Int &amp; FGs was further narrowed to those living in the Twin Cities 7-County metropolitan area. Non-smokers at 18 months were oversampled for Int &amp; FGs so that about half of participants fell into each group.</li> <li>The population of respondents for interviews was further narrowed to those that frequent bars or restaurants monthly or more.</li> <li>Note that the standards for generalizability differ for each method. Phenomeonological interviews seek invariant meaning that may be transferred to similar contexts; focus group findings are designed to be transferred to similar contexts.</li> </ul>

C • Response rates are provided for all methods. 55.3% for the survey; 47% for focus groups; 52% for interviews

U-S • Response bias analysis – external validity is moderately strong. The key threat of rate of response differing by abstinence status at 6 months was unwarranted. More stable eligible participants responded, consistent with similar studies. Represents greater precision of survey method.

*Topic: Respondent Characteristics*

C • Gender, age, race employment status, insurance status, highest education level completed and marital status was collected for all respondents to all methods. Clinical characteristics at intake and respondent satisfaction and use of QUITPLAN services at six months were also collected for all respondents in all methods. Most participant characteristics were similar across methods, except that more men participated in interviews.

C-FG&I • A small proportion of participants in FG & Interviews volunteered that they had multiple problems in their lives, such as other addictions, anger management issues, serious health issues, entanglements in unhealthy relationships, mental health issues, etc. The survey did not gather information in this area.

*Topic: Exposure to Ordinances*

C-FG&I, D-S • Specific survey item findings reveal that fewer survey participants were exposed to the ban (about 32.5%) as compared to about two-third of focus group and interview respondents as assessed via sampling criteria. The survey recruited participants statewide, while the focus groups and interviews recruited from the Twin Cities metropolitan area only, with two-thirds of participants from Hennepin or Ramsey counties (ban communities), and one-third from the 5-county suburban non-ban counties.

Dd-FG&I • Interviews had participants with the highest level of exposure to the ban due to sampling. Two-thirds resided in the Hennepin and Ramsey County ban communities, and all frequented restaurants

monthly or more. Focus groups had two-thirds of participants exposed, but at lower levels overall. The survey had the fewest participants exposed

*Topic: Findings Unique to Surveys - Smoking Outcomes*

- U-S • Comparison of outcomes at 6 and 18 months. The 7-day abstinence intention to treat (ITT) rate is statistically similar at 6 and 18 months.
- U-S • Abstinence rates at 18 months. The 7-day abstinence completer rate is 32.7%. This is due to the 17.7% of completers who stayed quit from 6 to 18 months, plus 14.9% who achieved a new quit from 6 to 18 months, and minus the 6.4% who relapsed from that time period.
- U-S • The relapse rate is 26.3%

*Topic: Support for the Ordinance*

- D • The survey reveals that respondents are about equally likely to support and not support the ordinance, which differs from other population-based studies showing 70% approval ratings. This strengthens warnings to generalize survey results to mostly hard core smokers. The focus groups avoided discussion of support for the ordinance because strong political feelings shifted group dynamics. They created divisiveness and distracted from participants' reporting on the impact of the ban. They created a challenge for all participants to freely express their opinions. Interviews did not result in the same vociferous opinions, perhaps because of less group posturing, more social desirability, and self-selection. Participants against the ban in interviews were less oppositional, likely because of the one-on-one nature of the method. Finally, the goal of the interviews was to gather experiences, which are related to but deeper and more personal than opinions.

*Topic: Travel to Bars / Restaurants in Communities with Different Regulations*

- C • Most respondents do not travel regularly to bars / restaurants with different regulations.
- U-S • Those in ordinance communities were more likely to travel than those in non-ordinance communities. The difference could be confounded by geographic location, as Hennepin and

Ramsey counties are the primary ordinance communities and are proximal to suburban non-ordinance communities, while non-ordinance outstate communities are generally not at all near ordinance communities. This could have been analyzed in focus groups but was not.

- U-S
  - Most respondents report traveling for reasons other than smoking regulations. General travel was not discussed in focus groups or the survey due to the more narrow focus of those methods.
- C-FG&I
  - Focus groups explain that bans have no impact on travel because (1) people already don't smoke when they go out with friends. Interview support that smokers adapt to non-smokers. They don't smoke in front of them either to protect them or hide their habit, or they go to smoke-free bars and restaurants with them and are OK with it. (2) focus groups and interviews explain lack of impact bans have on travel because the inconvenience of traveling outweighs the benefits. Focus groups & interviews explain the findings, survey more precisely catalogs it.
- C
  - All describe the range of travel for ban – occasionally to more regularly.
- C-FG&I
  - Explain the reasons for the range – travel occasionally for special occasions, at the request of friends, or when the mood hits them. Travel regularly with friends, to relax, to not feel like second class citizen.
- C-S&FG,  
Dd-S&  
FG
  - Those who travel to go to smoking communities are more likely to be smokers. Precision of comparison is unique to surveys. The effect in focus groups is more muted: report that smoking status effects travel, smokers travel to smoke, non-smokers travel to avoid smoke.
- C
  - Some smokers travel to smoking communities.
- C-FG&I
  - Some smokers also avoided non-smoking bars because smoking is relaxing to them and being in non-smoking facilities makes them feel like second class citizens. Avoiding non-smoking facilities wasn't specifically asked about on the survey. Interpretive methods explain the behavior.

- C-FG&I • Some smokers also avoid smoking bars during their quit attempt. Most can return to smoky bars, but for some new non-smokers, it is less frequent. This was a common response but not asked about on the survey.
- C • Some non-smokers reported traveling to non-smoking bars.
- C-FG&I • Non-smokers also avoided smoking bars (avoidance not asked about on survey). They avoided smoking or sought out non-smoking because they hate being in a smoky environment, or were actively trying to support their quit and avoid trigger situations. Interpretive methods explain findings.

*Topic: Impact of Cigarette Tax on Smoking*

- C-S&FG • The tax caused some smokers to think about quitting. Survey is more precise, but focus groups are more explanatory, see below. And not covered in interviews because of more narrow focus
- U-FG • The tax made some smokers think about quitting because of financial constraints in their lives – having to choose between cigarettes and shoes for your child, for example.
- C-S&FG • They tax helped some smokers to cut down and buy and smoke fewer cigarettes.
- C-S&FG • The tax had no meaningful impact on many smokers. FG more explanatory, see below.
- U-FG • The tax had no impact because of the power of addiction and they still wanted to smoke. Many also had enough disposable income so that the increase in price did not outweigh the benefits they gained from smoking.
- U-FG • The tax caused some smokers to use coupons, buy a cheaper brand, or travel to another state to buy cigarettes.
- C-S&FG • The tax caused some smokers to quit. The impact of the tax seemed stronger for focus groups than Surveys. Focus groups had explanatory power, see below.



- U-FG
  - Financial constraints in smokers lives made taxes effective in helping them quit. Some had to choose between cigarettes and other necessary household items.

*Topic: Impact of Exposure of Ordinances on 7-day Abstinence and New Quits*
- C-S&FG,  
D-I
  - The survey revealed that there is no significant association between exposure to the ban and 7-day abstinence and new quit attempts. The focus groups revealed that the ban was not helpful and had no impact on abstinence or new quits for some non-smokers and some smokers. The interviews did not produce these conclusions. The focus groups explains some reasons for the lack of impact; surveys provide support for one explanation, and the interviews provide some supporting evidence for others, see below. In interviews over half of participants reported that the ban was an aid to their quitting, including motivating them to quit. Interviews provide deeper underlying reasons for the outcomes they report
- U-FG
  - Explanation for lack of impact (1) focus groups reveal that some smokers and non-smokers hold political beliefs against government interference that makes them less likely to see and articulate the impact of the bans. The higher than average lack of a support for the bans shown in the survey supports that many QUITPLAN participants do not support the bans, likely for political reasons described above.
- C-FG&I
  - Explanation for lack of impact (2). Focus groups reveal that smokers have already adapted to smoking due to non-smoking family and friends. Interviews support this by showing that social norms are changing and smokers are adapting and considering non-smoking family and friends more.
- C-FG&I
  - Explanation for lack of impact (3). Focus groups reveal that seeing and articulating the positive impact of the absence of cigarette smoke is cognitively difficult. Interviews' focus on the smell of smoke produced similar conclusion
- C-FG&I
  - Explanation for lack of impact (4). Both focus groups and interviews show that those who do not support the ban are likely to not be exposed to it.

- C-FG&I • Explanation for lack of impact (5). Especially for smokers, the strength of their addiction outweighed any impact of the ban. The interviews described the apprehension, frustration and even panic that can result in trying to quit and that motivates those trying to quit to start again. The interviews describe how some weather these feelings, only to relapse later.
- U-FG • Explanation for lack of impact (5). Non-smokers proud of their willpower in quitting sometimes had difficulty seeing the impact of factors outside of themselves. Found but not explicated in interviews.
- U-FG • The many barriers to identifying the impacts of the ban suggest that the impact may be underreported.
- C-FG&I  
D-S • Surveys find no impact of exposure on abstinence or new quits. However, both focus groups and interviews found that those who reported no impact were more likely to have less exposure to the ban than those who reported some impact. This suggests that impact of the ban is associated with exposure.
- U-I • Smokers struggle with their cravings and addiction when they are prohibited from smoking due to bans. This can lead to apprehension, frustration and even panic. Some gained fresh insight into the control that nicotine has over them, and were motivated to quit. Additionally, some smokers experienced cognitive dissonance due to the ban because existing knowledge of the harms of tobacco are confronted by a person's active efforts to overcome the inconvenience to smoking that bans cause. This can motivate smokers to quit. Finally, the smell of smoke, especially in enclosed areas like a car, is a powerful motivator for some smokers to quit.

*Topic: Impact of Exposure to Ordinances on Relapse*

- C • The survey finds that exposure to the ban was marginally associated with relapse in a logistic regression ( $p=.061$ ). A bivariate examination shows a stronger, statistically significant relationship. Interviews and focus groups also find that bans help smokers to maintain their quit. In interviews, over half of participants reported that the ban aided their cessation, including helping them prevent

relapse. Focus groups and interviews provide explanatory power, see below.

- C-FG&I
- Both report that bans create conditions where maintaining a quit attempt was easier: they removed the temptation of cigarette smoke from bars and restaurants and at least partially removed the temptation of seeing groups of people, especially friends, smoking.

- U-I
- Interviews reported that the smell of smoke, especially in enclosed areas like a car, is a powerful motivator for some non-smokers to stay quit.

*Topic: Impact of Exposure to Ordinance on Bar / Restaurant Experience*

- C-FG&I
- Bans gave smokers and non-smokers a substantially better bar and restaurant experience because they prefer an authentically smoke-free environment.

- Dd-FG&I
- Interviews gathered more detailed and in depth information about smell and bar restaurant experiences

*Topic: Impact of Exposure to Ordinances on Thinking about Quitting*

- C
- All methods indicate self-report that bans make people think about quitting. Survey were most precise (19.2%).

- C-FG&I
- Focus groups and interviews explain why bans make people think about quitting. Smoke free ordinances motivated some smokers to think about quitting because drinking and smoking in bars is inconvenient.

- Dd-FG&I
- Interviews delve deeper into inconvenience. Inconvenience can be a result of the function of the ban – to enforce no smoking and reduce secondhand smoke from the air. Having to physically move to smoke is inconvenient and can cause cognitive dissonance where smokers’ knowledge of the harm of smoke conflicts with the extra actions they are taking to smoke.

- C-FG&I
- Focus groups and interviews explain why bans make people think about quitting. Smoke free ordinances motivated some to think about quitting because people must leave their friends to drink

and smoke, and being separated from their friends takes the enjoyment out of smoking. This can prompt some smokers to feel guilty about their habit.

Dd-FG&I

- Interviews confirm patterns of hiding & sneaking and report findings of shame & guilt in much greater depth. Feelings of guilt and shame are a central experience for many smokers. Shame is marked by feeling seen “in a powerfully diminished sense where the self feels exposed to both itself and others” (Kaufman, 1992, p. 45).

*Topic: Impact of Exposure to Ordinances on Smoking Less*

C-FG&I

- Both methods reveal that smoke-free ordinances motivated some smokers to smoke less, at least partially due to the inconvenience of having to get up and leave to smoke, especially when there is no drinking outdoors.

U-FG

- Unique focus on practical details: Smoking less was mitigated by a seasonal effect where it is easier to smoke the same amount in the summer because it is warm outside, and you smoke less in the winter because it is cold.

U-I

- Provide more underlying mechanisms. People smoke less in part because they experience cognitive dissonance because their knowledge that smoking is bad for them is confronted by the extra effort they’re choosing to take to go and smoke.

U-I

- Some smokers wish they could smoke less, but cannot. Some of this group experience the ban as a relief, because it is an external support for smoking less.

*Topic: Impact of Exposure to the Ban on Anger*

C-FG&I

- Some smokers are angered and frustrated by bans, because the bans curtail their autonomy to smoke. Many see smoking as a right, and feel resentful that this behavior is controlled by governments. Some react defiantly in a self-admittedly juvenile manner (i.e., I am going to smoke just because you told me not to).

U-FG

- People’s perceptions of their rights differ, some believe they should be able to smoke everywhere,

others in restaurants and bars, and others in bars and bowling alleys.

- U-FG • Smokers' defiance may be a reaction to their frustration of their addiction, which is highlighted by the ban, and which conflicts with their independent nature.
- U-I • Frustrated and angry smokers who are resentful and defiant of the ban defend and treasure their right to smoke in their home, even when doing so worsens smoking-related illnesses.

*Topic: Social Norms*

- C-FG&I • Everyone understands that smoking is harmful and feels social pressure to stop smoking. Respondents acknowledge the harm of secondhand smoke.
- C-FG&I • Some take social norms against smoking personally and feel like they are bad people / they are being told they are bad people because they smoke
- U-FG • Smokers report being glared at, condescended to, judged and treated rudely because they smoke. Media messages and their kids (who get anti tobacco education) also are ways smokers experience social norms.
- U-I • Interviews deepen understanding of guilt and shame from social pressure not to smoke. See other areas.
- C-FG&I • Smokers are developing new social norms about smoking that is considerate of non-smokers. Personally, they pride themselves on not smoking near others, not leaving ashes or butts, etc. They accept going to non-smoking establishments, although they may still hide or sneak cigarettes for themselves. Many don't like to be around smoke themselves. Most importantly, smokers seem to be protecting their loved one – especially children, but also spouses and elderly parents – by making personal bans and not smoking in their houses, cars, etc.
- C-FG&I • The social unacceptability of smoking appears to be a powerful motivator for smokers to quit and consider quitting. This was more directly stated in focus groups, but the interviews revealed this as well through the discussion of shame and guilt in smoking.

- U-FG • Younger smokers appear more amenable to social norm messages than older smokers. More research in this area is necessary.
- C-FG&I • Bans are external enforcement of social norms against secondhand smoke, and support and advance those norms. Bans and social norms against smoking are intertwined.
- C-FG&I • While subtle and not immediate, social norms against smoking appear to be the most promising mechanism to motivate new quit attempts, prevent relapse, and prevent individuals from starting to smoke in the first place.

*Topic: Smell of Smoke / Experience of Lack of Ordinances*

- C-FG&I • Focus groups & interviews reveal that smokers and non-smokers alike hate the smell of cigarettes.
- C-FG&I • Bans benefit all individuals who dislike the smell of cigarette smoke, regardless of their understanding or appreciation of it.
- Dd-FG&I • Interview subjects focused more strongly on the negative aspects of cigarette smoke. Unprompted discussion of the smell of tobacco was nearly ubiquitous, almost all interviewees discussed it. Their reactions were strongly emotional, guttural and unequivocally negative about how disgusting and nasty cigarette smoke was. Smoke was described as a physical experience where it is absorbed through the eyes, abrading the throat and assaulting the senses. Smoke had a weight and abrasion that lasted longer than the time a person was exposed to it. Smoke can even give a person a smoking hangover. Some interviewees described that smoky environments as a temptation, but the vast majority of the discussion was focused on the negative qualities of smoke. On the other hand, in focus groups, there was a greater emphasis on the smell of smoke as a sensation. Still, focus group participants reported better bar and restaurant experiences because of the ban as well.

*Topic: Findings Unique to Focus Groups*

- U-FG • Role of smoking accommodations. Smoking accommodations help smokers feel less like second class citizens. For some, they continue to smoke less. Others smoke just as much, like when drinks

are served on a patio, for example.

- U-FG • Some smokers are motivated to avoid smoking by enclosed smoking rooms. The rooms make the impact of smoking on your body more apparent by concentrating smell, smoke, particulate, etc.
- U-FG • Drinking & Bars. Drinking is a trigger to smoke, and drinking and smoking are activities associated with bars, especially because bars are places where friends gather and socialize, which is also associated with drinking and smoking.
- U-FG • Bars are a common site for relapse.
- U-FG • Impact of bans is due at least in part to relationship between drinking and smoking and bars. Smoking ordinances aid quit attempts and prevent relapse because they at least partially interrupt the association of drinking and smoking in bars.

*Topic: Findings Unique to Interviews*

- U-I • Impact of the Minnesota Clean Indoor Air Act (MCIAA). Almost all participants were exposed to the MCIAA and reported that it reduced their smoking
- U-I • Impact of personal bans. About half of participants reported encountering personal bans, most commonly home and car bans, often because of other family members' insistence. These taught them they could live without a cigarette and most adapted successfully to the change. Smokers who instituted personal bans to help them quit had mixed success because almost all hid or sneaked cigarettes in response.
- U-I • Meaning of the term regulation. The term regulation comprises several meanings that contrast with and confirm one another. Regulation suggests a rule, law or system that when initially instituted causes abrupt change, as well as what is customary or ordinary. Regulation can be understood as the process by which an agent seeks to adapt to a change: either by staying the same or moving towards development, growth and integration. Bans also seek to accomplish a variety of goals such as to equally protect all citizens from secondhand smoke in the workplace and in public places.

- U-I
  - Functions of the ban. To enforce specified smoking behaviors (no smoking in specified areas), and to reduce or eliminate secondhand smoke in specified areas. The bans make smoking inconvenient, especially where drinks are not allowed outdoors, because people must leave friends and their drink to smoke.
- U-I
  - Experience of relief. Some smokers experience relief at a policy that is an external support to smoke less in ways they wish they could change on their own, but cannot. Smokers and non-smokers experience relief at not being exposed to smoky environments.
- U-I
  - Experience of addiction. Face frustration, fear, even anger in face of the ban because prohibited from smoking. Mind works like a track trying to figure out where you can smoke. Change behavior, alter activity to make sure you can smoke. Always thinking about it.
- U-I
  - Experience hiding cigarettes. Hiding and sneaking cigarettes is a central experience for many smokers who cannot smoke in certain places and/or do not want to be seen smoking by others. This can both stem from and lead to feelings of guilt, shame and embarrassment for smoking.
- U-I
  - Experience of guilt and shame. Feelings of guilt and shame are a central experience for many smokers. Shame is marked by feeling seen “in a powerfully diminished sense where the self feels exposed both to itself and others” (Kaufman, 2005, p. 45). It is a common antecedent and consequence of addiction. Regulations may provide new opportunities to experience existing shame.
- U-I
  - Experience of cognitive dissonance. Some smokers experience cognitive dissonance due to the ban because existing knowledge of the harms of tobacco are confronted by a person’s active efforts to overcome the inconvenience to smoking that bans cause. Cognitive dissonance can result in reduction in smoking and efforts to quit.

<sup>1</sup> C=convergence, D=divergence, Dd=divergence by degree, U=unique, S=survey, FG=focus group, I=interview



APPENDIX D: EXTENT TO WHICH SINGLE METHODS

MEET THE JOINT COMMITTEE'S *PROGRAM EVALUATION STANDARDS*

Table D1

*Extent to which Single Methods Meet Utility Standards*

Standard	Survey	Focus Group	Interviews
U1 - Stakeholder identification	Client's need for quantitative evidence was identified	Clients' need to hear the stories of a substantial number of participants was identified	Clients' need to understand the some participants' experiences in depth was identified
U2 - Evaluator credibility	Principal investigator has established credibility with content and method	Same	Same. I have less experience with phenomenological methodology
U3 – Information scope and selection	Addresses primary concerns of the client's Board of Directors and staff	Addresses secondary concerns of the client's Board of Directors and staff	Addresses secondary concerns of the client's Board of Directors and staff
U4 – Values identification	Statistical interpretation was clearly explicated, but the perspectives and rationale underlying it were not	Interpretation based on frequency of comments was clearly explicated, but the perspectives and rationale underlying it were not	The perspectives, procedures and rationale used for interpreting findings were clearly explicated.
U5 – Report Clarity	The report clearly describes the intervention and the purposes, procedures and	The report clearly describes the program and its context, and the purposes, procedures and	The report clearly describes the program and provides extensive information on the

	findings of the evaluation. The context of the intervention is not addressed.	findings of the evaluation.	context of the regulations. The report clearly describes the purposes, procedures and findings of the evaluation.
U6 – Report Timeliness and Dissemination	To enhance utility of the evaluation, significant interim findings were reported orally before the final report was submitted.	Same	Same
U7 – Evaluation Impact	Through addressing travel, the evaluation was planned and reported to encourage stakeholder follow-through	Same	Same

Table D2

*Extent to which Single Methods Meet Feasibility Standards*

Standard	Telephone Survey	Focus Groups	Phenomenological Interviews
F1 – Practical Procedures	Evaluation negligibly disrupted the intervention and client organization	Same	Same
F2 – Political Viability	Staff from different divisions were engaged during planning, implementation and dissemination. The greatest political threat to the evaluation, activists backed by the tobacco lobby, were not engaged.	Same	Same
F3 – Cost Effectiveness	The most expensive method, this provided the information most desired by the Board of Directors	Very modestly priced, this provided information very useful to ClearWay staff.	Very modestly priced, this provided information very useful to ClearWay staff.

Table D3

*Extent to which Single Methods Meet Propriety Standards*

Standard	Telephone Survey	Focus Groups	Phenomenological Interviews
P1 – Service Orientation	The full range of targeted participants were included in the evaluation	The evaluation included a sufficient range of targeted participants	The evaluation included a sufficient range of targeted participants
P2 – Formal Agreements	A formal contract was signed for this method stipulating responsibilities of reach party.	Same	Same
P3 – Rights of Human Subjects	This method adhered very closely to the spirit of the IRB. Permission to use secondary data was obtained from the IRB to conduct the evaluation.	Same	Same
P4 – Human Interactions	Persons associated with the evaluation were treated with respect and dignity to my maximum capacity.	Same	Same
P5 – Complete and Fair Assessment	To my maximum capacity, I assessed the strengths and weaknesses of the intervention with completeness and fairness.	Same	Same
P6 – Disclosure of	Findings have been made	Same	Same

Findings	public within the client organization through presentations, at to the public through conference presentations. Reports have been provided to all interested parties		
P7 – Conflict of Interest	This study presented no conflict of interest	Same	Same
Fiscal Responsibility	Expenditures were accounted for and appropriate.	Same	Same

Table D4

*Extent to which Single Methods Meet Accuracy Standards*

Standard	Telephone Survey	Focus Groups	Phenomenological Interviews
A1 – Program Documentation	The intervention was clearly described and documented	Same	Same
A2 – Context Analysis	The context of the program was described clearly.	Attention to political context was given extra attention, and strengthened the understanding of the intervention.	Same. Attention to individuals' context was prioritized and substantially strengthened the understanding of the intervention.
A3 – Described Processes and Procedures	Processes and procedures were described in sufficient detail to be monitored and assessed.	Processes and procedures for conducting focus groups were described in sufficient detail to be monitored and assessed. The complex analysis procedures are fully understood by obtaining NVivo files.	Same as focus group
A4 – Defensible Information Sources	The sources of information were adequately described to assess the adequacy of information	Same	Same
A5 – Valid Information	Truth value was low; applicability was moderately strong; neutrality was strong	Moderate level of truth value and applicability; neutrality is unknown	Moderately strong truth value and applicability – stronger than focus groups. Neutrality is

			unknown, but confidence in conclusion may be higher
A6 – Reliable Information	Information was planned, implemented and reported to produce reliable inferences, but reliability is unknown	Same	Same
A7 – Systematic Information	Quality assurance processes were used to correct errors	Same	Same
A8 – Analysis of Quantitative Information	Data was appropriately and systematically analyzed using SPSS	Same. Refers to demographic and clinical information on respondents.	Same. Refers to demographic and clinical information on respondents
A9 – Analysis of Qualitative Information	N/A	Data was appropriately and systematically analyzed using NVivo	Data was appropriately and systematically analyzed using NVivo
A10 – Justified Conclusions	Conclusions were explicitly justified against statistical criteria	Conclusions were explicitly justified using categorizations of participants' comments	Conclusions were explicitly justified using categorizations of participants' comments and external sources on shame, addiction, etc.
A11 – Impartial Reporting	I Adhered to my role as external evaluator and standard procedures and practices.	I adhered to my role as external evaluator and used intensive and iterative analysis techniques in NVivo to reduce potential bias	I adhered to my role as external evaluator and used intensive and iterative analysis techniques in NVivo to reduce potential bias
A12 –	Metaevaluation of this method	Same	Same

Metaevaluation is being conducted in this section



## APPENDIX E. EXPERT CRITIQUE OF SINGLE METHODS

## Focus Group Expert Review

**Memorandum**

**April 26, 2007**

**TO: Anne Betzner**

**FROM: Richard Krueger**

**RE: Validation process**

The process you describe in your proposal makes sense. Undoubtedly you know how complex the habit of smoking really is and you will need to be sensitive to the challenges that these smokers are having.

The study looks reasonable and should yield helpful insight. I will comment on only several items where change ought to be considered.

**1. Sampling plan:**

The sampling strategy should be reconsidered. You are assuming that the current ordinances affect only residents of the county. This is likely untrue because residents of one county probably work, eat, drink and travel to other counties. Focus groups typically have homogeneous sampling and there are other sampling strategies that might make more sense. For example, consider separating them by the frequency with which they go to bars and restaurants. Or perhaps QUITPLAN can review past data and use demographics to categorize participants into categories such as: hard core, struggling or success and then see which group is most affected by the program. Then conduct 3+ groups with each category of people.

**2. Include a registration form:**

When participants arrive ask them to complete a short registration form where you ask some background information about their smoking history, their experience with trying to quit, as well as the county where they live, work, eat, drink. Also you might ask: age, gender, and other relevant factors. You don't want to ask these questions during the groups and you may not have this data from other sources, so the best bet is a registration table where they complete the information before beginning the focus group. In addition, this background data might be helpful in coding the responses.

**3. Identifying comments of individuals:**

For most effective analysis I suspect that you will be coding the responses for each individual in the focus group. You probably won't be able to do this by merely listing to the audio recording! You need a clear strategy for how to do this. For example, you will need a note taker who captures the first few words of each respondent and then later adds these names into the transcript. I've done thousands of focus groups and the following day when I transcribe I cannot accurately identify individual speakers. You need a strategy.

**4. Focus group questions:**

It is difficult to critique questions because you just won't know how well they work until you actually use them. But I would suggest some changes in the question sequence. Some of the questions you have should work well, but I would begin by asking a question where they indicate their current status on the path to quitting. Perhaps you show a continuum and ask them to identify where they are now, and where they were at some point in the past.

Then, if there is movement on the scale, ask what factors contributed to this change. Perhaps these are listed on a flip chart and the group discusses these factors.

Then move to QUITPLAN and ask about the influence of QUITPLAN. You can't assume that QUITPLAN alone made the change or even most of the change. You might ask the degree to which, or the amount of influence that QUITPLAN had.

I have no idea what the ingredients of QUITPLAN are, but likely some will be common sense steps that have been used for decades. Consider not asking about QUITPLAN and instead ask about the key ingredients of QUITPLAN.

I would include one or more questions that ask the respondents to give you a number or score indicating the strength of various factors that influenced their change. Words alone are imprecise here and a number is a better indicator. For example, if a person says that he / she has dramatically reduced smoking in the past year, you might ask. What were the factors that influence you? And what percent of change would you give to each factor? (I realize that this is mentally difficult and for some respondents it might not even be logical.) Yet, consider something that will get them to tell you the factors that influenced their behavior and then the strength of each of those factors.

Good luck with you study.

## Phenomenological Interview Expert Review

### External Review of the Research Methods

Noriko Ishihara <[ishihara@american.edu](mailto:ishihara@american.edu)>

American University

May 2007

Phenomenological methods being used in Betzner's study derive from a research tradition in which the essence of a phenomenon is sought in order to find meaning in the lived experience of the phenomenon. The underlying philosophy of this research tradition is that knowledge of a phenomenon (which often is the phenomenon that one has no first-hand experience of) can best be obtained by capturing it as closely as possible to the phenomenon. Husserl, the founder of the phenomenology, argues that this can best be accomplished by tapping into the knowledge of those who have experienced it directly without theorizing it. The central advantage of the study (as listed first below) is this directness to the participants' lived experience that the research methods bring. The other strengths of the research stem from the careful design and implementation of the study and researcher's tact in interviewing the participants.

- Description of direct and concrete lived experience of the phenomenon under investigation
- Comparative view gained through the stratified sampling (i.e., full, partial, and no ordinance conditions, and current and former smokers)
- Anonymity enhanced through the nature of the phone interviews and potentially less inhibition on the participants' part in describing their experiences (This point can greatly enhance the credibility of the data.)
- The sense of trust, warmth, and safety established through the researcher's conscious effort (e.g., verbalizing the non-verbal cues)
- Enhanced credibility of research through the field-testing and refining the selection of the interview questions
- Enhanced credibility of the research through the employment of *reduction* (i.e., suspending the researcher bias and preconceptions)

No research study is without disadvantages. Potential methodological weaknesses of this particular study include:

- Potential of not capturing the full range of the essence of the phenomenon due to the stratified sampling (Two participants in each category in the interview sampling may not be sufficient to demonstrate variant and invariant essences.)

- Less direct means of communication through the telephone (However, benefits of the telephone interviews discussed above appear to outweigh drawbacks.)
- Limitation of human consciousness, participants' limitation in ability to articulate only through verbal channel in a concrete, detailed manner (The assumptions listed on p. 6 seems optimistic, considering the demand of the phenomenological research methods.)

Another point that I am unable to assess from the given information is the effectiveness of the qualitative software program (NVIVO) employed for the “identification and transformation of meaning units” (p. 6). The way in which the program transformed the meaning units appears to remain unclear and warrants further information.

## Telephone Survey Expert Review

**Memorandum****August 27, 2007****TO: Anne Betzner****FROM: Michael G. Luxenberg, Ph.D.****RE: Validation process**

First, a disclaimer. As this was part of a Professional Data Analysts, Inc. funded study, I did have some input into the design. However, Ms. Betzner has been the principal investigator for this study.

This is a rigorous quantitative and qualitative evaluation using large sample sizes. It has a well thought out sampling plan, a very thoroughly designed follow-up survey, and carefully crafted methods of recruiting subjects to participate. A lot of effort was spent designing the consent protocol, as well as the training of program staff to get consent and respectfully encourage participation in the follow-up. High response rates were critical to the validity of the data captured in this study.

Since this study by its nature is based on an evaluation and not a controlled clinical design, there are certain design limitations inherent in this type of study. Most of these limitations are clearly delineated in Ms. Betzner's description. The study participants represent a very select population of smokers and former smokers who tried to quit smoking and contacted QUITPLAN for their services. Results cannot be generalized to Minnesota smokers as a whole. Participants obviously self-selected to be part of this study and obviously were not randomized to any particular ban condition, community or even the type of QUITPLAN intervention they selected. Therefore, when using least square models non-randomization must be take into consideration in its interpretation and generalization. Further, subjects had to provide consent to be followed-up and also to have their results aggregated for possible publication. So despite its high response rate, Ms. Betzner had to take this into consideration in her responder and consent bias analysis. There is also a confounding of ban type (full, partial or none) and community living in and geography (Twin Cities Metro, % County Metro and Outstate).

Also, unanticipated changes in the ordinance and those communities covered by them during the course of the study added to the confounding. The small number of partial ban communities required that the participants in these be grouped with those from full ban communities for analysis. Finally, border communities both within and between

states with differing ban ordinances are difficult to control for in terms of how much those people are exposed to the ban. Sample size was too small to assess the effects by QUITPLAN program type.

Nevertheless, this was a very ambitious study that provided a unique view and insight into smokers trying to quit or maintain a quit over an 18-month period with the ability to assess the impact of exposure to the ban, and the impact of the excise tax that went into effect during the study period. A side benefit of this study was the ability to estimate quit rates over an 18 month follow-up period with a very reasonable response rate. Despite some of the confounding of communities and ban type, it was possible to develop an "exposure to ban" variable that was able to mitigate the effect of some of the confounding variables mentioned.