Planned Missing Data Designs in Communication Research

A dissertation presented to

the faculty of

the Scripps College of Communication of Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Michael M. Parsons

December 2013

© 2013 Michael M Parsons. All Rights Reserved.

This dissertation titled

Planned Missing Data Designs in Communication Research

by

MICHAEL M. PARSONS

has been approved for

the School of Communication Studies

and the Scripps College of Communication by

Amy E. Chadwick

Assistant Professor of Communication Studies

Scott Titsworth

Dean, Scripps College of Communication

ABSTRACT

PARSONS, MICHAEL M., Ph.D., December 2013, Communication Studies

Planned Missing Data Designs in Communication Research

Director of Dissertation: Amy E. Chadwick

Prominent among the many methodological challenges communication research faces are the relative lack of longitudinal research conducted in the discipline and the threats to validity that arise from the complex instrumentation necessary for inquiry into human interaction. This dissertation presented planned missing data designs (PMDs) as solutions to these challenges because PMDs can make research less burdensome, cheaper, faster, and more valid. Three studies illustrate the use of PMDs in communication research.

Study one was a controlled-enrollment PMD investigation of the relationship between students' public speaking anxiety and communication competence in a semester-long longitudinal study. By using the controlled-enrollment design, this study had five measurement waves, but each participant was measured at no more than three measurement waves. Results indicated that the controlled-enrollment design was effective at minimizing participant loss due to attrition and reducing the risk of testing effects due to repeated measurements.

Study two was an efficiency-type PMD replication of Infante and Wigley's (1986) verbal aggressiveness scale validation study, in which each participant was presented with only 95 items from the 147 item survey instrument. Through the use of an efficiency design, this study was able to replicate the results of the original study with a

dramatically reduced time burden on the participants, indicating that efficiency-type PMDs are an effective tool for scale shortening.

Study three was an accelerated longitudinal PMD replication of Rubin, Graham, and Mignerey's (1990) longitudinal communication competence study, which measured change in students' communication competence over the course of a college career. Through the use of an accelerated longitudinal PMD, data collection was completed in just over one calendar year, far shorter than the three years the original study took to collect data. A flaw in participant retention procedures prevented data analysis from being conducted, but this study did effectively illustrate the increased methodological complexities caused by PMDs.

This dissertation concludes that PMDs can be of substantial benefit to communication research, and should be adopted in the discipline. Special attention must be paid, however, to the increased design complexity added by the use of these methods.

DEDICATION

For Genevieve Monteverde.

Because it will make her smile.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Amy Chadwick, for your guidance, patience, and support in turning a rough idea into an actual finished work. I am a better writer, thinker, and scholar because of the time and attention you have put into this dissertation. I would like to thank my committee members: Dr. Scott Titsworth, for helping me take earliest spark of a question and turn it into something worth developing, and for knowing exactly who I needed on my committee to guide me through this process; Dr. Carson Wagner, for having big ideas, and making me come back up out of the rabbit hole to wrestle with them; Dr. Gordon Brooks, for being a skeptic and making me convince you. To all four of you: you were exactly the committee I needed to get this project done.

I would also like to thank all of the friends and colleagues in the COMS community, who over the past four years have been a constant reminder of why I came here in the first place. In particular, I want to thank Drs. Laura Black, Raymie McKerrow, and Ben Bates for being mentors during my time in COMS, and Heather Grove and Gayle McKerrow for always knowing the answers to my questions.

I cannot underestimate the help I have gotten from my family. My mother and father, Jean and John, and my brother, Kevin, have all been unending in their support and patience for me. Having such a supportive family has made all the difference in the world.

I want to thank the good friends I have made here. Sarah, Christian, Steve, Anna, Adriana, Jessica, each one of you has, in your own way, helped me finish this dissertation, and I am forever thankful for being your friend.

And Jeff, well, you are truly the best friend I could ever have asked for. Has it really been 13 years already? Thanks for the laughs, the advice, and for sharing the best office ever. And yes, now that this is done, maybe we could finally dust off the Xbox.

Finally, I need to thank Jennifer, my lovely and talented wife. She is the love of my life, and the rock that keeps me sane. She is as responsible for this getting finished as I am.

TABLE OF CONTENTS

Page

Abstract iii
Dedication v
Acknowledgments vi
List of Tables xiii
List of Figures xiv
Chapter 1: Introduction 1
Defining the Challenges 1
Longitudinal Research 2
Long Survey Instruments 12
Planned Missing Data Designs 14
Chapter 2 – Review of Literature 16
Types of Missing Data 18
Strategies for Handling Missing Data 20
Strategies that Impart Error or Reduce N 21
Study Designs with Planned Missing Data 24
Accelerated Longitudinal Designs 25
Types of Accelerated Designs 28
Advantages Conferred by Accelerated Longitudinal Designs 32
Circumstances not Suited for Accelerated Designs 36
Efficiency Designs 39
Between-Block Split-Form Design 40
Within-Block Split Form Design 41
Controlled-Enrollment Longitudinal Design 42
Advantages Conferred by Efficiency Designs 44
Hypotheses 49
Goals Related to PMDs 49
Chapter 3: Study 1 - Public Speaking Anxiety 52
Hypotheses 53
Study Design 53

Sample Size 53
Planned Missingness 56
Recruitment 57
Participants 57
Procedures 58
Scale Reliability 61
Measures 62
Data Analysis 65
Study-Specific Results 71
Preliminary Analysis 71
Substantive Analysis 75
PMD-Related Results 77
Discussion 78
PMD-Related 79
Additional Findings 83
Limitations 84
Conclusion 87
Chapter 4: Study 2 – Verbal Aggressiveness Scale 89
Introduction 89
Hypothesis 90
Study Design 91
Sample Size 91
Recruitment 92
Participants 92
Procedures 93
Measures 94
Planned Missingness 98
Block Design 98
Simulation Comparison 100
X Set 102
Data Analysis 103
Subsamples 105

Results 107
Study-Specific Results 108
Planned Missing Data Results 112
Discussion 117
Introduction 117
Summary of Findings 117
Reduction in Participant Burden 118
Multiple Imputation Process 120
Replication Findings 121
Summary 122
Limitations 123
Conclusion 124
Chapter 5: Study 3 - Accelerated Longitudinal Communication Competence 126
Introduction 126
Hypotheses 126
Study Design 127
Sample Size 127
Recruitment 128
Participants 130
Procedures 132
Measures 132
Planned Missingness 136
Data Analysis 137
Results 137
Analysis 142
Discussion 145
Chapter 6: Discussion 147
Finding One 148
Finding Two 151
Finding Three 153
Strengths, Limitations, and Future Research 156
Strengths 157

Limitations 157
Future Research 159
Conclusion 163
References 165
Appendix A – Study 1 - Recruitment Emails 186
Recruitment Email for Participants at First Data Collection 186
Recruitment Email for Participants at Second Data Collection 187
Recruitment Email for Participants at Third Data Collection 188
Appendix B: Study 1 - Consent Form 189
Appendix C: Study 1 – Measures 192
Personal Report of Public Speaking Anxiety 192
Self-Perceived Public Speaking Competency Scale 194
Appendix D: Study 2 – Consent Form 196
Appendix E: Study 2 - Measures 199
Verbal Aggressiveness Scale 199
Argumentativeness Scale 200
Subscales of the Hostility-Guilt Inventory 202
Social Desirability Index 203
The Janis & Field Personality Questionnaire 205
Personal Report of Communication Apprehension 207
Appendix F: Study 3 – Consent Form 209
Appendix G: Study 3 – Recruitment Messages 213
Initial Recruitment Tool 213
Pre-Time Point #3 Recruitment Tool 215
Appendix H: Study 3 – Measures 217
Interaction Involvement Scale 217
Personal Report of Communication Apprehension-24 218
Communication Competency Self-Report Questionnaire 220
Previous Experiences Scale 221
Appendix I: SPSS Recoding Scripts 223
Study 1 – Parcel and Augment Creation Script 223
Study 2 – Item Recoding and Scale Creation 224

Study 3 – Parcel and Augments Creation Script 225
Appendix J: Mplus Scripts 228
Study 1 – Imputation Syntax 228
Study 1 – Analysis Syntax 229
Study 2 – Imputation Syntax 232
Study 2 – Analysis Syntax 232
Study 3 – Imputation Syntax 233

LIST OF TABLES

Page

Table 1: Accelerated Longitudinal – Example Measurement Pattern 34

Table 2: Accelerated Longitudinal – Example Cohort-Sequential Pattern.....35

Table 3: Data Collection Schedule for Controlled-Enrollment Design 60

Table 4: PRPSA Coefficient H Scores.....63

Table 5: SPPSC Coefficient H Scores 64

Table 6: Means and SDs of Days between Survey Completion and Speech Delivery...65

Table 7: Correlations between Verbal Aggressiveness and Comparison Variables.....109

Table 8: Repeated Measures ANOVA Results Table for Scale Variables 114

Table 9: Change in Standard Deviation as Sample Size Increases.....116

Table 10: Alpha, Mean, and Standard Deviation scores for the PRCA.....133

Table 11: Alpha, Mean, and Standard Deviation scores for the IIS 134

Table 12: Alpha, Mean, and Standard Deviation scores for the CCSR.....135

Table 13: Alpha, Mean, and SD scores for the Previous Experience Scale 136

Table 14: Number of Participants by Measurement Wave and Grade Year.....140

LIST OF FIGURES

Page

Figure 1: Accelerated Longitudinal Study, Measurement Design.....30

Figure 2: Reconstructed Accelerated Longitudinal Study Design.....31

Figure 3: Reconstructed Accelerated Longitudinal with Confounding Event.....37

Figure 4: Initial Hypothesized Structural Equation Model 74

Figure 5: Final Trimmed Structural Equation Model.....75

CHAPTER 1: INTRODUCTION

Quantitative communication research faces real challenges, and yet there are new and powerful tools available which can improve the quality of research produced within the discipline. The goal of this dissertation is to present a set of those tools, a family of research designs which employ a planned missing data design strategy, and which have the potential to significantly improve the way quantitative communication research is conducted. The long-term goal of this study is to introduce PMDs into the canon of quantitative designs used in communication studies.

The first chapter in this dissertation will present the status quo in communication research, explain why the status quo is inadequate to meet the needs of our research goals, and briefly introduce planned missing data designs as a powerful alternative to current research designs. Later chapters will present planned missing data designs in much greater depth, and will present three studies that demonstrate how planned missing data design principles can be applied to communication research.

Defining the Challenges

Communication research is currently facing two major methodological challenges. First, we need to do more longitudinal research. Communication is a discipline that researches communicative processes and interactions between people. To accomplish this, we need to be properly equipped to conduct studies that allow us to account for time in the interaction. However, as the review of literature presented below will show, this is a need that is well known, but not as well met in the literature.

Second, communication research deals with complex processes that frequently involve complex instrumentation. Unfortunately, as the length of survey instruments increases, eventually so too does the threat of reduced validity as a result of the instrument length (Andrews, 1984). As the review of literature below will show, the options for dealing with this challenge that communication researchers currently have to choose from are not always good ones, and sometimes threaten validity themselves.

Longitudinal Research

In the 2007 Presidential Address to the International Communication Association, Jon F. Nussbaum discussed what he considered to be one of the central challenges in communication research:

> I do not think it would be an overstatement to assert that the great majority of communication scholars have attempted to incorporate the importance of process, though not necessarily change, into their theories or descriptions of communication but fall far short of designing studies to test any notion of change or process. Once the data analysis is complete, though, we seem to have no problem discussing the results of our theory building as if we have tested for change or process. As a collection of scholars, we have been talking a very good game of process for over 45 years but, with very few exceptions, have not followed our process talk with much empirical evidence documenting that process (Nussbaum, 2007, p. 4).

Although these comments were couched in a presentation about communication about the lifespan, they were directed at the discipline as a whole. Nussbaum (2007) noted that

> The great majority of our research designs and statistical procedures are in no way appropriate to capture change across time…To be rather straightforward, to move beyond simple descriptions of differences or correlations and to capture meaningful change across time, one needs to move beyond one-shot survey/experimental designs or cross-sectional designs and conduct longitudinal investigations (p. 5).

Longitudinal Research in the Literature

Nussbaum's (2007) assertions that communication exhibits a disconnect between the rhetoric and the reality of longitudinal research is supported by a review of the highest-impact quantitative research journals in communication. Although longitudinal research is regularly done in some areas of communication (such as political campaign research, health campaign research, or media effects research), they are an infrequently used method for a discipline concerned with interactions and message exchange. This problem can be seen in the major journals, which regularly publish scholarly works that call for their particular area of research to be extended longitudinally, and do so in numbers that outweigh the amount of longitudinal research actually performed.

I conducted a thorough review of three leading outlets of quantitative communication research: Journal of Communication, Human Communication Research, and Communication Research, from January 2007 until mid-2012. In sum, across the three journals I reviewed during the relevant time period, there were a total of 530 scholarly works published. In that time, approximately 11% of works could be categorized as longitudinal. This includes any works that involve the element of time and are not simply cross-sectional in nature. Some of these were pre-test/post-test designs

(e.g., Moyer-Gusé & Nabi, 2010), while others were described as full longitudinal designs with multiple time points and an extended chronological design (e.g., Anderson et al., 2008; Peter & Valkenburg, 2009; Valkenburg & Peter, 2009).

The literature indicates that scholars regularly point out that their work should be extended longitudinally, and they do so with a frequency disproportionate to the amount of longitudinal research that is actually conducted. Sometimes these scholars are suggesting that this longitudinal extension is the next logical step in the development of their ideas. For example, Hargie, et al. (2008) concluded that their "findings highlight the need for further research in this area, and in particular longitudinal studies, to investigate the precise effects" (p. 813) found in their study.

In other cases, scholars are pointing out that their ability to make any causal claims is inhibited by the cross-sectional nature of their design, and that to make their claims fully would require longitudinal work. For example, as Warren, Hecht, Jung, Kvasny and Henderson (2010) pointed out, one limitation to their study was that "a cross sectional self-report study has inherent limitations in testing a causal model. Future longitudinal studies, including those with experimental designs, are needed to establish causality and refine these conclusions" (p. 694-695).

One common subject for longitudinal work is health communication, where interventions and campaigns require the involvement of time as a variable (e.g., Anderson, et al., 2008). Another is media effects, where a frequent operating premise is that there is a long-term, accrued impact of media consumption on behavior and cognition (e.g., Aubrey & Taylor, 2009). However, there are many topics that are not such obvious candidates for being tested in a longitudinal research setting. Examples of

these types of topics include decision-making (Henningsen & Henningsen, 2007), the effect of technology on argumentation-style (Ellis & Maoz, 2007), the addition of voice to online gaming interactions (Williams, Caplan, & Li, 2007), or a cross-cultural comparison of messages in favor-asking interactions (Lee & Park, 2011). Obvious or not, these topics do warrant testing in a longitudinal context.

Whether the operating metaphor for communication is the transmission model, of communication as information delivery, or the social constructionist model of communication as creating reality, research should acknowledge that communication is a process that involves time. Therefore, any investigation of a communicative phenomenon could potentially investigate how that phenomenon occurs over time. With this in mind, whenever a scholar notes that future longitudinal research is needed in a particular area, they are acknowledging not only that this longitudinal research could add to the knowledge in their particular research context, but also that the current study was unable to fully explore what was happening in the research situation. Thus, when scholars suggest future longitudinal research, they are indicating a missed opportunity, a place where current research leaves us with a less-than-ideal understanding of the communicative situation.

What is particularly concerning is the number of times published works make repeated calls for longitudinal research, which indicates a disconnect between the reality of conducting longitudinal research and the ideal of how frequently it should be done. As an example of this disconnect, consider one particular series of publications, all discussing a common subject (antismoking media campaign efficacy and social influence), with a common lead author, which was published in three successive years

across three journals, using four separate data sets (Paek & Gunther, 2007; Paek, 2008a; Paek, 2008b; Paek, 2009). This author is clearly establishing an effective program of research, yet each publication states that in the future, this line of research should be conducted longitudinally. The first study (Paek & Gunther, 2007) cautions against too-strong an interpretation of the causality of the model, noting that "panel data should clarify this causal direction more clearly in future studies" (p. 425). The next article in this sequence (Paek, 2008b) suggests that "future studies that attempt to examine campaign effects should include appropriate measures of specific antismoking campaigns (e.g., recognition and recall) with longitudinal data for more rigorous causal inferences" (p. 535). The third article (Paek, 2008a) notes that "as with any cross-sectional survey study, some attention is due to the causal direction of the direct and indirect relationships in the final model…panel data should verify the causal directions more clearly in future studies" (p. 98). Finally, the fourth article in this program of research (Paek, 2009) notes the limited ability of a cross-sectional study to make causal claims, saying that "future studies should develop a more rigorous causal model that integrates various peer perceptions. Analysis of longitudinal data could identify a clearer causal direction between peer perceptions and individuals' own behavior" (p. 451). Clearly, the need for longitudinal research on this particular subject is understood, but it is only after four years of stating the importance of longitudinal research for addressing the ability to test these hypotheses that this longitudinal research eventually happens (Paek, Reid, Choi, & Jeong, 2010).

This example is not a criticism of the particular scholar. Indeed, Dr. Paek is by no means the only author to call for future longitudinal research repeatedly, even within this

sample of articles (e.g., Schrodt & Ledbetter, 2007; Ledbetter, 2009; Ledbetter, et al., 2011; Ledbetter & Kuznekoff, 2012; Covert & Dixon, 2008; Zhang, Dixon, & Conrad, 2009), and unlike many other lines of research, Dr. Paek's series of articles ultimately includes a longitudinal study. Rather, Dr. Paek's research is an example of a disconnect in communication between the rhetoric and the reality of longitudinal research. When authors repeatedly note that their work would be stronger if it was tested longitudinally, but do not as often do that longitudinal research, there is a challenge that needs to be addressed.

Barriers to Longitudinal Research

If this need for more longitudinal research is clearly understood in the discipline, but the actual research is not nearly as forthcoming, then there must be real, non-trivial reasons for the disparity between the calls for longitudinal research and the reality of conducting longitudinal research. Nussbaum (2007) posits three causes for this relative dearth of longitudinal research in communication. Specifically, he notes that longitudinal research requires (a) "rather massive grants," (p. 5) (b) "career-long dedication" (p. 5), and (c) a knowledge of statistical analyses that are too new to be known by most established scholars and too advanced to be taught to most current graduate students. Each of these three reasons: cost, time and training, are discussed in detail below.

Cost. One significant barrier to longitudinal research can be cost, as the expense incurred for administering a cross-sectional study must then be multiplied by the number of measurement points in a longitudinal design. In Rubin's (1982, 1985) Communication Competency Assessment Instrument, which is one of the measures employed in Rubin, et al.'s (1990) study of longitudinal communication competence, four trained judges each

watched a videotaped speech from each study participant, and rated the speech on each of 57 separate assessments. Viewing a 3-minute speech, followed by a 57-point assessment, could take as long as 15 minutes. The aggregate N over all time periods in the Longitudinal Communication Competence study (Rubin, et al., 1990) was 127, suggesting that each judge spent approximately 30 hours rating speeches. In addition, the CCAI takes "approximately one-half hour per student to administer" (Rubin, 1982, p. 24), time which needs to be supervised by a (presumably paid) assistant, and the potential budget for this study runs in the range of about $2,000, assuming $10/hour for paid assistants & judges. And that is a conservative estimate of study cost, one that could run much higher if the number of participants had not decreased greatly over the course of measurement points. One of the self-noted problems with that study (Rubin, et al., 1990) was that they encountered significant attrition over their design. Had they not lost a single participant – maintaining an N of 50 for each measurement point – they would have had 200 speeches to assess, and the total cost for the study might have reached as high as $3,000. Finally, theirs was not a large study in terms of number of participants. Rubin, et al.'s (1990) initial N was 50. Because costs would increase linearly with N, had the authors determined that a power analysis required they start with an initial N of 200 (not an unusually large N for a study in communication), their total study budget could have been as high as $12,000.

For many scholars, that kind of money may not be in the budget. Unfortunately, the costs to conduct high-quality research apply widely. Sundar (The National Youth Anti-Drug Media Campaign, 1999) estimated the costs of conducting a three-group, 65-participant, two-measurement experiment for a media effects based master's thesis at

$10,000. Fortunately, planned missing data designs present a way to reduce cost, and are broadly relevant to the entirety of communication research.

Time. Longitudinal studies inherently take longer to conduct than cross-sectional designs, because the measurement phase itself involves the passage of time. Using as an example the longitudinal study replicated in this dissertation (Rubin, et al., 1990), we can compare the time required to conduct the original study to various time-frames in the career of a scholar, and see the difficulty with conducting longitudinal research at various points. The original study collected data from college students early in their first year, and then at the beginning of each of the subsequent three years. The total time for data collection was just over three calendar years.

The long duration of longitudinal studies make them challenging for all scholars to conduct, but particularly challenging for graduate students and early-career faculty. Consider a doctoral student in a traditional four-year program. Depending on the school, intensive planning for the dissertation may begin immediately, but it is more likely that most doctoral students start to design their dissertation study at some point during their second or third year in the program. A well-organized and ambitious student might begin data collection at the start of year two. However, even that student is not capable of doing a three-year longitudinal study and still graduating on time. Given the pressures of doing a dissertation, it is unlikely that a doctoral student would undertake an additional three-year non-dissertation study, in which they (presumably) would leave the institution midway through to take a job somewhere else.

The same problems exist for the early-career scholar. If we assume that it takes several months to design a study and get IRB approval, three calendar years to collect

data, several months to write the resulting article, and at least a year between initial submission of the article and publication, it may take upwards of 5 years for a publication to emerge from a three-year longitudinal study. Conceivably, a young scholar in their first or second year might design a study that is eventually published, but not in time for *their tenure submission*. Given the pressures of publishing for tenure, it is easy to understand why the challenges of longitudinal research make cross-sectional or short-term experimental studies more appealing in terms of time.

*Training***.** Communication is caught in a predicament with regards to the teaching of advanced longitudinal methods. Fewer established faculty know these methods, so they are less likely to teach them to the current generation of graduate students. Thus, the communication graduate students who are most likely trained in longitudinal methods are the ones who are committed enough to search out other academic units on campus who *do* teach such courses, frequently educational research or quantitative psychology departments.

I reviewed quantitative methods course descriptions from an informal survey of twelve top communication doctoral programs. What I found supports Nussbaum's assertion about the lack of training in longitudinal designs and indicates that longitudinal designs do not hold the same place of importance within our methodological curricula as do, for example, correlational designs, factor analyses, nonparametric analyses, multiple/logistic regression, and (M)ANOVAs. Obviously, course descriptions may not cover all the topics that are actually taught in a course, but it is worth noting that whereas there were numerous mentions in the descriptions of the methods listed above, there was exactly one mention of any specific training in longitudinal designs (Monge, 2011). It

appears that methodological training may be one cause of the lack of longitudinal research in communication. Of course, this could be a classic chicken-and-egg situation. Few graduate programs effectively teach skills to conduct sophisticated longitudinal analyses, and so there are relatively few people using them in practice. Since there are relatively few longitudinal studies in communication, longitudinal analyses appear to be less critical to teach in graduate programs.

There may well be other reasons for the existence of this disconnect as well. For example, the confines of APA writing style demand that authors hedge their causal claims (Madigan, Johnson, & Linton, 1995) and include a statement about future research. It may be that many authors have simply gotten into the habit of hedging their own claims by pointing out the limited ability of cross-sectional research to make causal claims, for which a logical follow-up in the future research section would be a longitudinal study, thus leading to many completed cross-sectional studies and many suggested longitudinal ones. Although the calls for longitudinal research may be more automatic than sincere, nevertheless they demonstrate the need for and important role of longitudinal research in communication scholarship. PMDs can make conducting these longitudinal studies cheaper, quicker, and more valid through a reduction of participant burden.

While planned missing data designs cannot directly contribute to the improved training of quantitative scholars in communication, I hope that this dissertation can make more advanced research designs more accessible, which might in turn encourage them to be taught to current and future scholars.

Long Survey Instruments

In addition to the need for longitudinal research to examine the process of communication and change over time, communication researchers often face the need to employ lengthy survey instruments in their research. This may be caused by using many scales to measure covariates, modeling a complex interaction, or it may be a case of using few, but long, scales. There comes a point of diminishing returns, where the longer a survey instrument becomes, the smaller the likelihood of obtaining quality information from a participant. Longer surveys are less likely to be started by a participant in the first place, and if started, less likely to be completed (Galesic & Bosnjak, 2009). Longer surveys are associated with higher rates of nonresponse (Heberlein & Baumgartner, 1978; Yamarino, Skinner, & Childers, 1991; Crawford, Couper, & Lamias, 2001) and are more likely to induce bias in the form of straight-line responding (Herzog & Bachman, 1981) and satisficing (Krosnick, 1991).

This can be partially ameliorated through the use of multiple, re-ordered forms of an instrument (Herzog & Bachman, 1981), but the net effect in both cases is still a reduction in validity if the instrument is too long. The reduction is just spread across all of the measures instead of concentrated in whichever measure comes last in order. Still, these approaches all ultimately lead to a less valid instrument. Currently, there are three ways to deal with this design problem. One option is to go through the full process of designing and then validating a short form version of an existing scale. Another, less time-consuming choice is to create an ad-hoc short form version of the existing scale for use in the new study. Finally, there is always the option to use the full, original scale, and accept the risk of threats to construct validity that come along with a too-long survey.

Current Solutions

The first potential solution to this predicament is the creation and proper validation of a short form version of an existing scale (e.g., McCroskey, 1978; Hensley & Batty, 1974; Cheney, 1983; Marteau & Bekker, 1992; Strahan & Gerbasi, 1972; Wrench, 2005). This is done in the hopes that a shorter scale will measure the same construct while placing a lower response burden on the participant. This can reduce administration time, which can make the collection of data a cheaper, less complicated task. A short form instrument can be just as effective as the original if it is fully validated in the same way that a normal scale is validated. Short-form instruments may be used to dramatically shorten a truly unwieldy measure (Kelly & Keaten, 2007) and, if created from a truly unidimensional measure, may be isomorphic to the original (Maloney, Grawitch, & Barber, 2011). Designing and validating short-form scales is a difficult process, one which may require a significant time commitment, and there is the potential to create a scale less valid or reliable than the original. For example, consider the warning that accompanied the short-form of the original PRCA measure: "[The] short form of the PRCA appears to be a useful instrument when time constraints do not permit employing the longer form. However, since reliability and precision are reduced by the use of the short form, the long form should always receive preference" (McCroskey, 1978, p. 203).

The second potential solution is to create an ad-hoc short form, without any rigorous re-establishment of validity (Lapinksi & Orbe, 2007; Kerkhof, et al., 2011). The use of an ad-hoc short form scale forces the readers of the study to make a judgment, one that they are often unequipped to make, about the impact that altering the original scale has had on the validity of the remaining items. Effectively, after choosing a subset of

items from the original scale, the researcher draws conclusions from this subset of items as if they have perfect item-total correlations to the full scale, which is a highly unlikely scenario. Assuming the chosen items do *not* have perfect item-total correlation, than the ad-hoc short form measure is measuring something different than the original measure.

Sometimes, depending on the demands of the particular research context, extremely short forms, sometimes as short as one- and two-item measures, are employed, and these types of instruments pose analytical challenges. When time demands are imposed on a study such that "single-item measures and shortened scales [are] used because of time limitations imposed by the schools and the specific age group" (Kam, 2011, p. 470), the first instinct may be to use an extremely short form. This use of single- (or even double-) item scales poses a serious analytical problem beyond the validation problems posited above, because the use of any SEM-family analyses are severely curtailed when there are fewer than three items for any given construct included in the measurement model, as it results in a locally underidentified model (Little, Cunningham, Shahar, & Widaman, 2002). Ultimately, the use of such measures negatively affects the sophistication of analyses possible.

Planned Missing Data Designs

A family of research designs called planned missing data designs (PMDs) have the potential to allow researchers to conduct longitudinal designs in shorter timeframes and with less cost as well as allow researchers to use long survey instruments while maintaining construct validity and reducing error. PMDs can be broadly categorized into two types of designs: efficiency and accelerated longitudinal. In an efficiency design, questionnaire items (or entire collection time points) are intentionally not presented to

certain participants to reduce cost and burden. In an accelerated longitudinal design, data are collected from multiple cohorts simultaneously, leading to a study design that can draw conclusions about change over a longer period than it took to collect the data. These designs are more fully explored in the next chapter.

CHAPTER 2 – REVIEW OF LITERATURE

Communication research often requires longitudinal designs and lengthy survey instruments to appropriately study the influential factors involved in the communicative process. However, longitudinal designs present challenges in terms of the time, cost, and training required. Similarly, long survey instruments present challenges of increased respondent burden and decreased validity. Planned missing data designs provide an opportunity to conduct longitudinal designs and studies using unusually long instruments at reduced time, cost, and participant burden while maintaining validity. It is clear that there is an opportunity to improve the way that quantitative data collection is conducted in communication. Planned missing data designs can provide that improvement. To illustrate the ways that planned missing data designs can help improve communication research, the following chapter will detail the types of designs that comprise the family of planned missing data designs and how they can be leveraged to improve communication research.

Research designs that employ a planned missing data design strategy have the potential to significantly improve the way quantitative communication research is conducted. These designs, which allow researchers to increase the efficiency of their research design in terms of time, money, instrument length, or number of participants, present the opportunity for communication researchers to conduct their studies at lower resource cost and reduced threats to validity. At present, planned missing data designs are very rarely discussed or used in the communication discipline, despite a growing body of research in related disciplines which indicates both the statistical validity of these designs and the clear benefit they can have for social scientific research. By employing planned

missing data designs, we have an opportunity to improve the quality of the quantitative research produced in communication.

Ideally, this study will lead to PMDs becoming commonly used designs, because the advantages conferred by PMDs are applicable across the continuum of communication research. In the short-term, this dissertation will argue the merits of PMDs, show how these designs have been effective in other social sciences, and demonstrate empirically how previously-published communication research could have been conducted more efficiently using PMDs.

This is dissertation argues for and provides evidence in support of the use of a research design that is currently not being employed in communication. Despite the evidence from other disciplines that PMDs confer the expected efficiencies at minimal statistical cost, these designs have not been used with any depth in the communication literature. Though PMDs are not a cure-all, they offer benefits that are agnostic to the type of research being conducted. Though there are limits to the *type* of questions that can be addressed using PMDs, there is no limit to the *subject matter* those questions deal with.

The expected outcome of this dissertation is increased knowledge about the application of planned missing data designs in communication research. This dissertation will benefit the communication discipline by serving as evidence that planned missing data designs can have an observable positive impact on research in the discipline. Though PMDs can advance a wide range of quantitative study designs (e.g., experimental, quasi-experimental, survey research) in this dissertation they are being presented in a survey research context, where they allow communication researchers to use longer

questionnaires without increasing validity threats due to the instrument length, and also allow researchers to conduct accelerated longitudinal research which produces equivalent results in a dramatically reduced timeframe. This chapter will present a brief discussion of the types of missing data, strategies for handling missing data, the process of multiple imputation, and will then explore in detail the family of research designs that comprise planned missing data designs, and how they can be employed to the benefit of communication research.

Types of Missing Data

To understand planned missing data designs, one must first understand the patterns of missing data, the options available for handling missing data, and the impact that these missing data strategies will have on the data set. There are three types, or patterns, of missing data: Missing Completely at Random (MCAR), Missing at Random (MAR), and Not Missing at Random (NMAR) (sometimes referred to as Missing Not at Random). The distinction between the three patterns is based in what caused the values to be missing (Rubin, 1976). Data missing completely at random (MCAR) exhibit what Enders (2010) refers to as "haphazard missingness" in which "the observed data points are a simple random sample of the scores you would have analyzed had the data been complete" (p. 7). MCAR occurs when data are missing for a reason completely unrelated to any variable in the model. For example, in a pre-test/post-test design, if a respondent failed to return for the second collection because he/she had the flu, or had a professor that kept class 10 minutes late and therefore missed the bus to the data collection site, those would be considered MCAR because the reason the student's data are missing is not related to the study or the variables being measured.

When data are missing at random (MAR), the missingness is caused by a variable in the model, but not by the variable that has the missing data. In a MAR pattern of missingness, "a systemic relationship exists between one or more measured variables and the probability of missing data" (Enders, 2010, p. 6). MAR patterns arise when the pattern of missingness on a particular variable is related to another measured variable. For example, consider a university that is using Qualtrics to survey employee workplace satisfaction, and it finds a high rate of missing data. It turns out that the response rates for the survey differ dramatically based on the age of the employee, because younger employees are more willing to respond to a survey online. In this example, the missingness on the job satisfaction variable is caused by another variable (employee age). Assuming the survey asked about employee age, then the variable causing the missingness on the job satisfaction variable would also measured in the model, and the missingness on the job satisfaction variable would be a MAR pattern.

The final pattern of missingness is Not Missing at Random (NMAR), in which the pattern of missingness is caused by the variable with missing data themselves, even after controlling for other variables. Consider a study of communication competence, using as an instrument the Communication Competence Assessment Instrument (CCAI) (Rubin, 1982). One component of this instrument is a short extemporaneous, persuasive speech given by the participant (and evaluated by a panel of reviewers). In this case, whether a participant gives the speech (and whether his or her value for that item is missing or not) is directly related to what the participant's value on that item would have been had it been present. Students who are less competent communicators are more likely to have missing communication competence data because these individuals would be less willing to give

the persuasive speech. Thus, participants who would have scored low in communication competence are more likely to not complete the measure. This situation, where the missingness is caused by the variable which exhibits the missingness, is a NMAR pattern. In an NMAR pattern, the remaining data are heavily biased. In this particular example it is biased because only highly-competent communicators remain in the study to complete the measure, as those participants who would have scored low in communication competence have left the study and do not complete the measure.

Strategies for Handling Missing Data

Because missing data frequently occur in research, dealing with the complications that missingness can produce is an important consideration, and the strategy that should be employed to deal with the complications will change with the type of missingness. If the missingness is *truly* MCAR, then the missingness can essentially be ignored, because the remaining data are unbiased (the remaining complete cases are a random subsample of the original cases). This can be determined through the use of statistical tools, such as Little's (1988) test of MCAR for multivariate data. However, if the missingness is MAR or NMAR, the researcher must deal with the missingness, because "the complete cases are unrepresentative of the population, and biases are substantial" (Schafer & Graham, 2002, p. 157).

Missing data will almost always be present, to some greater or lesser degree, in any data set. If the missingness is not MCAR, which is likely, then scholars are presented with a choice about how to handle the missingness that occurs in their data. The options available can be broken down into two categories: 1) strategies that impart error and/or reduce the N; 2) newer strategies that reproduce the missing data using Maximum

Likelihood or Multiple Imputation procedures. Little (2013) likens this choice to a surgical decision. He states,

> These [older] approaches (listwise or pairwise deletion) are akin to surgery to remove the injured parts of our data. The modern approaches are akin to reconstructive surgery to restore the affected area to its original condition. Importantly, modern imputation is not plastic surgery to change or disguise the look of something—it is a restorative and reconstructive procedure. (p. 54)

Strategies that Impart Error or Reduce N

Listwise deletion

Listwise deletion, also known as case deletion, is the process by which any case that has at least one value missing is deleted in its entirety from the analysis, regardless of whether the variable(s) with the missing value for that case are part of the analysis. This can lead to a sizable reduction in N. Despite the dramatic bias that the procedure can introduce into a data set, and the incorrect conclusions that can result, listwise deletion is still commonly used in communication scholarship (e.g., Price, Nir, & Cappella, 2006; Cohen, 2008; Eveland & Hively, 2009; Roberto & Goodall, 2009; McComas, et al., 2010). In a situation where the missingness in a data set is purely random, such as in many planned missing data designs, listwise deletion will not have a biasing effect on the data. However, it will still effect a reduction in the overall study N, which will result in a loss of statistical power. Whether the initial study N is large enough to make this reduction in N due to listwise deletion non-problematic is a decision for the individual researcher.

Pairwise deletion

A second common strategy for dealing with missing data is pairwise deletion. In pairwise deletion, cases with missing values are removed from the analysis only if they have missing values for the variables being analyzed. For each analytical procedure, missing values are ignored, and the number of missing values are subtracted from the N for that analysis. This procedure results in different tests having different N values. Like listwise deletion, pairwise deletion can still produce biased (sometimes heavily so) results. The reduced N will also have a deleterious effect on statistical power. These effects can lead to incorrect conclusions being drawn from data, but pairwise deletion is still commonly employed in communication research (e.g., Pearson, Child, Carmon, & Miller, 2009; McComas, Trumbo, & Besley, 2007; Ballard & Seibold, 2000; Qi, Fink, & Cai, 2008).

The one exception to this assumption of introduced bias is when the only pattern of missingness present is MCAR. If the missingness is truly MCAR, then neither listwise nor pairwise deletion will impart any bias into the data, and the only negative consequence is the reduction in N. However, if the missingness is MAR or NMAR, listwise and pairwise deletion will reduce N, and impart some degree of bias, potentially serious levels (Enders, 2010).

Multiple Imputation

More effective methods of dealing with missing data, from a data validity perspective, are multiple imputation (MI) and maximum likelihood. Full Information Maximum Likelihood (FIML) is an estimator function that repeatedly attempts to fit different combinations of population parameter values until it finds the set that best fits

the data (Arbuckle, 1996). Because FIML is based on an expectation maximization algorithm, like multiple imputation, the analytical results garnered from using the two missing data procedures are frequently similar. Previous research has shown that multiple imputation and FIML "will always yield highly similar results when the input data is the same" and that "neither approach is inherently better than the other" (Collins, Schafer, & Kam, 2001, p. 349) This dissertation employs multiple imputation rather than FIML as a matter of preference, rather than statistics – the author was already somewhat familiar with multiple imputation before beginning this dissertation, and so employed it here, rather than FIML.

Imputation is the process of "replac[ing] the missing items with plausible values and proceed[ing] with the desired analysis" (Schafer & Graham, 2002, p. 158). This is a process "in which each missing value in the data set is replaced by a predicted value from a regression analysis based on complete cases plus a random residual term" (Sinharay, Stern, & Russell, 2001, p. 319). MI goes through this process multiple times, creating multiple complete data sets. In each data set, the missing values have been filled in with estimates of what the actual value would have been. The researcher then conducts the desired analyses on each data set (modern software packages make this a relatively seamless process), and then the multiple result sets are recombined into one averaged set. This process more accurately estimates both parameters and standard errors of the full data set. Sinharay, et al. (2001) explained that while MI cannot magically reproduce data that is missing, "MI retrieves some of the lost information" (p. 327).

The MI technique can provide impressively robust protection against lost cases due to a severe amount of missing data, because with an MI procedure, the goal is to

recreate, not dispose of, missing data. In simulations, Graham, Taylor, and Cumsille (2001) demonstrated that a data set with 44% missing values can be successfully reconstructed using MI, losing only 9% statistical power in the process. This technique applies to longitudinal designs, as well. Hecht, Graham, and Elek (2006) suggest that "even with relatively substantial attrition and other forms of drop out (e.g., students drop out at one wave but reappear at a later wave), the biasing effect on the validity of statistical conclusions may be trivial, provided the analysis incorporates accepted missing data procedures" (p. 270).

Usually, data collection involves going to great lengths to ensure that every case is complete, and that no datum is left missing. However, because MI is such a robust process for reproducing missing data, it affords researchers flexibility in data collection priorities, such that getting every participant to answer every item is not the only factor that can be considered. If one can reliably reproduce missing data, then considerations such as cost, time, and complexity become worth considering as well. In fact, some study designs assume that parts of the data will be left uncollected, and the researcher plans to have missing data.

Study Designs with Planned Missing Data

If multiple imputation is an appropriate method for repairing a data set after missing data has occurred, it is not a significant logical leap to consider that there may be circumstances in which the power of this procedure to re-create missing data can be leveraged to the researcher's advantage. In a planned missing data design study, the researcher intentionally induces missing data at the point of collection. Depending on the type of study, the missing data could take the form of items in a scale that are not

presented to a participant, participants who are not measured at a particular time point, or whole measurement waves that are skipped in a longitudinal design. Planned missing data designs fall into two broad categories: *accelerated longitudinal* and *efficiency* designs. This dissertation will present two types of efficiency designs: *split-form* designs, which are employed in a cross-sectional manner, and *controlled enrollment* designs, which are employed longitudinally.

Accelerated Longitudinal Designs

In a traditional longitudinal design, a single group of participants is measured at multiple time points in an effort to assess change in some variable over time. Scholars who attempt to implement a longitudinal design are faced with challenges particular to the design. Most significantly, whereas cross-sectional studies only require participants to complete a questionnaire (or other data-collection procedure) a single time, longitudinal studies require the participants to complete a questionnaire multiple times. At each subsequent point of data collection, there is risk of losing participants who had completed the previous data collection. This loss of participants, called attrition, can have an impact on the validity of results. For example, Ishii (2006) implemented a design with a two-year lag between data collections, and encountered a 33.7% attrition rate, leading to a conclusion that "the panel respondents may be less representative than the total respondents of the first wave" (p. 352). While weighting procedures can be undertaken to mitigate some of the error introduced by this attrition, these procedures can significantly complicate a study design.

Similarly, Rubin, Graham, and Mignerey (1990) conducted a study in which they assessed students' levels of communication competence to see if these levels changed

over the course of a college career. The researchers measured a group of students (N = 50) upon entry to the college (freshman year), and then measured them each fall for the ensuing three years. Rubin et al.'s study took more than three years to complete and, at the last time point, had an N of 20 students. The rest of the participants were lost to various types of attrition (leaving the study, leaving school entirely, etc.). As a result, this study was only partially successful in rejecting its null hypotheses. Some of the differences the researchers had hoped to locate were not seen, and they attributed this to the small N, which (potentially) prevented them from seeing significant results.

These problems with attrition come about despite the best efforts of scholars to convince participants to return. Valkenburg and Peter (2009) reminded their respondents "twice by e-mail [and] once by surface mail. [Then participants] were finally offered an extra bonus of 5 Euro, in addition to the 3 Euro that they received for filling in the [initial] questionnaires" (p. 84). Despite these multiple attempts, Valkenburg and Peter still had a 30% loss rate between data collections. The result is that, in a longitudinal design, "not only must one wait longer to obtain findings, but sample attrition will tend to accumulate over many waves of data collection, undermining the validity of inferences" (Miyazaki & Raudenbush, 2006, p. 45).

Scholars who use a longitudinal design are faced with the challenge of balancing competing needs in design choices. Lag (time between measurement points) needs to be long enough to allow change in the construct of interest to take place. If the lag in the study design is too short, no change will be detectable. With increased lag comes increased risk of losing participants. In a study on usage of sexual internet material by Dutch teens, Peter and Valkenburg (2009) decided that "because hardly any research on

optimal time lags exists in our specific field, we based our choice of a 6-month time interval between the waves on two pragmatic considerations," the second of which was that "choosing inappropriately long time lags might have increased the risk of losing too many respondents" (p. 177-78).

There are two key challenges to conducting longitudinal designs that accelerated planned missing data designs help solve: attrition and practice effects. Practice effects are what happens when "the number of times a respondent has taken the [instrument] has an adverse effect on the quality of answers they provide" (Backor, Golde, & Nie, 2007, p. 20). The inherent danger in administering the same instrument multiple times is that previous completions of the instrument may affect how a participant responds on subsequent measurements, which negatively affects the researcher's ability to access that participant's true score. It is in the interest of the researcher conducting longitudinal research to minimize the number of times any given participant takes the same instrument. Unfortunately, that goal is in direct opposition to the idea that ability to see complex levels of change is directly related to number of measurement points.

The solution to this problem is the *accelerated longitudinal* design, also called the *cohort-sequential* design (Bell, 1953; 1954). In this design, the researcher collects data in segments split by time and cohort, which are then reconnected in software. In effect, the researcher is "approximate[ing] a long-term longitudinal study by conducting several short-term longitudinal studies of different age cohorts simultaneously" (Duncan, Duncan, & Hops, 1996, p. 237). This design has become more popular in recent decades as computer hardware has become powerful (and cheap) enough to make the use of such software accessible to most researchers.

The eventual result is a reconstructed growth curve connecting a single trend from the youngest age at the first measurement, to the oldest age at the last measurement. For example, in a study investigating the efficacy of a new core curriculum amongst communication majors by measuring change in communication apprehension, the cohort might be class year, and data collection might happen at four time points: fall, winter, spring, and fall (in year two). With this design, the researcher has data for first-year students in the fall (entry to college), seniors in the spring (graduation), and all data points (i.e., time/school year combinations) in between. From this information, the researcher can reconstruct a single growth curve just as effectively as if each data point had been collected in chronological order over the course of four years. By compacting the time frame for data collection, several benefits can be realized. First, the shortened time will reduce attrition of participants. Second, the reduced number of measurement times can reduce the cost to conduct the study. Third, the compressed time span will reduce the likelihood that some confounding event (e.g., a natural disaster that closes the school for a week) will affect the data (Schaie, 1986).

Types of Accelerated Designs

There are numerous contexts in communication research in which time is an important variable. Any research questions involving change, whether it is related to relational development, self-disclosure, technology usage, or any number of other topics, could benefit from the use of a longitudinal design. As previously mentioned, the lack of longitudinal designs in communication research is evidence of a challenge facing the discipline and that there are a variety of structural (funding, timeframe, training) difficulties causing this challenge. I will now focus on explaining how the application of

planned-missing designs to a longitudinal context poses an opportunity to alleviate some of those challenges.

An accelerated longitudinal designs (Bell, 1953; 1954) is based on an analysis which involves collecting simultaneous line-segments and then reassembling them into a coherent line. To understand how this type of design works, it might be easiest to start by thinking about what a traditional longitudinal design looks like when the results are plotted. When a traditional longitudinal design is plotted, there is a line demonstrating change in a variable over the course of several measurement points. These measurement points on the plot could be thought of as representing boundaries between segments of the line, or the place where line segments meet each other. In an accelerated longitudinal design, these line segments are collected simultaneously (see figure 1), instead of sequentially, and are used as building blocks for the final plotted line. Essentially, this is a difference between thinking of longitudinal change as a single line, or as multiple line segments reassembled into a single line (see figure 2). Duncan, Duncan, and Strycker (2005) describe accelerated longitudinal designs as:

> provid[ing] a way to link adjacent segments of limited longitudinal data from different age cohorts to determine the existence of a common developmental trend, or growth curve. In this way, the researcher approximates a long-term longitudinal study by simultaneously conducting and connecting several short-term longitudinal studies of different age cohorts" (p. 74).

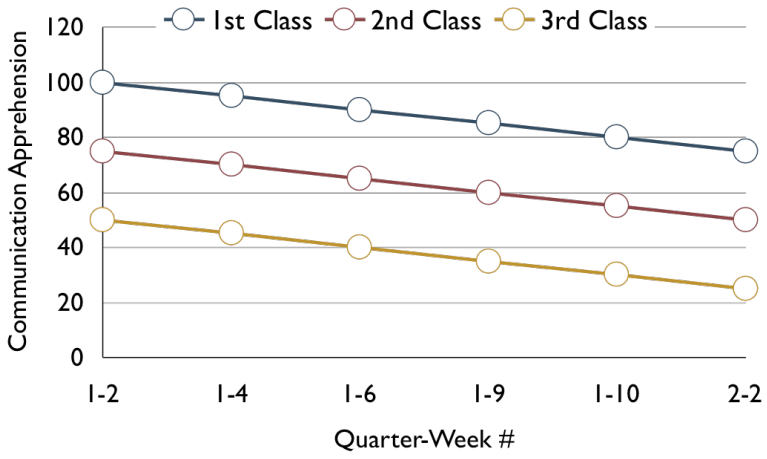

*Figure 1.*Accelerated Longitudinal Study, Measurement Design. Class number refers to how many communication courses that group of participants has enrolled in. Measurement points on x-axis indicate time during the academic quarter. Note that quarter 2 - week 2 is the second week of the *next* communication course, such that quarter 2-week 2 for the "1st class" group is the same measurement point as quarter 1-week 2 in the "2nd class" group.

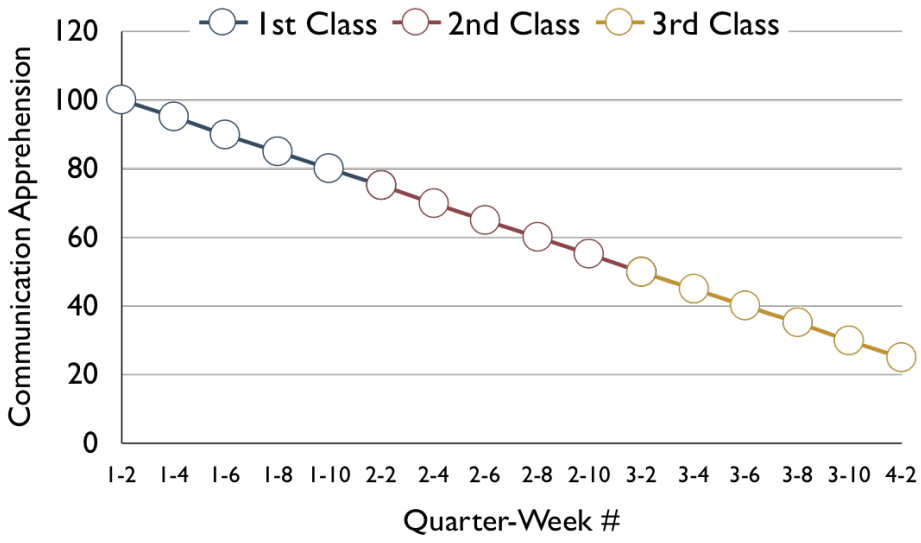

Figure 2. Reconstructed Accelerated Longitudinal Study Design. Class number refers to how many communication courses that cohort of participants has enrolled in. Note that quarter 2 - week 2 is the second week of the next communication course, such that quarter 2-week 2 for the "1st class" group is the same measurement point as quarter 1-week 2 in the "2nd class" group.

When planning an accelerated longitudinal design, the researcher is best served by thinking of the eventual (reconstructed) line in terms of line segments, because the statistical power of the study is dependent less on pure sample size at the final measurement point and dependent more on the number, structure, and overlap of the line segments that will be used to reconstruct the line (Moerbeek, 2011).

Longitudinal designs face all of the same challenges that cross-sectional designs face and more. Both types of designs must fight participant attrition (which in a cross-sectional design is thought of as *dropout*), and both must fight the tension between building a large, thorough, and powerful research design and the cost that such a design

will incur. What longitudinal designs add are the problems specific to the elongated timeframe of the research design. Whereas a cross-sectional design must only deal with time in a very basic sense (i.e., how long will the instrument take to complete), a longitudinal design must also consider everything that takes place outside the study itself over the course of the research. This might include participant loss over time, or the impact of confounding variables that affect the study design during the research timeframe. Ultimately, while the longitudinal design affords the advantages that come with being able to see change over time, "alternative approaches are needed to reduce study time, subject attrition, and the cost of continual assessment" (Duncan & Duncan, 2004, p. 350).

Advantages Conferred by Accelerated Longitudinal Designs

Accelerated longitudinal designs confer two main advantages over traditionally structured longitudinal designs. First, the shortened timeframe of the study design decreases the likelihood of participant attrition across measurement points. Also, the shortened timeframe itself is a significant advantage. Being able to conduct research in a dramatically reduced amount of time will increase the amount and quality of research that communication researchers can conduct. Finally, in the unfortunate event that a problem occurs during the research process, and an entire study is lost, the cost in time wasted for an accelerated longitudinal study is greatly reduced, sometimes saving years of potentially lost time.

Reduction in Attrition

The number of participants who fail to fully complete all of the instruments in a research design can vary greatly. In cross-sectional studies, the dropout rate (participants

who start – but do not finish – the instrument) could be as low as a few percent. In longitudinal research, the problem is greatly exacerbated by the need to call participants back to complete measures at a later time. It is not unheard of for longitudinal research to lose a large percentage of participants between the first and last measurement. For example, Frymier (1994) was able to use data from only 178 of 523 initial participants (34%) in a study that lasted one semester. Hecht, Graham, and Elek (2006) employed a design with a pretest and 3 posttest measures spread over the course of more than 16 months. The pretest was completed by 4,199 participants, but only 1,479 participated in all four measurement points, for a total of 65% attrition. Miller, et al. (2005) employed a design with a six-month time frame, and lost more than 37% of participants over the course of the study. Peter and Valkenburg (2009) lost 46% of participants in a one year design with three measurement points, despite an effort to "avoid inappropriately long lag times [that] might have increased the risk of losing too many respondents" (p. 178).

Though there are a variety of factors that affect attrition, including studying sensitive topics (Peter & Valkenburg, 2009), employing an accelerated longitudinal design can reduce the risk of attrition by dramatically reducing the time between first measurement and final measurement. For example, Prinzie, Onghena, and Hellinckx (2006) used an accelerated-longitudinal design to replicate change in children between the ages of 4-9. Their study design included three time points and they lost only 13% of participants to attrition over the two-year timeframe of the study.

Time saved

Nussbaum (2007) suggested that one reason for a lack of longitudinal studies in communication was the immense time commitment that long-term designs can require.

Accordingly, one of the benefits of an accelerated longitudinal design (regardless of type) is that by collecting data on multiple cohorts simultaneously, the data collection phase takes much less time than a traditional design. For example, in Tables 1 and 2, I present the design plan for a college-career-long study.

Table 1 presents the measurement points for this study design and the organization of the cohorts. In this study, measurement takes place over two school years, and the participants are three cohorts of students who are in their first, second, and third year of school (respectively) at measurement time wave W1. Each cohort is measured six times, three times per year, for two successive school years.

Table 1

Accelerated Longitudinal – Example Measurement Pattern

	Year 1 – Fall Wave 1	Year 1 – Mid-year Wave 2	Year 1 – Spring Wave 3	Year 2 – Fall Wave 4	Year 2 – Mid-year Wave 5	Year 2 - Spring Wave 6
First-year	X	X	X	X	X	X
Sophomore	X	X	X	X	X	X
Junior	X	X	X	X	X	X

Note: Participant age at first measurement on Y axis, Measurement time on X

Once data collection is completed at time wave W6, the conversion to a cohort-sequential accelerated longitudinal can be conducted, and the result of that process is shown in Table 2. This conversion is made possible because of the occurrence of overlapping measurement points. Participants who entered the study as first-year students, by the time they are measured at time point W4, are in the fall of their sophomore year – the same as sophomores measured at time point W1. Similarly, participants who entered the study as sophomores, when they are measured at time point W6, they share a common measurement point with participants who entered the study as juniors who were measured at time point W3. The end result is that, by reconstructing the discrete line segments into a coherent line, a continuous pattern of change across the measured timeframe can be analyzed, as detailed in Table 2.

Table 2

Accelerated Longitudinal – Example Cohort-Sequential Pattern

	First Year			Sophomore			Junior			Senior		
	F	M	S	F	M	S	F	M	S	F	M	S
First Year	1	2	3	4	5	6						
Sophomore				1	2	3	4	5	6			
Junior							1	2	3	4	5	6

Note: Cohort on Y Axis, Age on X axis, measurement wave number in cells. F = Fall, M = Mid-year, S = Spring.

By building this line, comparisons can be made between new first-year students upon entry, seniors about to graduate, and every point in between, having only taken two years to collect data, rather than the four years that a traditional longitudinal design would have taken.

Circumstances not Suited for Accelerated Designs

Despite their efficacy in mitigating some of the potential problems inherent to longitudinal research, there are circumstances in which accelerated designs are not effective. For example, the possibility exists that a confounding event may take place during the timespan of the study design. This event can impart serious error into a growth curve constructed from an accelerated design.

Accelerated longitudinal designs involve collecting data from multiple cohorts simultaneously, but then treating that data as if it had been collected chronologically. Because of this treatment of the data, any event that has a dramatic impact on all of the cohorts at the same time will appear to have had a repeated impact on the participants in the study, when in reality the event had a single impact – it just had that impact upon multiple cohorts. Figure 3 demonstrates what the reconstructed accelerated longitudinal design would look like if such a confounding external event took place at week six of a ten-week academic quarter. Note the change in longitudinal trend at week six, how it is repeated in each segment, and how these segments form a less coherent line than the line segments from Figure 2.

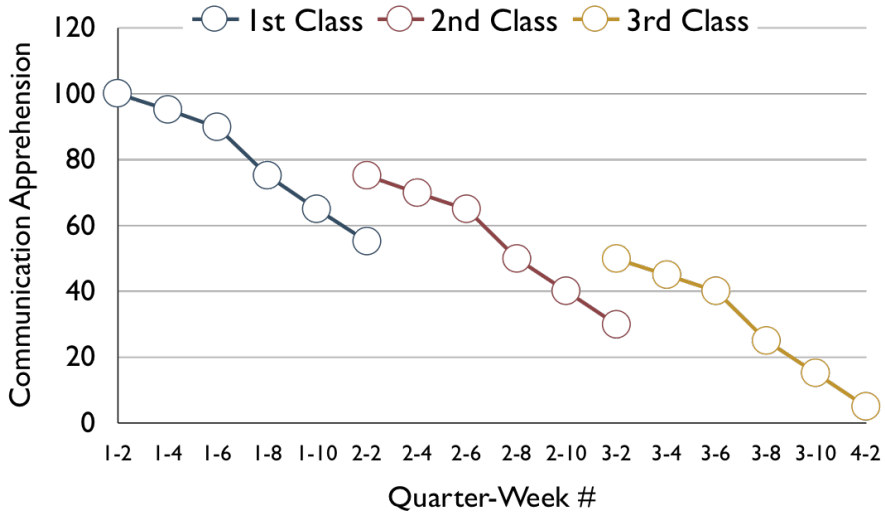

Figure 3. Reconstructed Accelerated Longitudinal with Confounding Event. Class number refers to how many communication courses that cohort of participants has enrolled in. Note that quarter 2 - week 2 is the second week of the next communication course, such that quarter 2-week 2 for the "1st class" group is the same measurement point as quarter 1-week 2 in the "2nd class" group.

Not every event that takes place in the situation surrounding an accelerated longitudinal design rises to the level of a confounding event. While some events are "far more parochial in nature and affect only certain localities or specific subsets of the general population" (Schaie, 1986, p. 258), there are also events which have an impact more globally. These are the events can threaten the integrity of an accelerated design. For instance, the "introduction of technological changes that achieve virtually instantaneous and universal acceptance" (Schaie, 1986, p. 258) might introduce significant error into the study, as would any event that had a significant impact on

participants in the study such as September 11th. In the hypothetical new communication curriculum study mentioned above, the introduction of a new teaching technology mid-study, such as recording equipment in every classroom, could threaten the results. This is because, in an accelerated design, that disruptive event would appear in the data not one time, but one time *per cohort*.

In a traditional longitudinal design, this type of dramatic change in communicative technology would appear as a knot point in the growth curve. Knot points are "points in time at which one phase is ending and another phase is beginning" (Grimm & Ram, 2011, p. 415). They form the intersection of distinct (nonparallel) line segments of the growth model. There are a variety of situations where knot points can be intended as part of the study design. For example, in a study of drug use, the knot point may be the implementation of the treatment design, where there is an intended before-after dichotomy in change trajectories. In essence, while any longitudinal model will show change over time, a knot point is an indication of a change in the *rate of change* over time.

Knot points can also be unintentional. For example, in a study of foreign language acquisition over a high school career, an unanticipated change in instructors (due to firing, moving away, sickness, pregnancy, etc.) might result in a clear knot point in the growth curve, where the rate of language acquisition changes. In a traditional longitudinal design, this event might not necessarily be a hindrance to achieving the desired results from the study, because it is possible that the effect due to that unanticipated change can be accounted for in the statistical analysis.

In an accelerated design, those knot points are multiplied by the number of cohorts, because that disruptive change will have happened to all four class years in the high school simultaneously. While the introduction of a disruptive communicative technology might not threaten a traditional longitudinal design, it is much more likely to induce error in the findings of an accelerated longitudinal design. One single event in time would appear multiple times on the resulting growth curve, because the change will have happened to multiple cohorts simultaneously.

Efficiency Designs

The second main type of planned missing data design is efficiency designs. Efficiency designs all have in common the assumption that there are advantages to be gained by strategically not collecting all possible data points for each participant. While there are an almost-infinite number of particular forms of efficiency design, there are two broad categories that are distinguished by the research questions they are designed to address, and by the form the missing data will take. One category is designed for cross-sectional studies and the missingness results from presenting only some items to all participants (in what is known as a split form, or x-form design) (Raghunathan & Grizzle, 1995). The other category is designed for longitudinal studies and the missingness results from asking participants to complete some, but not all, measurement points (known as a controlled enrollment design) (Little, Bovaird, & Card, 2007; Little, 2013).

Furthermore, the split-form designs can be subdivided into two main types, based on whether the researcher chooses to use forms with parts of measures split across multiple item blocks, or whether they use forms which keep measures intact, so that no

measure is split across multiple blocks. This is a choice between "between-block designs" and "within-block designs" (respectively) (Adigüzel & Wedel, 2008, p. 609).

Between-Block Split-Form Design

The basic form of an efficiency design is an x-ways split-form design, where *x* different versions of the questionnaire are designed, each containing a subset of the full instrument. In the between-block design (Adigüzel & Wedel, 2008), the researcher creates blocks of items by splitting measures across blocks, such that every measure contributes to every bock, and within each measure, approximately the same number of items are assigned to each block (Raghunathan & Grizzle, 1995). Each participant is then presented with one of the forms, containing a subset of the total number of blocks on the questionnaire. If the blocks have been designed well, the observed blocks will contain items that correlate very highly with items that are in the missing blocks, and vice versa. It is under the condition of high correlation between blocks that the MI engine performs the best. If the within-block correlations remain low, and the between-block correlations remain high, Raghunathan and Grizzle (1995) find that "there is little loss of efficiency" (p. 60), which would translate to little loss of power. Other research has suggested that with appropriate sample sizes (at least 200 participants), there exists only a very narrow range of effect sizes (.03 wide) in which a traditional design has greater than .80 power, but an efficiency design has less than .80 power. (Graham, Taylor, Olchowski, & Cumsille, 2006). In other words, there are only very limited conditions under which a traditional study would have enough power to detect an effect, but a PMD-based efficiency design would not. This indicates that with appropriate samples, properly

constructed split-form designs are only marginally less powerful than complete-form designs.

Finally, it is usually preferable to split measures across groups, which results in a between-block design, because the higher the inter-item correlations are for a particular measure, the larger the between block correlations will be, creating ideal conditions for MI (Graham, Hofer, & MacKinnon, 1996; Graham, et al., 2006). However, there are also particular advantages conferred by maintaining intact measures in a split-form design, which results in a within-block design (Adigüzel & Wedel, 2008).

Within-Block Split Form Design

In a situation where the primary obstacle to an appropriately large N is cost, within-block designs offer a useful tool to minimize the sample size required for any particular group of participants (Adigüzel & Wedel, 2008), by reducing the cost incurred to collect data on measures for which fewer participants are required. If the instrument to be used is comprised of multiple measures, and there is a difference in cost to administer them, a within-block design allows the sample size for participant groups to be adjusted as necessary based on which sections of the total instrument the group will receive. In situations where there are expensive, but powerful measures, those measures do not need to be given to every participant, thus reducing costs. This design allows for the researcher to customize the minimum number of participants on a per-measure basis, not on an omnibus basis. In other words, it "allows information on data items with relatively low variability and high cost to be collected from fewer units than data items with relatively high variability and low cost" (Chipperfield & Steel, 2009, p. 227-228).

Controlled-Enrollment Longitudinal Design

Finally, there are longitudinal studies in which an accelerated design is not ideal. For instance, when there is no logical cohort structure into which to organize the participants, or when the entire timeframe of the longitudinal design is short enough to make the accelerated design not worth the added complication. In these situations, the application of a controlled enrollment efficiency design could still confer advantages over a traditional design. In a controlled-enrollment design, the "blocks" of an efficiency design are measurement points in a longitudinal design. Participants are measured on fewer than all of the blocks. Though participants complete the full instrument when they are measured, they are not measured at every time point, so the missingness takes the form of complete measurement points.

For an example of how this design might be implemented, consider the work of Knobloch and Theiss (2011), who used six measurement points at one-week intervals to measure levels of relational uncertainty and relationship talk over time. If the researchers had been interested in implementing a planned-missing design, an accelerated longitudinal design would not have been appropriate for their study, because they had no easy way to arrange their participants into cohorts. Their hypotheses did not speak to how long the couples had been together (at least not as an analytical variable), how old they were, what grade they were in, etc. Without an easy way to divide participants into cohorts, there was no way to construct a cohort-sequential accelerated longitudinal design. Instead, they could have employed a controlled enrollment design.

In a controlled enrollment design, the efficiency is found in the reduced demand on participants and the reduced likelihood of measurement effects and participant

attrition because participants are asked to complete the instrument at fewer than all of the measurement points. A controlled enrollment design might have benefitted Knobloch and Theiss's (2011) study, because they made design choices which indicate they were aware of the risks of putting too much of a demand on their participants. They noted that their six-week design was chosen to be "sensitive to people's day-to-day experiences of relationship talk" (p. 10) while not requiring too large of a time commitment from participants. A controlled enrollment design could have allowed them to reduce demand even further, requiring each participant to complete as few as four measurement points, while still having a total of six measurement points in the study design. This reduces demand, but not the ability to draw conclusions from results.

Controlled enrollment designs can also allow the researcher to extend a longitudinal design without adding additional demand on participants. For example, if Knobloch and Theiss (2011) had decided that six measurement points was an acceptable level of demand on participants, but that they were interested in obtaining a more finely detailed understanding of the relationship between relational uncertainty and relationship talk over time, they could have expanded their study design using a controlled enrollment design with more than six measurement points, in which each participant completes no more than six measurements. Using a controlled enrollment design, they could have expanded their design to a nine-week study without any additional burden on participants. All participants would complete the first collection point, but then complete a randomly assigned five of the following eight measurement points. The result is a design in which each participant still completes six measurement points (no additional burden on participants) and do not complete three. The resulting 33% missingness is not

likely to impact their study in a negative way. Based on Graham, et al.'s (2006) guidelines for sample size, missingness, and resultant power loss, Knobloch and Theiss (2011) would probably have not lost any significant results that they found in the traditional design, but they could have seen additional results from weeks seven through nine that the original six-week design was not designed to see.

Advantages Conferred by Efficiency Designs

Efficiency designs present several different types of advantages relative to traditional study designs. First, they hold the promise of an increased response rate, which can dramatically increase analytical power. They can also reduce the cost of a study. Finally, they can increase the validity of a study by increasing the quality of the data collected.

Response rate

Because efficiency designed-instruments can be between 10-50 percent shorter than their full-length equivalent, it is possible to dramatically reduce the deleterious impact of questionnaire length. Studies have found that the length of a questionnaire affects the level of nonresponse to the questionnaire (e.g., Dillman, Sinclair, & Clark 1993; Deutskens, de Ruyter, Wetzels, & Oosterveld, 2004). Ganassali (2008) found that an increase in questionnaire length from 20 items to 42 items caused a 6% increase in survey dropout (i.e., the participants start – but do not finish – the questionnaire). This result is particularly troublesome, because 42 items is a fairly short instrument. In fact, it is almost impossible to conduct some types of questionnaire-based designs (such as pilot testing items for a proposed scale with more than 2 theorized subscales) with fewer than

42 items. An efficiency design can keep questionnaires short for participants, while not limiting the number of items included in the overall study.

Cost

Efficiency designs can potentially reduce the cost of measurement. A controlled-enrollment design significantly reduces cost as there are entire measurement points where participants are not measured. For cross-sectional research, insofar as an efficiency design is used for the same purpose as, but is shorter than the equivalent traditional measure, costs are reduced. Cross-sectional efficiency designs mimic the form and goals of short-form scales, which are created because they take less time to administer, put less demand on the participant, and are cheaper and easier to use (Wardenaar, et al., 2010). Because an efficiency designed-instrument will have fewer items than an original scale, it can be deployed at less cost to the researcher. This can be of great benefit to communication research, especially in circumstances when the research design involves scoring by a trained coder, or other fixed-cost expenses such as political communication or other public opinion research, where questionnaire costs are often based on a per-question basis (Montgomery & Cutler, 2013).

Validity

Communication research is based largely in asking participants to give responses to a variety of questions about particular behaviors, but we often cannot directly measure the constructs we are interested in. Nearly all of the items on the numerous scales that our studies include are actually designed to access some latent construct within the mind of the participant. If the questions are designed well by the researcher, and the participant answers all questions truthfully, then the answers (in aggregate) give us information

about how the participant feels about some construct that we cannot measure directly, like happiness, relational satisfaction, or willingness to communicate. The ability of the researcher to access the latent construct can be dramatically reduced if participants do not accurately answer the questions they are given. Previous research has shown that item positioning late in a long questionnaire can cause an increase in straight-line responding (i.e., always selecting the same response), especially in item sets with identical response options, leading to dramatic changes (greater than .10) in correlations between items in a scale (Herzog & Bachman, 1981). Though computer-based survey systems can ameliorate this problem through random ordering of items, much of our research is still performed offline. Other research has noted that in longer questionnaires, participants are more likely to provide a response inconsistent with a previous response (which would affect validity) (Deutskens, et al., 2004). Questionnaires of extended length can impart fatigue effects (see Porter, 2004; Krosnick, 1999), and fatigue effects can have a deleterious impact on ability to access the desired latent constructs for each participant. Because of this deleterious effect, it is likely that even with the slight loss to statistical power due to imputation, the data from a very long questionnaire that is shortened through a split form design (which is then multiply imputed) could be *more valid* than it would be had each participant been asked to complete each item.

A logical assumption might be that as a participant progresses later into an instrument, they increase the randomness with which they respond to items, exhibiting a lack of caring. Instead, research has shown that "responses seemed to be less discriminating, rather than being more random" (Kraut, Wolfson, & Rotherberg, 1975, p. 775). Participants more commonly choose the central answer, indicating not randomness,

but a lack of cognitive effort. By employing an efficiency-design, researchers can ameliorate these various problems, to some greater or lesser degree. The use of split questionnaire designs "decrease completion time, fatigue, boredom, and nonresponse and are evaluated more positively by respondents" (Adigüzel & Wedel, 2008, p. 616). Thus, efficiency designs are a way to achieve shorter, cheaper, and more valid results than complete-form designs.

Simulation studies have shown that the use of efficiency designs can adequately replicate the results from a full data set. One study artificially induced missingness into a complete-case data set containing psychological and educational-related measures from a sample of Dutch law students (Smits & Vorst, 2007). Tests of this data with one-third and one-half missingness inserted show that the imputed sets' estimates exhibit equivalent levels of reliability and validity, and do not contain any systemic bias relative to the original data they are derived from. The only negative impact from the imputation process was a larger confidence interval around the resulting estimates.

In practice, efficiency designs have allowed researchers to conduct research that would otherwise have been unfeasible. Batra, Lenk, and Wedel (2010) surveyed participants on their perceptions of various brand personalities. Ideally, they would have been able to have each participant complete a 15-item assessment for each of the 30 brands they were investigating. However, because that would result in a 450-item instrument, that was clearly not a realistic option. To make their instrument more manageable, the authors employed a split-questionnaire design, and imputed the missing sections. In their design, each participant only rated 10 brands, with one additional brand common to all participants. This resulted in a 160-item instrument, which, while still very

long, equaled a 63.3% reduction in items that each participant was responsible for, and had no negative impact on their results. Because the researchers employed an efficiency design, they were able to conduct research covering an impressive array of brand perceptions, but were able to do so while dramatically reducing the burden on their participants.

Another example shows how studies that had not previously used a planned missing design can benefit greatly from their use. In fact, sometimes the act of replacing a destructive missing data solution (like listwise deletion) with a planned missing data design can result in seeing effects that might always have been present – but are only now observable. Aasland, Olff, Falkum, Schweder, and Ursin (1997) surveyed approximately 9,000 Norwegian physicians, and the instrument they needed to use was far too long to use in its entirety. They employed a split-form design, with each participant receiving a common part A, and 3 of the remaining 15 parts. The authors note that they could have used the normal method of analysis, which is to use "only the respondents who had valid answers on all variables" which totals to "220 subjects, or 1/22 of the total sample" However, had they used this approach, their results would have been that "none of the predictors reach[ed] significance" (p. 1626). Instead, through a single, regression-based imputation process, their effective sample size became 6,100, and their logistic model performed well. Incidentally, the authors acknowledge that multiple imputation may have been a superior method for handling missing data than the single-imputation process that they used, but that "relatively heavy machinery" (p. 1620) was required to perform the multiple imputation, so they did not conduct the procedure. As modern computer

hardware is more than sufficient to conduct multiple imputation procedures on even the largest data sets, this is no longer a limitation.

Hypotheses

Like all quantitative dissertations, this one presents hypotheses as a way to benchmark which results are expected and in concert with previous research, and which results are unexpected and differ from previous research. However, this dissertation differs from many others in that the expectations for what will result from the quantitative analyses have to be considered from two different perspectives. On a broad, conceptual level, success must be defined by the degree to which planned missing data designs can help communication researchers deal with the challenges facing them. In each individual study contained within this dissertation, however, there must also be local hypotheses. As such, based on the previously reviewed literature, this dissertation will test both overarching goals about the impact that planned missing data designs will have on the process and results of quantitative research in communication, but also individual study-level hypotheses about the particular expectations of each example study used to illustrate the practice of planned missing data designs.

Goals Related to PMDs

For this dissertation, I would like to define success by directly comparing planned missing data designs to fully observed designs, to highlight the advantages of the planned missing data designs. The use of a planned missing data design can potentially affect data in two distinct ways relative to a fully observed dataset: bias and increased uncertainty surrounding parameter estimates. Unfortunately, the only way I can directly test a planned missing data design against a fully observed design is through the use of

simulation studies that generate artificial datasets, and then induce missingness into the data. In real-world conditions, it is simply not possible to use multiple study designs, simultaneously, on the exact same sample of participants.

Simulation studies done previously indicate that multiple imputation produces only minimally biased results relative to fully-observed data sets by setting the desired parameters *a priori*, then imposing a pattern of missingness on thousands of generated datasets, and multiply imputing the created gaps (Graham & Schafer, 1999; Schafer & Graham, 2002). This lets them compare the imputed parameters to the generated ones they are replicating. However, in the real world, there is no way to prove a lack of biased data through imputation. It is simply impossible to both collect and not collect data simultaneously on the same sample of participants.

Ultimately, previously published studies in other disciplines that use PMDs and simulation studies that support the math behind the method may provide the strongest argument that PMDs are mathematically sound. If we accept that evidence, then the question of whether they are useful in a communication context becomes one of advantages conferred in terms of time, money, reduced burden, and validity, and whether the studies that the planned missing data designs are employed in can stand on their own merits (in terms of power, effect sizes detected, hypothesis testing, etc.). Taking all of the above into consideration, I developed a series of goals specifically regarding planned missing data designs. Each of the three studies in this dissertation are designed to achieve one of the following goals, which directly address the challenges facing communication research as presented in the previous chapter:

- The PMD-related goal of study one is to reduce participant burden in terms of measurement points.

- The PMD-related goal of study two is to reduce participant burden in terms of instrument length.

- The PMD-related goal of study three is to reduce the time between the beginning and end of data collection in a longitudinal design.

CHAPTER 3: STUDY 1 - PUBLIC SPEAKING ANXIETY

The first study in this dissertation was an original study applying a planned missing data design to the assessment of self-reported public speaking anxiety and communication competence over the course of a semester-long public speaking class. The study employed a controlled enrollment design (Little, 2011), in which all participants received all measures, but were not measured at all time points. In this design, each student completed the questionnaire at three of the five measurement points. The resulting missing data, in the form of completely missing time points, was then multiply imputed. The resulting data had all of the analytical advantages of a longitudinal study with five measurement points, with a lower likelihood of burden and testing effects due to repeated exposure to the instrument.

The primary goal of this study was to use a controlled-enrollment planned-missing data design to test the relationship between public speaking anxiety and communication competence over a semester-long public speaking course. Within this primary goal, there are two objectives. The first objective is to advance an understanding of how both public speaking anxiety and communication competence function longitudinally. The second objective is to demonstrate the utility of PMDs as powerful methodological tools with which to conduct future studies in this, and other areas of communication research. In this chapter, I will present the longitudinal structural equation model used to test these hypotheses, followed by results relevant to each specific hypothesis.

Hypotheses

In addition to the planned missing data design hypotheses presented in chapter two, I tested hypotheses specific to this study. These hypotheses were consistent with the literature which indicated that public speaking anxiety would decrease over the course of time (e.g., Duff, Levine, Beatty, Woolbright, & Park, 2007) and communication competence would increase (e.g., Rubin, Rubin & Jordan, 1997):

H_1: Self-reported public speaking anxiety will decrease over the course of the academic term.

H_2: Self-reported communication competence will increase over the course of the academic term.

H_3: Self-reported public speaking anxiety will be negatively related to self-reported communication competence.

Study Design

Sample Size

Because the data analysis employed SEM-family analyses (i.e., confirmatory factor analysis, structural equation modeling, longitudinal growth curve modeling), I calculated the sample size using a power analysis tool designed for SEM-family analyses (Preacher & Coffman, 2006). To conduct a power analysis using this tool, several parameters must be known about the data, including degrees of freedom, null hypothesis RMSEA, and alternative hypothesis RMSEA. Because decisions made in the later stages of model fitting (e.g., parceling of indicators) affected the degrees of freedom in the final model, a definitive power analysis could not be made prior to data collection. Rather, a

range of power analyses, covering potential scenarios, resulted in liberal and conservative estimates of sample sizes needed for sufficient statistical power under potential analytical circumstances.

The first step in determining the necessary sample size was determining degrees of freedom. Across all the scales included in this study, participants responded to a total of 53 items. Putting that in terms of indicators, depending on the exact model specified, a 53-indicator model would have had upwards of 1,300 degrees of freedom (Rigdon, 1994). Because 53 indicators would result in a globally-overidentified model, which could impair model fit (Kline, 2010), it was likely that indicators would be parceled during the analysis phase to better approximate the ideal just-identified solution (Little, Cunningham, Shahar, & Widaman, 2002). Parceling is the process of combining individual scale items into groups of items, and using these resulting groups, or parcels, as indicators to load on to latent constructs in a structural equation model. This can have a stabilizing effect on parameter estimates, which also leads to better model fit (Matsunaga, 2008). The potential downside to parceling is that it transforms scale parameters into "arbitrary metrics" that cannot readily be compared to the parameters of the original, "untransformed metrics" (Little, et al., 2002).

The parceling resulted in fewer degrees of freedom than the initial estimate of 1,300 presented above. A parceling scheme that reduces the number of indicators in the model to 1/3 of the original number of indicators, or 18 indicators, would result in a model with 110 degrees of freedom. Because power decreases as the degrees of freedom decrease, using 18 indicators and 110 degrees of freedom to calculate sample size is a conservative estimate of how many participants are required.

Model fitting is an iterative process, and as the shape of the model changes, the degrees of freedom do as well. As a result, there was no way to know ahead of time exactly how many participants would be required. However, I was able to make a conservative estimate about the potential degrees of freedom that the final model would have, and assuming a worst-case scenario in terms of degrees of freedom and statistical power in the final model led to a safer estimate of necessary sample size at the time of recruitment. As a result, I assumed 110 degrees of freedom for this model.

After estimating degrees of freedom, the next step was to estimate null and alternative RMSEAs. MacCallum, Browne, and Sugawara (1996) suggest that the appropriate construction of hypotheses for testing model fit is to test a null hypothesis of not-good-fit against an alternate of good fit. Accordingly, the power analysis for this study employed a null hypothesis value of RMSEA = .08, and an alternative hypothesis value of RMSEA = .01, based on a conservative definition for the boundary of acceptable fit and poor fit (Kline, 2010; Browne & Cudeck, 1993). The more conservative RMSEA = .08 boundary was chosen instead of the more liberal RMSEA = .10 boundary of acceptable fit and poor fit (Bollen & Curran, 2006) because I have tried to consistently be conservative in making estimates during the power analysis so as to avoid underestimating the eventual necessary sample size.

I conducted the power analysis[1], based on assumptions of alpha = .05, power = .80, degrees of freedom = 110, null RMSEA = .08, alternative RMSEA = .01, produces a suggested N of 64. This provides a conservative lower bound for the sample size needed

[1] Using Preacher & Coffman's (2006) R code derived from MaCallum, et al.'s (1996) SAS code.

for good model fit, however, that lower bound is based on a complete-case sample. To calculate a power analysis for a planned missing design, a further correction is necessary (Enders, 2010). Because each participant in this design will have 40% missing and 60% observed data (3 out of 5 collections), any correlation between two variables will contain 36% observed data (.60 * .60). To correct the sample size for the missing data, I divided the complete-case N (64) by the fraction of observed data (.36), which resulted in a corrected sample size of 178 (Enders, 2010). Calculating a precise power estimate with missing data requires simulations based on the fraction of missing information (which is distinct from the percentage missing, and measures the missing data's influence on the sampling variance of parameter estimates) (Liu, 1994), but the rough percentage missing correction provides "a viable option for generating conservative power estimates" (Enders, 2010, p. 32).

This corrected sample size of 178 would have been the minimum desired sample at each time point, accounting for the use of SEM in the analysis and the presence of missing data during the collection process. However, that sample size of 178 assumed a perfect response rate from the recruited participants at every collection point. To again be conservative, I estimated the response rate at each time point to be 80%, resulting in a minimum goal of 220 participants to be recruited at each measurement point.

Planned Missingness

In a controlled-enrollment PMD, whole time points are missing for each participant (Little, Bovaird, & Card, 2007). In this study, each participant completed the questionnaire at only three of the five measurement points and the remaining two measurement points were multiply imputed. Because of the planned missingness, this

study was able to see longitudinal change with the level of detail of a five-measurement-point design, but with the lower potential for measurement effects of a three-measurement-point design. Also, the lower levels of demand on each participant reduced the possibility of attrition during the study.

Recruitment

Students taking specific communication courses, including the public speaking course, earn 2% of their grade by participating in research or an alternative assignment, creating a pool of potential research participants. From this pool of students, the research pool administrator randomly assigned 500 participants enrolled in the public speaking course to this study. Because research participation is required for course credit, research completion rates from recruited students have historically been very high, about 90%. The recruitment email I sent to potential participants clearly stated that this study was associated with the official research participation system, and gave them a brief explanation of the research study, instructions on what the extent of their participation would be, and a link to the Qualtrics web-hosted questionnaire where the data collection took place. Because this study is longitudinal, recruitment emails for secondary and tertiary data collections also reflected that participants had already given generously of their time, and that I was appreciative.

Participants

Participants for this study were 458 students enrolled in the basic public speaking course during the Fall 2012 semester. This number exceeds the minimum number of required participants to have sufficient power in the study. Participants ranged in age from 18-37, with 83.8% being 18-19 (M = 18.75, SD = 1.34, Mn = 18.00, Md = 18.00),

because the basic public speaking course is most commonly taken during the first year on campus. The vast majority (n = 416, or 91%) of participants reported being in their first or second year of school. The sample was 55% female, which is a slightly higher representation of females than the university student body, which is about 52% female (Ohio University Office of Institutional Research, 2012).

The public speaking course is a standardized course in which all participants complete the same five graded speeches. All classes completed the speaking assignments at approximately the same time, though there was some unavoidable variation due to the difference in class meeting length and days [i.e., Tuesday-Thursday classes (80 minutes) and Monday-Wednesday-Friday classes (55 minutes)]. In general, every student in the public speaking course presented each speech within a few days of one another. Finally, participants were asked to list how many classes they had previously taken which involved public speaking (M = 1.78, SD = .84), and to rate their level of public speaking experience on a scale from 1-5 (M = 2.79, SD = .1.14).

Procedures

The next step in this study was to assign each participant to a group to coordinate at which measurements each participant would be asked to provide data, and at which measurements they would not. The research pool administrator assigned an initial pool of 500 students to this study, and I then randomly assigned each of them to groups using a random number generator. I created 10 groups, each with 50 participants. The group assignment process was completely transparent to the participants, and only affected them insofar as it determined at which three of the five possible data collection points they completed the questionnaire. At each measurement point, six out of the 10 groups,

or 300 of the 500 participants, were asked to complete the instrument, and the remaining four groups (200 people) were not contacted to complete the instrument. Table 3 presents the complete schedule of which groups were contacted before which speeches.

Table 3

Data Collection Schedule for Controlled-Enrollment Design

Group	Type of Speech				
	Special Occasion	This I Believe	Informative	Persuasive	Impromptu
1	X	X	X		
2	X	X		X	
3	X	X			X
4	X		X	X	
5	X		X		X
6	X			X	X
7		X	X	X	
8		X	X		X
9		X		X	X
10			X	X	X

Note: X *denotes the presence of a data collection for a particular group for a particular speech. The absence of an* X *denotes no data collection for that group for that speech.*

Three days before the first day of each speech, I sent emails to all participants assigned to groups from which I was collecting data that round. The email included a link

to the Qualtrics questionnaire where data collection took place. The only exception to this procedure was the first collection point. A delay in acquiring the participant list prevented the emails from being sent to participants until approximately 36 hours before the first speeches of the semester. The three-day lead time was chosen to give participants enough time to complete the questionnaires, but not so much time that they were not feeling anticipatory anxiety for their upcoming speech.

I customized the recruitment email based on participant group, because different participant groups proceeded through the study design at different times. At each measurement point after the first one, the study included participants at differing stages of the data collection process. My reminder emails reflected the particular stage of the study that participants were in, having completed either zero, one, or two questionnaires already. Examples of each version of these reminder emails are included as Appendices A (first collection), B (second collection), and C (third collection) respectively.

Scale Reliability

To assess the reliability of the indicators used to model public speaking anxiety and communication competence across the five measurement waves, this study employed the maximal reliability coefficient, or Coefficient H (Hancock & Mueller, 2001). Coefficient H is an estimate of the correlation the latent factor is expected to have with itself over repeated administrations. Values above .70 are considered desirable (Mueller & Hancock, 2010). This measure of indicator reliability is preferable to the traditional measure, Cronbach's Alpha, because α produces a biased underestimate of the reliability of the composite scale under the circumstances found in this study – when the items in the scale are congeneric (load on the same construct, but without correlated errors

between the items or equal loadings), but not tau-equivalent (congeneric with equal loadings) (Miller, 1995; Hancock & Mueller, 2001).

Measures

This study included two measures: McCroskey's (1970) Personal Report of Public Speaking Anxiety (PRPSA) and Ellis' (1995) Self-perceived Public Speaking Competency Scale (SPPSC). These scales respectively measure anxiety and competence in a public speaking context and are self-report measures. I included short series of demographic questions as well.

PRPSA

The PRPSA is a frequently cited measure of public speaking anxiety (e.g., Hensley & Batty, 1974; Burgoon, 1976; Daly, 1978; Lohr, Rea, Porter, & Hamberger, 1980; Beatty, 1988; Scott & Rockwell, 1997). It is a 34-item, Likert-type scale, with five response options from strongly agree to strongly disagree. It has been validated as a one-dimensional measure of public speaking anxiety (Hensley & Batty, 1974). Example items include "I breathe faster just before starting a speech" and "I face the prospect of giving a speech with confidence." Table 4 details the Coefficient H (Hancock & Mueller, 2001) values for the PRPSA at each of the five measurement waves in this study.

Table 4

PRPSA Coefficient H Scores

	Coefficient H
Wave 1	.96
Wave 2	.96
Wave 3	.95
Wave 4	.96
Wave 5	.97

SPPSC

The Self-Perceived Public Speaking Competency Scale (Ellis, 1995) is a self-report measure of public speaking competence. It is a 19-item, Likert-type scale, with five response options from never to very often. Example items include "I use language that is extremely clear" and "I dress to enhance my credibility." The scale was created as a self-report version of the Competent Speaker Speech Evaluation Form (CSSEF), published by the Speech Communication Association, (now the National Communication Association) (Morreale, 1990). The original CSSEF underwent a rigorous validation process supervised by a panel of communication scholars, who concluded that the CSSEF "is a viable instrument for assessing public speaking competence" (Morreale, Moore, Surges-Tatum, & Webster, 2007, p. 31). The items on the SPPSC were designed to parallel the language in the CSSEF to retain validity (Ellis, 1995). Of particular note for its application in this analysis, the SPCCS was designed to be used in a longitudinal

setting (Ellis, 1995). Table 5 details the Coefficient H (Hancock & Mueller, 2001) values for the SPPSC at each of the five measurement waves in this study.

Table 5

SPPSC Coefficient H Scores

	Coefficient H
Wave 1	.86
Wave 2	.87
Wave 3	.87
Wave 4	.89
Wave 5	.90

Demographics

Participants responded to demographic questions about age, sex, and grade level. Participants also indicated how far in the future (in days) their next speech for their public speaking class was. This information allowed for measures (especially measures of anxiety) to be corrected for time-until-speech. Table 6 details this information for each speech.

Table 6

Means and Standard Deviations of Days between Questionnaire Completion and Speech Delivery

	Special Occasion	This I Believe	Informative	Persuasive	Impromptu
Mean	1.91	2.11	3.49	6.41	1.39
SD	1.29	1.62	2.79	5.47	1.21

Data Analysis

Data Cleaning

The first phase of the analytical process was a brief data cleaning procedure. Because data collection was conducted entirely online through Qualtrics, and there was no free entry, there were no out-of-range values to screen for. Thus, in this case data cleaning was limited to screening for blank records. Only one blank record was found at one data collection. In most communication research, the data cleaning process would continue with consideration for assumptions for the normality of the data. Unlike in most other communication research, assessing the impact of skewness and kurtosis in PMD studies is both difficult and, at this point in time, uninformative in the decision-making process, because imputing under transformed conditions leads to *more* biased results than the untransformed imputation of nonnormal data (von Hippel, 2013).

Skewness and kurtosis statistics can be obtained for observed data from any software package, but in the presence of unplanned missing data (as in the case of all three studies in this dissertation), any statistics based on observed-data only could be biased. Currently, skewness and kurtosis statistics for imputed data are not available in any output in Mplus (Muthén & Muthén, 2012), the data analysis software used in this dissertation, so there is no effective way to obtain an unbiased statistic for skewness or kurtosis in the presence of any unplanned missing data. Thankfully, this lack of information does not affect the data analysis process.

Enders (2010) suggests that violations of normality assumptions will have a limited impact on the parameter estimates that result from the MI process, and transformations to reshape the data towards normality may do more harm than good by altering the covariance structure of the data. Other research suggests that pre-imputation transformations may not be appropriate because of the altering of the covariance structure (Demirtas, Freels, & Yucel, 2008). This leads to a conclusion that it is usually best to impute under an assumption of normality, regardless of whether the data is actually normally distributed or not. As unsatisfying as not transforming potential violations of normality might be, it is current best practices, though Enders (2010) emphasizes that the decision to transform or not requires further methodological research.

Multiple Imputation

Data analysis continued with the process of multiple imputation, using Mplus v7 (Muthén & Muthén, 2011). In Mplus, multiple imputation is a two-step process, in which the Markov chain Monte Carlo (MCMC) algorithm creates a distribution of potential values from which imputed datasets are then generated. First, the MCMC algorithm must

converge to a satisfactory level before any imputations can be conducted. An MCMC algorithm achieves convergence when it has generated a distribution that is the same as the posterior distribution (in Bayesian terms, the distribution informed by the observed data) (Sinharay, 2003), and the boundary between satisfactory and unsatisfactory can be defined on the basis of potential scale reduction (PSR) (Gelman & Rubin, 1992). This is a measure "reflecting how much sharper the distributional estimate might become if the simulations were continued indefinitely" (Sinharay, 2003, p. 11). The more the Markov chains rely on the posterior distribution to generate their sample distribution, instead of the prior distribution, the lower the PSR value is. The PSR is measured on a per-parameter basis (i.e., each parameter in the model has a separate PSR value, and the highest individual parameter PSR is the PSR value for the entire procedure). The minimum value is 1.0. Gelman, et al. (2004) suggest values under 1.1 for all parameters. The default PSR to continue to the imputation phase in Mplus is 1.05. To be conservative, I set the maximum PSR for this study at 1.01. This maximum PSR has the effect of requiring dramatically more MCMC chain iterations before convergence – and sometimes convergence never is reached. However, setting the PSR at 1.01 also results in more precise estimates in the imputation phase.

There are various recommendations of how many imputations should be used in the multiple imputation procedure. Schafer' (1999) suggests that "Unless rates of missing information are unusually high, there tends to be little or no practical benefit to using more than five to ten imputations" (p. 7). On the other hand, more recent research suggests that there are benefits, including reductions in standard error inflation and statistical power falloff, to increasing the number of imputations as high as 40 when the

rate of missingness is greater than 0.50, as it was this study ($\gamma = 0.40$ planned, plus $\gamma = 0.14$ unplanned) (Graham, Olchowski & Gilreath, 2007). In the last five years, computational power has increased to the point that with data sets of the size used in this study, the difference in processing time between ten imputations and 40 imputations becomes less of a limiting factor than it was previously. On very large datasets, Schafer's (1999) advice might establish the lower bound for number of imputations, but for the purposes of this analysis, the number of imputations was set at 40. The limits of computer hardware and software to run the imputation procedure are, of course, contextually bound. An analysis that taxes the abilities of a home computer may be well within the abilities of a university mainframe, and computers are constantly being upgraded.

Because I did not have access to a mainframe, and did all of the analysis in this dissertation on my home computer, I had to take efforts to make the imputation procedure less computationally taxing by reducing the number of parameters in the imputation model (Asparouhov & Muthén, 2010), I employed a process called duplicate-scale imputation to reduce the number of parameters that needed to be multiply imputed (Enders, 2010; C. Enders, personal communication, March 1, 2013).

Duplicate-scale imputation is the process of imputing at the scale (in this case, parcel) level, not the item level, but using an auxiliary variable to avoid the loss of information. First, I created parcels from the raw items, such that each construct had exactly three parceled indicators per time point. Three parcels was the ideal choice following the recommendation of Little (2013), who suggested "to the extent possible, that just-identified constructs be utilized" (p. 85). The PRPSA was parceled into three indicators, using 11, 11, and 12 items in each of the parcels, respectively. The SPPSC

was parceled into three indicators, using seven, six, and six items in each of the parcels, respectively.

Any participant who did not complete all items in the parcel received a missing value for the entire parcel, meaning that the participant's value for that parcel would be imputed. However, because many participants did not complete *all* items in a parcel, but did complete many items in the parcel, simply assigning them a missing value would serve to throw away the answers they gave on the items they had completed. One of the central values underlying this dissertation is the proper handling of missing data, and throwing away data in this fashion is akin to listwise deletion – dealing with missing data through the further destruction of existing data.

Fortunately, duplicate-scale imputation provides a way to preserve the data that participants did provide. This process is based in the creation of an auxiliary variable, which is the mean of all of the items in a scale that the participant completed. The final version of the data had two scores (per parcel) for each participant. If the participant completed every item that contributed to that parcel, then they will have a value for the parcel, and they will have a value for the auxiliary variable that is equal to the value for the parcel (because it is the mean of all of the items contributing to the parcel that they completed – in this case, all of them). If the participant completed fewer than every item that contributed to that parcel, then they will have a missing value for the parcel, and they will have a value for the auxiliary variable that is equal to the mean of all of the items which contribute to that parcel, and which they did complete. This information is then used to impute the missing scale score.

The duplicate-scale imputation process was necessary because the overall design of this study had too many parameters to successfully impute with only 458 participants. Unparceled, the imputation model for this dataset would have contained 63,545 parameters, and would have taken weeks to impute, assuming the MCMC sequence ever converged. By using the duplicate-scale imputation process and a parceled dataset, the imputation model used in this study contained 2,277 parameters, and entire imputation process (MCMC sequence plus imputed data generation) took four hours and 41 minutes. Finally, this process also improves upon another potential alternative procedure, creating parcels from the mean of existing information. That process is conceptually similar to ad-hoc short form scale creation, in that it assumes that the items which contribute to that meaned parcel measure the construct centroid in the same manner that a parcel constructed from all of the items in the scale would.

Following the multiple imputation process, I continued with the main analytical process, which took place in two phases, and was conducted using Mplus v7 (Muthén & Muthén, 2011). First, I conducted a confirmatory factor analysis to establish measurement invariance over time, which is a critical step in assuring that the scales functioned in the same way at all time points (Brown, 2006). Next, I built a structural model to assess the degree of change over the course of the semester for each of the dependent variables to test H_1-H_3, respectively**.**

Study-Specific Results

Preliminary Analysis

Measurement Invariance

Establishing measurement invariance is a critical preliminary step in the process of conducting a longitudinal structural equation model. One of the assumptions behind a longitudinal structural equation model is that the instruments employed in the model function in the same fashion at each time point – that they are invariant across time. It is a procedure that was originally used in a cross-sectional, multi-group context, but has been adapted for use in the longitudinal setting (Little, Preacher, Selig, & Card, 2007). Failing to meet this assumption can be a threat to the validity of the structural equation model, because if a measure is not invariant over time, there is no way to determine whether measured change in a construct is due to changing levels of the construct or differences in how the measure itself is functioning over time (Little, 2013). Establishing measurement invariance is comprised of a series of progressively more restrictive tests, in which each test limits the ways that the measure can vary over time relative to the previous test. Because each test is comprised of the same model and data as the previous test with only the addition of parameter restrictions, the models are nested, and can be directly compared against each other using a ΔX^2 test (Geiser, 2012).

The first step in establishing measurement invariance is to establish *configural invariance* (Horn & McArdle, 1992), which is a test to establish equal factor structures. Because structural equation modeling involves constraining some parameters in the model to equal zero, constraining others to equal one, and letting others be freely estimated, the configural invariance step establishes whether the same pattern of fixed

and free parameters exists over time. Following that, weak factorial invariance, or *loading invariance* should be established (Widaman & Reise, 1997), in which the patterns of indicator loadings are constrained to be equal over time. For example, if at the first measurement, indicators 1 and 2 each loaded approximately equally on the latent factor, but indicator 3 loaded slightly less, weak factorial invariance would constrain that same pattern of indicator loadings to exist at each of the following time points. The third step in establishing measurement invariance is to establish strong factorial invariance, or *intercept invariance* (Meredith, 1993). In this level of invariance, the intercepts of each latent variable's indicators are constrained to have the same relationship to each other at each time point. There is a fourth level of invariance, strict factorial invariance or *residual invariance* (Brown, 2006; Meredith, 1993), in which the residual variances of corresponding indicators are equated across time. This level of invariance is generally not recommended to be used, though, as it is unreasonably strict, and unlikely to be met (Brown, 2006; Little, 2013).

In this study, results of the invariance testing procedure indicate that the configural model fits the data well, X^2 (300, 458) = 255.20, p = .97, RMSEA = .00 (.00 - .00), CFI = 1.00, TLI = 1.05. Constraining the loadings to be equal across time to establish weak factorial invariance results in a model that also fit the data well, X^2 (316, 458) = 264.62, p = .98, RMSEA = .00 (.00 - .00), CFI = 1.00, TLI = 1.06. ΔX^2 (16, 458) = 9.42, p = .90, indicating that the weak invariance model is not significantly worse fitting than the configural invariance model. Constraining the intercepts to be equal across time to establish strong factorial invariance results in a model that also fit the data well, X^2(332, 458) = 279.94, p = .98, RMSEA = .00 (.00 - .00), CFI = 1.00, TLI = 1.06. Δ

X^2 (16, 458) = 15.32, p = .50, indicating that the strong invariance model is not significantly worse fitting than the weak invariance model. As a result, I can conclude that both of the instruments used in this study function in the same way across measurement points, and that the "fundamental meaning of the construct has not changed across the different developmental periods" (Little, et al., 2007).

Initial Hypothesized Model

Previous research has suggested that the relationship between communication competence and public speaking anxiety may be more complex than a simple inverse relationship, including evidence that previous speaking experience and initial levels of public speaking anxiety and communication competence can affect the relationship between the two constructs (Rubin, Rubin, and Jordan, 1997). These findings are supported by my own extensive experience as an instructor, having taught public speaking to more than 1,000 students. This experience tells me that public speaking anxiety and communication competence are intricately intertwined constructs. Given this previous research and guided by personal experience, I decided to start with a model that hypothesized a high degree of complexity of relationships between communication competence and public speaking anxiety and to build the final model by removing non-significant paths. Thus, the initial model included three types of latent paths: Auto-regressive paths between latent constructs (e.g., competence at wave one on competence at wave two), nondirectional (covariance) paths between communication competence and public speaking anxiety (e.g., competence at wave one and anxiety at wave one), and cross-lagged paths regressing communication competence on public speaking at the next

successive time point (e.g., competence at wave two on anxiety at wave three), as seen in Figure 4.

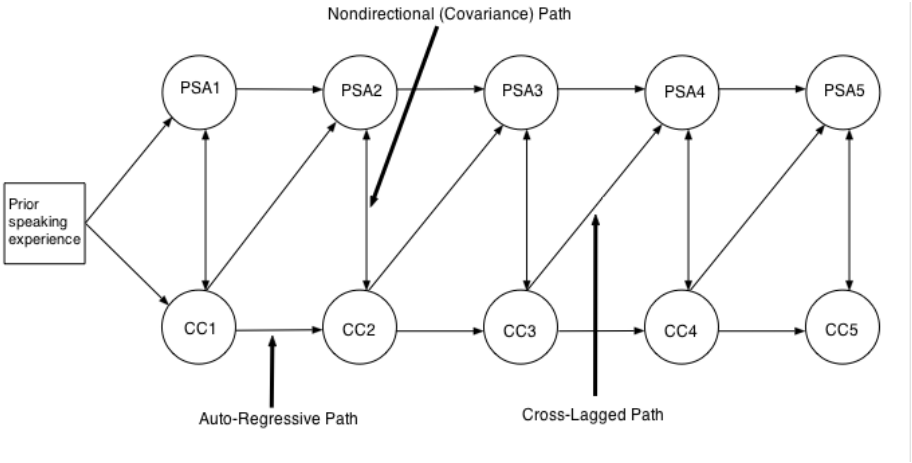

Figure 4. Initial Hypothesized Structural Equation Model, with auto-regressive, cross-lagged, and nondirectional (covariance) paths indicated. PSA constructs are public speaking anxiety (at time points 1-5), and CC constructs are communication competence (also at time points 1-5).

Additionally, previous public speaking experience (measured by a self-report 1-5 scale variable) was entered into the model as an exogenous covariate predicting only the first wave of endogenous latent constructs. This structure posits previous speaking experience as a covariate without further down-stream influence on the latent constructs, which is an appropriate assumption when participant differences in the exogenous covariate "are already established and won't change with later measurements" (Little, 2013, p.16). This saturated model demonstrated excellent model fit, X^2 (388, 458) = 305.46, p = .999, RMSEA = .000 (.000-.000), CFI = 1.000, TLI = 1.081, however there

were a number of non-significant regression paths. I proceeded with the standard iterative process of model re-specification by constraining to zero the latent path with the least association between the constructs, each time re-testing the model (Kline, 2005). This process ended when no non-significant latent paths remained in the model. The final trimmed model (shown in Figure 5) also had excellent fit to the data, X^2 (391, 458) = 309.96, p = .999, RMSEA = .000 (.000-.000), CFI = 1.000, TLI = 1.079. A chi-square difference test revealed no significant difference in model fit, ΔX^2 (3, 458) = 4.5, p = .21.

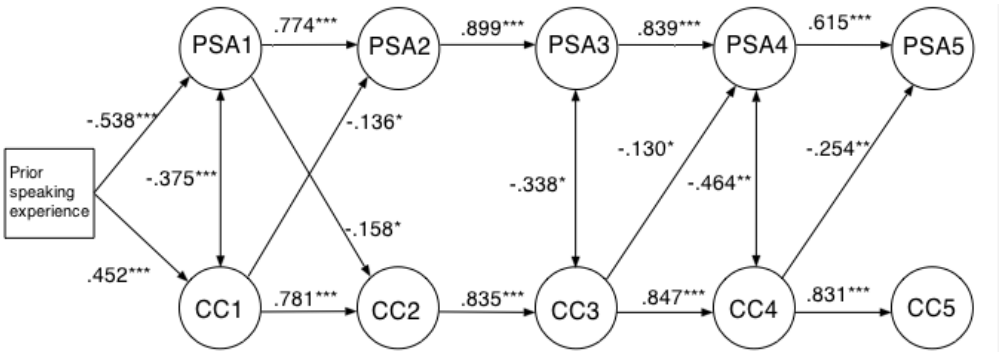

Figure 5. Final Trimmed Structural Equation Model, with nonsignificant paths sequentially removed.

Substantive Analysis

Hypothesis 1

The first hypothesis predicted a reduction in public speaking anxiety over the course of the semester. This hypothesis is supported by an examination of the latent intercepts, which indicate a significant overall decrease in public speaking anxiety over the course of the semester ($\Delta\alpha$ = -.43, z = -3.67, p < .001). Further examination indicates

that a significant decrease exists between waves one and two ($\Delta\alpha$ = -.25, z = -2.51, p = .01), but no significant difference exists between waves two and three ($\Delta\alpha$ = .001, z = .01, p = .99), waves three and four ($\Delta\alpha$ = -.004, z = -.04, p = .97), or waves four and five ($\Delta\alpha$ = -.19, z = -1.06, p = .29), indicating that the only significant change in mean level of public speaking anxiety occurred between waves one and two.

Hypothesis 2

The second hypothesis predicted an increase in self-reported communication competence over the course of the semester. This hypothesis is supported by an examination of the latent intercepts, which indicate a significant increase in communication competence over the course of the semester, ($\Delta\alpha$ = .26, z = 2.71, p = .007). Further inspection of the latent intercepts tells a more complicated story, however, because the difference in communication competence between waves one and two was non-significant ($\Delta\alpha$ = .06, z = .55, p = .58), as was the difference between waves two and three ($\Delta\alpha$ = .07, z = .08, p = .42). There was a significant *decrease* in self-reported communication competence between waves three and four ($\Delta\alpha$ = -.19, z = -2.07, p = .04), followed by a significant *increase* between waves four and five ($\Delta\alpha$ = . 30, z = 3.23, p = .001).

Hypothesis 3

The third hypothesis predicted a significant negative relationship between self-reported public speaking anxiety and self-reported communication competence. This hypothesis is supported by an examination of the cross-lagged latent paths, which indicate that communication competence at wave one significantly predicted public speaking anxiety at wave two (β = -.14, z = -2.0, p = .046), communication competence

at wave three significantly predicted public speaking anxiety at wave four (β = -.13, z = -1.97, p = .049), and communication competence at wave four significantly predicted public speaking anxiety at wave five (β = -.25, z = -2.87, p = .004). Additionally, public speaking anxiety at wave one was a significant predictor of communication competence at wave two (β = -.16, z = -2.09, p < .037).

Other Predictors

The final model contained previous speaking experience as an exogenous variable predicting both public speaking anxiety at wave one (γ = -.54, z = -12.78, p < .001), and communication competence at wave one (γ = .45, z = 8.80, p < .001). Previous public speaking experience as measured by number of classes taken that involved public speaking was found to be unrelated to any other variable in the model. I also tested time until speech as a predictor of both public speaking anxiety and communication competence. Surprisingly, it was not significantly related to either construct at any of the five measurement waves. This is directly in conflict with previous research (Behnke & Sawyer, 1998) which suggests that anxiety should increase as the speaking event gets nearer in time.

PMD-Related Results

PMD Goal 2

In addition to testing study-specific hypotheses, this study also addressed goals related to planned missing designs. The results of this study indicate that a properly constructed controlled-enrollment planned missing data design (Little, et al., 2007) allows data to be collected at more time points than a standard longitudinal design without increasing participant burden or attrition. In this study, there was a rate of

planned missing data of $\gamma = .40$ with two of five measurement points not collected for each participant. Additionally, there was a $\gamma = .16$ rate of unplanned missing data because participants failed to complete some questionnaire items. The design used in this study was able to minimize the attrition frequently present in longitudinal research. A total of 458 of the 500 initial participants contacted for this study eventually participated, for an overall loss rate of 8%. This compares favorably with previous longitudinal research in the classroom. Previous research has demonstrated, for example, loss rates of 66% of participants over the course of one semester (Frymier, 1994), 23% over the course of one semester (Duff, et al., 2007), and 39% over an 8-week period (King & Witt, 2009). Because this study was able to use data from every participant, whether they provided data at all three requested waves, or fewer than three, participant loss rates do not translate exactly. However, by reducing the burden that is put on participants, the use of a PMD increases the likelihood of participant retention relative to a fully-observed design.

Discussion

This study had two overarching objectives. First, one objective of this study was to test the efficacy of a controlled-enrollment type planned missing data design in a communication context. Second, this study sought to explain the nature of the longitudinal relationship between self-reported communication competence and self-reported public speaking anxiety over a semester-long public speaking course. Viewed as a whole, the results of this study strongly support the conclusion that the relationship between communication competence and public speaking anxiety is highly dynamic, and should be measured at a level more detailed than pre-test/post-test, or before/during/after

the semester. The results of this study also show that a controlled-enrollment planned missing data design is an effective way of measuring such a relationship.

PMD-Related

The first main findings relate to this study's implementation of a controlled-enrollment PMD. Most importantly, a controlled-enrollment design can be effective at reducing the burden of repeated measurements on participants in a longitudinal design without any negative impact on the study's ability to achieve its goals. This study was able to successfully test all of its hypotheses, in the process building a five-wave longitudinal structural equation model with near-perfect model fit, using a methodology that required participants to contribute data at only three of the five measurement points. In addition, the results from this study suggest that controlled-enrollment designs can reduce the likelihood of participant attrition. This study finished a semester-long design, including five measurement points, with a loss rate of 8%, which is lower than comparable previous research (Frymier, 1994; Duff, et al., 2007).. The other main findings of this study relate to the hypotheses specific to this study.

Hypothesis One - Public Speaking Anxiety.

The first hypothesis anticipated that there would be a reduction in public speaking anxiety over the course of the semester. This hypothesis was supported by the data, which indicated an overall decrease in public speaking anxiety from wave one to wave five. However, a closer inspection of the data indicates that what appears on the surface to be a reduction in public speaking anxiety over the course of the semester is actually a reduction from wave one to wave two and no significant change after wave two. Previous research in this area (Ellis, 1995; Rubin, Rubin, & Jordan, 1997; Duff, et al., 2007) has

consistently found a decrease in public speaking anxiety over the semester, but has yet to investigate *when* in the course of a multi-speech public speaking course the reductions in anxiety occur.

Knowing when during the semester changes in anxiety occur will have a real and immediate impact on public speaking anxiety research and pedagogy, by improving course design and increasing the efficiency of future research in this area. The public speaking curriculum in which this study was conducted does not focus anxiety-reduction efforts solely on the first presentation. It is an ongoing part of the curriculum throughout the semester. However, results from this study indicate that the first presentation is the only time that anxiety-reduction actually occurs. This suggests that the anxiety-reduction efforts at later points in the semester are either ineffective or effective only in a maintenance capacity. Future research should explore the efficacy of these late-semester anxiety reduction treatments. The increased ability to locate change outcomes in public speaking anxiety, as demonstrated in this study, will make it easier for future researchers to conduct the many-wave longitudinal studies necessary to precisely locate change in public speaking anxiety during the semester, leading to a better understanding of how particular anxiety-reduction curricula are (or are not) functioning.

Hypothesis Two - Communication Competence

The second hypothesis predicted that there would be an increase in communication competence over the course of the semester. This hypothesis was supported by the data, which indicated an overall increase in communication competence over the course of the semester. However, similarly to anxiety, competence appears to have a more complicated pattern of change across time than a simple pre-test/post-test

study would have revealed. In this case, the data suggests that the trend line of communication competence is curvilinear, such that there is a significant decrease in communication competence from waves three to four, and then a significant increase from waves four to five, though this curvilinear trend should be confirmed by future research using growth curve modeling (Duncan, Duncan, & Strycker, 2006). Complex growth curve models can describe quadratic change and piecewise change which might better capture individual patterns of change throughout a semester.

The results from this study have implications for public speaking instructors and course directors. Contrary to what is assumed by many public speaking instructors, the data suggest that repetition and exposure to the speaking event do not monotonically increase students' communication competence. The findings from this research could help course directors and public speaking instructors.

Hypothesis Three – Anxiety and Competence

The third hypothesis predicted that self-reported public speaking anxiety would be negatively related to communication competence. This hypothesis was supported by the data, which indicated that communication competence was a statistically significant negative predictor of public speaking anxiety at measurement waves one, three, and four, such that students who were higher in communication competence were lower in public speaking anxiety at the following measurement point. This finding reinforces existing literature showing that communication competence is a strong negative predictor of public speaking anxiety (Sallinen-Kuparinen, McCroskey, & Richmond, 1991; Ellis, 1995; Rubin, Rubin, & Jordan, 1997; Donovan & MacIntyre, 2004).

However, whereas previous research had investigated this relationship in the framework of a pre-test/post-test method, this study was able to see that relationship at a more granular level due to the use of a planned missing data design, and do so with an extremely low participant loss ratio. This indicates that a controlled-enrollment PMD is an improvement on previous methods of inquiry into the relationship between public speaking anxiety and communication competence.

Why was communication competence not a significant predictor of public speaking anxiety only from wave two to three? One potential explanation may be within the nature of the wave three speech – the informative speech – and how that speech differs from the other speeches during the semester. The informative speech in this public speaking course is not designed to be the most difficult speaking assignment – that is the persuasive speech, which takes place at wave four. From wave two to wave three, at the informative speech, however, is the point at which the most significant *increase* in difficulty takes place. In this particular public speaking course, speaking assignments generally get more challenging throughout the semester (i.e., speech two is more difficult than speech one, speech three is more difficult than speech two, etc.), and the speech two to speech three step has the greatest rise in difficulty (speech five – the impromptu speech - is the exception to this). During this time period – in initial preparation for the informative presentation, students are first faced with the challenge of doing outside research for their presentation, of constructing logical arguments, and of constructing a speech that is more than double the length of their previous speeches. Those same levels of preparation, each of which would prompt anxiety about the speech, are not necessary to the same degree in speeches one, two, and five. Although the preparation is necessary

for the persuasive speech at wave four, it comes after students have already completed these tasks once for the informative speech at wave three, so they are not as new or imposing.

It is reasonable to argue that the nature of the speaking assignment, in addition to its relative chronological placement during the semester, might affect the levels of public speaking anxiety felt by students giving speeches (Witt & Behnke, 2006), leading to the relationship being muted at this particular place in the semester. Unfortunately, if variables related to the speaking assignment itself are affecting this relationship, as opposed to its chronological placement during the semester, then this study does not include the data needed to explain the lack of a relationship. There were no items on the questionnaire that asked about any features of the particular speech.

Additional Findings

In addition to the findings mentioned above which directly relate to a stated hypothesis, there was one particular result of note. Although not predicted specifically in any hypothesis, results indicated that public speaking anxiety at wave one was a significant predictor of communication competence at wave two. This suggests that the students who were the most nervous in anticipation of their first speech reported themselves to be the least competent communicators before their second speech. However, this relationship only existed at that specific point in the semester.

This result indicates that there could be an unmeasured variable moderating the relationship between public speaking anxiety and communication competence: speech grade. In the public speaking class in which this study was conducted, the first graded speech is very simple. There are few requirements, the necessary components are highly

prescribed, and it is very short. In my experience as an instructor, students tend to succeed or fail at that first speech on the basis of little more than their prior speaking experience and how well they can manage public speaking anxiety. Students who fail to manage their anxiety get a poor grade, and view themselves as incompetent communicators in the future (and vice versa).

As anxiety decreases significantly between speeches one and two, it becomes less of a factor in speech grades, and thus less of a predictor of competence. At the same time, the speeches get more complicated, and speech grades (and competence judgments) become about more than just anxiety – they are informed by research, argumentation, and delivery skills as well. Future research should investigate this connection further by measuring speech grade to confirm whether it functions as a mediator between public speaking anxiety and communication competence.

Limitations

Despite the potential this study has for extending knowledge of public speaking anxiety and communication competence, there are some limitations to address. Because the primary focus of this study was on the methodology, the nature of the relationship between public speaking anxiety and communication competence was not completely explored through the inclusion of additional covariates. This resulted in an inability to fully explain a significant predictor of public speaking anxiety at wave one on communication competence at wave two in the final model. Future research in this area should ensure that information regarding the nature of each speaking assignment and the participants' perceptions of that particular assignment is collected, including information about speech grades.

There is also a minor generalizability concern to address. Though this study is only generalizing to the college/university public speaking classroom, the study was conducted at a particular campus, using a randomized assignment of 500 students who happened to be taking the public speaking course that semester (i.e., a randomized subdraw from a convenience sample). These students were engaged in a particular public speaking curriculum, and giving certain speeches, which may not be required at other universities. Thus, these results may not necessarily translate to other campuses, and they should be tested in other public speaking classrooms before the results can be assumed to be true in all situations.

One surprising result was the lack of any covarying effect from the days-until-next-speech variable. It is possible that this failure to covary could have been due to a measurement problem. I only used a single indicator to measure the amount of time between when participants completed the questionnaire and when they were going to present their next speech. The wording of that item might not have been optimal, however, and may have resulted in less-than-ideal data. Instead of asking "how many days from today do you present your xxx speech?" (where xxx was replaced with the name of the speech they were about to present), it would have been more theoretically consistent with the Tailored Design Method (Dillman, et al., 1993) to ask "when is your next speech?" Then, in combination with timestamp data from Qualtrics, I could have calculated time-until-speech myself. This would have shifted some cognitive load from the participants, potentially resulting in better data.

The days-until-next-speech variable might also have been unrelated for an entirely different reason. Although I posted each questionnaire, and sent recruitment emails

approximately three days before the beginning of each round of speeches, the actual time between when the recruitment email was sent and when each student delivered their speech varied greatly. Most of the speeches were spread across multiple class sessions, and I had no access to information about when individual participants were speaking, so I had no way to more specifically time the recruitment emails.

Complicating this measure even further was the fact that the persuasive speech (speech number four) was spread out over a much greater time period because it was scheduled across the Thanksgiving holiday. This resulted in some participants taking the questionnaire *much* further in advance of their speech for that particular measurement than they did at any of the other speeches. For some students, this was as much as a three-week difference, which limits the ability of the survey to measure anticipatory speech anxiety.

Finally, given the very high proportion of missing data in this study, it would be ideal to be able to directly compare estimates from the final, imputed dataset against the original, observed-only dataset. This would allow me to distinguish the effect that the imputation procedure had on the final estimates from the observed data. Unfortunately, that is not a possibility in this study.

The missing data as a result of study design in an efficiency-type PMD will be MCAR, assuming proper randomized assignment of participants to groups (Enders, 2010). In this circumstance, the use of listwise deletion, though still not advised (due to the potential losses of statistical power), could be unbiasing because the location of the missing data is randomized. However, results in this study indicate that the missing data in this study was not *purely* planned. Based on the design, this study should have had a

proportion of missing data $\gamma = .40$. The total proportion of missing data however, was $\gamma = .57$. This additional, unplanned missing data was not distributed evenly throughout the data collections. It appeared at a greater rate in collections three and four, suggesting that timing or speech related variables might have affected participants' response rates to the study. This missing data was not randomly distributed, and therefore is not MCAR, and cannot be assumed to be unbiasing towards the remaining data. In this circumstance, using listwise deletion could be highly biasing (in addition to the reduction in statistical power).

Unfortunately, it is difficult to directly compare the observed and imputed datasets in this study because of the high rate of planned missing data in the study design. This design results in a situation where every case is missing values for 40% of the variables in the dataset, and using listwise deletion in this analysis would result in zero valid cases. As a result, a direct comparison between observed-only and imputed data cannot be conducted under a listwise deletion scenario.

Conclusion

The ultimate purpose of this study was use a controlled-enrollment longitudinal planned missing data design to examine the relationship between public speaking anxiety and communication competence. After establishing strong measurement invariance, an iteratively trimmed structural equation model showed that there was a reduction in public speaking anxiety over time, an increase in communication competence over time, and that communication competence was a statistically significant negative predictor of public speaking anxiety at each measurement wave. These results are consistent with existing research on public speaking anxiety, communication competence, and the

relationship between the two constructs. Finally, the cross-lagged regressions of communication competence on public speaking anxiety, and the single cross-lagged regression of public speaking anxiety on communication competence suggest that there are still questions in this area of study that lack complete answers and points to the next logical area of future research in this area.

The results also provide key evidence for the efficacy of planned missing data designs as a method and for the continued need for research into the complex relationship between public speaking anxiety and communication competence. Using the methodology employed in this study, multi-wave longitudinal studies become less burdensome on the participants and therefore more likely to succeed, giving the researcher more and better information about the relationships between these latent variable structures at specific points in time.

CHAPTER 4: STUDY 2 – VERBAL AGGRESSIVENESS SCALE

Introduction

The second study in this dissertation was a replication study applying a planned missing data design to the scale validation process conducted by Infante and Wigley (1986). This study employed an efficiency design, in which all participants completed a questionnaire containing less than every item on all of the scales contained within the instrument. The resulting missing data was then multiply imputed. The goal of this study was to illustrate how planned missing data designs could effectively shorten scales without affecting construct validity, and reduce the burden on study participants, thereby reducing the likelihood of testing effects. Infante and Wigley's (1986) study was chosen for replication in this dissertation because in the original study, the authors engaged in a particularly robust process of validating their Verbal Aggressiveness Scale against existing scales of related communication constructs, a process that resulted in a study design that placed a high burden on participants. This burden can be dramatically reduced through the use of an efficiency-type planned missing data design.

In the original study the authors compared the Verbal Aggressiveness Scale to five established scales totaling 143 items, plus a Cognitive Complexity (Crockett, 1965) measure, which is free writing-based. To reduce the anticipated participant burden, they asked their participants to complete this instrument in three pieces, at different times, separated by two weeks. In this dissertation, through the use of a split form planned missing data design, their scale validation process was replicated, but participants completed the data collection in one (online) session.

In an efficiency design, each participant is presented with fewer than all of the items in an instrument (Graham, et al., 2006). In this case, each participant was presented with 95 of the 143 total items from the self-report scales. This resulted in an instrument which, while still long, was short enough not to induce measurement effects due to instrument length (Andrews, 1984). Having each participant complete the instrument in one session obviates the need to have participants return multiple times, which prevents participant attrition between measurement points, and it avoided complicating the design by having to "mak[e] the study appear to be three separate and independent studies" (Infante & Wigley, 1986, p. 65) as the original study did.

Hypothesis

In addition to the hypotheses about planned missing data designs that were presented in chapter two, this study also tested hypotheses which were originally presented in the Verbal Aggressiveness scale validation study (Infante & Wigley, 1986). Previous research suggested that verbal aggressiveness should have had positive associations with the assault and verbal hostility subscales of the hostility-guilt inventory (Buss & Durkee, 1957) and with communication apprehension (McCroskey, 1982). Previous research also suggested that verbal aggressiveness should have had negative associations with social desirability (Crowne & Marlowe, 1960) and inadequacy (Hovland & Janis, 1959).

$H_{4:}$: Verbal aggressiveness will be positively associated with assault, verbal hostility, and communication apprehension, and it will be negatively associated with social desirability and inadequacy.

Study Design

Sample Size

The first step in the study design was to establish the necessary sample size to properly conduct this study. For that, I needed to conduct a power analysis. In this case, the power analysis was a straightforward exercise, because this was going to be a very simple analysis - Pearson correlations only – and pre-established values for the strength of those correlations already existed in the original Infante and Wigley (1986) study

To conduct the power analysis, I used G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009), using alpha = .05, power = .80, H_0 = .00. For H_1, the alternative hypothesis, there were a variety of choices, because this study was simultaneously testing multiple Pearson correlations. I decided that this study needed at least enough statistical power to detect the weakest relationship that had been detected in the original study, which was the correlation between Verbal Aggressiveness and Cognitive Complexity, though that specific measure was not used in this study. As a result, H_1 was set at .22. The power analysis in G*power resulted in a suggested sample size of 159.

This power analysis provided a conservative lower bound for the sample size needed for good model fit, however, that lower bound was based on a complete-case sample. To calculate a power analysis for a planned missing design, a further correction was necessary (Enders, 2010). Because this design had at least 35% missing data (95 out of 143 variables observed for each participant), correlations between verbal aggressiveness (which was fully observed) and the other construct variables contained only 65% observed data. Dividing the complete-case N from the power analysis by the fraction of observed data resulted in a corrected N = 245 (Enders, 2010). Fortunately,

based on the recruitment pool available to me for this study, I was able to exceed the minimum required N by approximately 750.

Recruitment

Students in specific communication courses earn 2% of their grade by participating in research or an alternative assignment, creating a pool of potential research participants. From this pool of students enrolled in participating COMS classes, the research pool administrator randomly assigned approximately 1,150 participants to this study. Due to logistical problems in the administration of the research pool, the exact number of participants recruited by the COMS research system cannot be verified, but 1,113 records were created in Qualtrics indicating a survey session had been started. Of those 1,113, a total of 93 records indicated no attempt to enter data, and the record was discarded. Because research participation is required for course credit, research completion rates from recruited students have historically been very high, about 90%. The recruitment email sent to students from the research system clearly stated that this email was associated with the official research pool, gave them a brief explanation of this specific study, instructions on what the extent of their participation would be, and a link to the Qualtrics web-hosted questionnaire where the data collection took place.

Participants

Participants for this study were 1,020 students enrolled in communication courses during the Fall 2012 semester. This number greatly exceeds the minimum number of required participants to have sufficient statistical power in this study. Because these COMS classes participate in a school-wide research pool, the participants in this study received course credit for their participation in the study. Participants ranged in age from

18-57, with 97% being between 18-22 (M = 19.52, SD = 2.47, Mn = 19.00, Md = 19.00). A total of 75.6% of participants reported being in their first or second year of university, even though participation in this study was not age restricted (beyond being limited to adults over 18). This is likely a result of 100-level courses being disproportionately represented in the research pool, even though there are many upper-level courses that participate. The sample was about 60% female, which is noticeably higher than the university student body, which is about 52% female (Ohio University Office of Institutional research, 2012).

Procedures

Participants entered the online survey system through a link made available to them electronically. Clicking on that link took them to a landing page before the questionnaire where each participant was presented with an informed consent document, and given a yes/no option to proceed. The yes option took them into the questionnaire, the no option took them to the end of the questionnaire, where they were presented with a message of thanks for their time. The survey system then presented the participants with a randomized series of scale items. Items were contiguous within their scales, but their order within the scale was randomized, and the order in which each scale was presented was randomized. The only exceptions to this were the verbal aggressiveness scale itself, and the brief series of demographic questions. Because it was the focal point of this analysis, the verbal aggressiveness scale was always the first scale presented to every participant. The brief series of demographic questions were always presented last. Participants were presented with reminders to complete unanswered questions, but were never forced to answer any unanswered questions to proceed.

Measures

This study included one measure as the focus of analysis, the Verbal Aggressiveness Scale (Infante & Wigley, 1986), and also used five of the six scales against which the scale was originally validated. The other instruments were the Argumentativeness Scale (Infante & Rancer, 1982), the Assault and Verbal Hostility sub-scales from Buss and Durkee's (1957) Hostility-Guilt Inventory, the Crowne and Marlowe (1960) Social Desirability Scale, the Feelings of Inadequacy subscale from the Janis and Field Personality Questionnaire (Hovland & Janis, 1959), Personal Report of Communication Apprehension-24 (McCroskey, 1982), and Cognitive Complexity (Crockett, 1965).

This study will generally adhere to the original instruments used by Infante and Wigley (1986) to validate their scale, with one large exception, and one small exception. The large exception is the measure of Cognitive Complexity (Crockett, 1965). Though Cognitive Complexity was found by Infante and Wigley to be significantly correlated with their scale, and it is considered to be an effective instrument (O'Keefe & Sypher, 1981), the nature of the instrument makes it unwieldy to use in this context, because this study design was a large-sample online questionnaire. Cognitive Complexity is a free-writing instrument which must then be evaluated by trained coders. That makes it both cost and time prohibitive to use in this context.

It might seem paradoxical, in a study highlighting the ability of planned missing data designs to decrease the cost and time necessary to conduct research, to have to eliminate a scale from a study because it was cost and time prohibitive. This happens to be an unfortunate byproduct of the design, that planned missing data designs, as a result

of the inflated standard errors associated with multiple imputation, have slightly less power than an equivalent-*N* complete-case design (Graham, et al., 2002; Rubin, 1996). This can usually be accounted for by increasing *N*. However, an increase in *N* is inversely related to the ease of collecting open-ended, free-writing data, such as the Cognitive Complexity scale. As a result, there are some situations where planned missing data designs actually can make data collection more impractical than the equivalent complete-case design.

The small exception to the original instrumentation is the use of the PRCA-24. In the original, due to instrument length constraints, the authors chose to use the PRCA-Short Form (McCroskey, 1978). In this dissertation, through the use of a planned missing data design, this study was able to employ the PRCA-24, which McCroskey considers to be "preferable above all other earlier versions of the instrument" (McCroskey, 2007).

Verbal aggressiveness

Verbal aggressiveness (Infante & Wigley, 1986) is a unidimensional measure of aggressiveness, defined as "attacking the self-concept of the other person [in a dyad] in order to hurt the person psychologically" (p. 67). It is a 20-item, Likert-type scale with five response options from almost never true to almost always true. Example items include "I am extremely careful to avoid attacking individuals' intelligence when I attack their ideas" and "when others do things I regard as stupid, I try to be extremely gentle with them." Coefficient alpha in the validity studies was .81. In this study, $M = 2.54$, $SD = .54$, alpha = .87. Scale $M = 49.04$, $SD = 10.83$.

Argumentativeness

Argumentativeness (Infante & Rancer, 1982) is a two-dimensional measure, consisting of *approach* and *avoidance*, in which the general trend for argumentativeness is conceptualized as an interaction between a trend towards approaching arguments and a trend away from avoiding arguments. It is a 20-item, Likert-type scale with five response options ranging from almost never true to almost always true. Example items of the *approach* dimension include "arguing over controversial issues improves my intelligence" and "I enjoy a good argument over a controversial issue." Example items of the *avoidance* dimension include "I enjoy avoiding arguments" and "when I finish arguing with someone I feel nervous and upset." Previously measured internal reliability of the approach factor was extremely high at alpha = .91, and of the avoidance factor alpha = .86. Test-retest reliability r values of .87 and .86 for the two dimensions (respectively) suggest that this is a highly stable measure. Convergent and divergent validity relationships with existing measures suggest it is a valid measure (Infante & Rancer, 1982). In this study, the approach factor had $M = 3.03$, $SD = .67$, alpha = .86, while the avoidance factor had $M = 3.20$, SD =.60, alpha = 81.

Assault and verbal hostility subscales

The Assault and Verbal Hostility subscales from the Hostility-Guilt inventory measure tendency towards physical assault and verbal assault of another person (Buss & Durkee, 1957). They are 10 and 13-item (respectively) true/false scales. Example items from the assault subscale include "once in a while I cannot control my urge to harm others" and "if I have to resort to physical violence, I will." Example items from the

verbal hostility subscale include “When I disapprove of my friends’ behavior, I let them know it” and “ I demand that people respect my rights.”

This scale was one of the most heavily used self-report measures in psychology for almost 30 years, though the construct validity has been questioned in the intervening years since its use in Infante and Wigley’s (1986) original study, based on 1950’s-era scale-development practices that are now known to be inferior (Buss & Perry, 1992). Reliabilities have been reported for this scale ranging from .46 to .82 depending on the subset of items used (Velicer, Govia, Cherico, & Corriveau, 1985). In this study, the assault subscale had KR-20 reliability of .66, and the verbal hostility subscale had KR-20 reliability of .56.

Social desirability

The Social Desirability Scale (Crowne & Marlowe, 1960) is designed to measure the needs of participants to present themselves in a favorable light. It is a 33-item true/false scale. Example items include “at times, I have really insisted on having things my own way” and “I’m always willing to admit it when I make a mistake.” The KR-20 reliability of the measure has previously been reported at .88, and the test-retest reliability at .89 (Crowne & Marlowe, 1960). In this study, the KR-20 reliability was .63.

Feelings of inadequacy subscale

The Feelings of Inadequacy subscale from the Janis and Field Personality questionnaire (Hovland & Janis, 1959) is a heavily cited (see Fleming & Watts, 1980) measure of self-esteem. It is a 23-item Likert–type scale. Answer options differ slightly based on the working of the item, and could take the form of either very often to practically never, or very to not at all. Example items include “How often do you feel

inferior to most of the people you know?" and "do you ever think that you are a worthless individual?" The subscale has shown excellent reliability at .83 (Fleming & Watts, 1980). However, previous research has argued that there are serious problems with the subscale's construct validity (Church, Truss, & Velicer, 1980). In this study, $M = 3.02$, $SD = .55$, and alpha = 88.

Personal Report of Communication Apprehension

The PRCA-24 is a stable, valid, highly reliable measure of communication apprehension (McCroskey, 1982). It contains 24 Likert-type items, with five response options from strongly agree to strongly disagree. Example items include "I feel relaxed while giving a speech" and "I face the prospect of giving a speech with confidence. " Though the original verbal aggressiveness study used the PRCA short form, McCroskey (1978) notes that "since reliability and precision are reduced by the use of the short form, the long form should always receive preference" (p 203). In this instance, the use of planned missing data designs afforded me the opportunity to use the longer scale, an opportunity that the original authors did not have. In the original, the authors found this scale to have a reliability of .85 (Infante & Wigley, 1986). In this study, $M = 2.76$, $SD = .60$, alpha = .92.

Planned Missingness

Block Design

Efficiency designs are the methodological descendants of a class of designs employing Matrix Sampling, which at its most basic level is the practice of administering different items to different participants (Graham, et al., 2006). In the efficiency design, this has developed into the use of multiple forms of the questionnaire, in which items are

combined into blocks, participants receive fewer than all blocks on their measurement instrument, and participants are organized groups, which are each given a unique form of the instrument containing a unique selection of item blocks.

Split form designs can be created in two ways. First, scales can be split across blocks, such that several items from the scale appear in each block. Second, scales can be kept complete, such that items from a single scale never cross into multiple blocks. After establishing sample size and degree of missingness, this decision probably the most important one in planned missing data designs. Unfortunately, this decision is also less clear-cut, as there are competing opinions. Graham, Hofer, and MacKinnon (1996) determined in simulation studies that standard errors for estimates of variances and covariances were always smaller when forms used a split-scale design (putting questions from scales across blocks) as opposed to the complete-scale design (including complete scales in discrete blocks). They found that this gain in efficiency was strongest when the within-scale correlations were above .5. This suggests that using a split-scale design will result in least-inflated standard errors. Ultimately, "the impact of missing data on power will diminish as the correlations among the variables increase in magnitude" (Enders, 2010, p. 27).

In contrast to the split-scale designs, researchers can also choose to maintain scales as intact units in the design, such that entire scales are either observed or missing, but not partially missing. Graham, et al. (1996) tested this type of design in simulation studies, and found that complete-scale designs were most powerful when *inter*-scale correlations were high. That condition would have caused problems for this particular study, however. Because Infante and Wigley (1986) included the Argumentativeness

Scale (Infante & Rancer, 1982), which they hypothesized to have a null relationship with their Verbal Aggressiveness Scale, using a complete-scale design would have necessitated including scales which are known to have a minimal correlation to other scales in the study. This could have dramatically inflated the standard error, leading to problematic estimates. As a result, a split-scale design was the most appropriate choice for this study.

Because this study was deployed entirely online, through Qualtrics, I was presented with the opportunity to extend beyond existing discussions of the particulars of block types and sizes, and employ a design with constant random selection of items. The result of this process is, essentially, an "infinite" number of blocks, each with an approximately equal degree of missing information. As a result of this, the design discussion stops being about blocks, and becomes one purely about missingness and power. The number of blocks is limited only by the number of participants.

Simulation Comparison

To better understand how much missingness would be possible in this study, and how large of a sample would be necessary, and how powerful of a design would result, I located a simulation study that approximated the anticipated conditions under which this study was to be conducted. The study that best balanced these competing demands was the design used by Raghunathan and Grizzle (1995) in the simulation study component of their article. What that study provided for this dissertation was a model in how to structure a large sample (N of at least 1,000), high missingness study (at least 35% planned missing), and what parameter estimates would result from such a design. In this

study, they analyzed the impact of missing data on the mean, standard error, and confidence intervals of the data.

Their simulation study employed a three out of five block design, resulting in 40% missingness, and an effective increase of 60% in the number of items that can be collected without increasing demand on participants. Their results indicated that the efficacy of the design, the bias introduced, and the increase in 90% confidence interval depend most heavily on the correlation amongst items in different blocks and the skewness of the distributions of variables in the model. Given their simulation *N*=1,000, ideal conditions led to bias as low as .01%, with an inflation of 90% confidence interval as low as .6%. If the study conditions were less than ideal, especially the presence of low correlations across blocks and extreme skewness, sometimes the bias and CI inflation levels increased dramatically. However, middle estimates of between-block correlations suggest that biases of 2.2% and increases in 90% confidence intervals of 45% were a realistic potential outcome (Raghunathan & Grizzle, 1995, p. 61).

The correlation amongst items in different blocks was largely affected by item-total correlations in a scale, and how effectively items were spread across blocks in the measurement design. The randomized blocks approach that I employed in the study in this dissertation was effective at ensuring that approximately the same number of items from each scale are presented to each participant. This ensured high cross-block correlations and coverage of all measured constructs when the scales were unidimensional.

With a multidimensional construct, there is a chance that the randomization approach might not be as effective. It is possible –unlikely, but possible – that the

randomization measure could consistently select one dimension's items more frequently than another, which would force an over-reliance on imputed data for the under-selected items, and therefore the construct measured by those items. To avoid this potential scenario when using a multidimensional construct in a randomized blocks approach, the researcher must set the randomizer to draw items equally (or proportionally, as appropriate to the design of the scale) from each dimension.

Previous research suggests that an analysis using SEM with N=1,000, even with a 90% confidence interval increased by as much as 45%, will result in statistical power no worse than a fully-collected, non-imputed set with N=200 (Little, 2013). This is still large enough to have acceptable power to fit models of mundane size (df= 100) (Preacher & Coffman, 2006). Ultimately, this means that a properly constructed between-block planned missing design can function with only minimal bias and inflation of confidence interval relative to a fully observed study.

X Set

While each of the alternative scales was used in one comparison – against the Verbal Aggressiveness scale – the Verbal Aggressiveness scale itself was used in comparison against each other scale. Because of this, it was the most important scale in this analysis. When an analysis includes an item set that is distinctly more important to the overall analysis than any other item set, that set should be included as an X set. In a design that includes an X set, the X set is the group of items that is administered to each participant, unlike all other items, which are missing for some participants. The X set is distinct from the other item sets because the items in the X set are more critical to the analysis. Not all split form designs include an X set. The X set is used when "virtually

everyone should provide answers to the questions in this set" (Graham, et al., 2006, p. 327). In case of this study, the Verbal Aggressiveness scale functioned as the X set. Therefore, there was no planned missingness induced for this scale.

Data Analysis

After retrieving the raw data files from Qualtrics, I conducted a basic data cleaning process. Because the data was collected entirely online, and there was no free entry, there were no manual data input errors to screen for, or out-of-range values. There were blank cases to eliminate, however, because Qualtrics creates a case every time the questionnaire is opened on a web browser, regardless of whether any data are entered into the questionnaire. In total, 93 Qualtrics records indicated questionnaires created at least to the point of indicating "yes" on the informed consent page, but without data entered. These records were eliminated.

The next step in the data analytical process was multiple imputation, using Mplus 7.0 (Muthén & Muthén, 2011). Mplus uses a two-step imputation process, in which the first step is the generation of a distribution of potential values for imputation through the use of a Markov chain Monte Carlo (MCMC) algorithm. The second step is the process of actually imputing probable values using the generated distribution. The process of generating the distribution of potential values is an iterative process, in which the algorithm successively creates sample distributions. The first of these distributions are based largely on the Bayesian prior distribution, which in this case is an uninformed prior, and random. As the process continues, the algorithm relies more on observed data, and less on the uninformed prior. The process completes when the Markov chains achieve *convergence*, which occurs when the algorithm has generated a distribution based

on the observed, or posterior distribution (Sinharay, 2003). Although there are many measures for convergence, the built-in measure in Mplus is potential scale reduction (PSR), (Gelman & Rubin, 1992), which compares the current state of the distribution to a potential indefinitely-run algorithm, to establish how much more accurate the distribution could potentially become if it is left to continue forever. The minimum value for the measure is 1.0. In this dissertation, I saw starting PSR values of 1.2-1.5 depending on the dataset and the exact model specified. Gelman, Carlin, Stern, and Rubin (2004) suggest values under 1.1 for all parameters to be acceptable. To be conservative, I used a maximum PSR of 1.01 for this study. The cost to such a conservative PSR comes in the form of dramatically longer time to convergence – sometimes on the order of days. It was necessary to be that conservative, however, because the studies in this dissertation are designed to test the boundaries of the method, and it is important to be able to show the efficacy of PMDs even under stricter-than-necessary operating conditions, such as a maximum PSR value of 1.01.

The Gibbs (PX1) sampler used to generate the MCMC sequence from which the imputations were generated is one of the "most widely used Markov chain simulation methods" (Gelman, et al., 2004, p. 283). It is one of the default algorithms built into Mplus 7.0 (Muthén & Muthén, 2011), and is particularly effective at generating parameterizations with low mean square error in many-variable/low-*N* situations (Asparouhov & Muthén, 2010). For a point of comparison, Asparouhov and Muthén (2010) note that the Gibbs (PX1) sampler is ideal for imputations with large numbers of variables requiring imputation, and they define large as "50" (p. 7). This study had 146 variables used in the imputation model.

I set the number of imputations at 40. Previous research indicates that when the number of imputations is at least *m*=40, the relative efficiency of the multiple imputation procedure approaches optimal, even in high missingness ($\gamma > 0.50$) situations (Graham, Olchowski & Gilreath, 2007). In other words, the loss of statistical power is as low as .76% even when the rate of missing data is high, if sufficient imputations are used (Graham, Olchowski & Gilreath, 2007).

I then conducted the main analysis. This process also used Mplus 7.0 (Muthén & Muthén, 2011). After imputing the data, I created scale scores for each participant for each instrument, and conducted Pearson zero-order correlations between verbal aggressiveness and each construct it was being validated against.

Subsamples

Finally, to facilitate addressing PMD-related hypotheses, I also created a series of subsamples from the original cleaned data, pre-imputation. This entailed randomly selecting N cases from the full sample, and exporting those cases to separate files, where N = 100, 200, 400, 600, and 800 cases, respectively. Though this did not entail a full bootstrapping procedure with thousands of repetitions (as is done in Monte Carlo simulations), random draws (with replacement) from a larger sample each constitutes approximately random samples of smaller size. This subsampling was done before the multiple imputation process. I generated these subsamples for two reasons. First, I generated them to account for the fact that the full sample in this replication study was dramatically larger than the sample in Infante and Wigley's (1986) original study (their original N = 104). By looking at the relationships between the scales used in this study at smaller sample sizes, I was able to investigate whether the relationships found at the full

sample size were an artifact of that much larger sample size, or were relationships that would have existed in a smaller sample as well. Second, generating smaller subsamples pre-imputation allows me to comparatively test the robustness of the multiple imputation procedure at progressively larger sample sizes, by observing the change in Standard Deviation of parameter estimates at various sample sizes.

Standard Deviation provided me with a real-world measure that can be applied across subsamples, but which does not tend to change with sample size. This means that changes in Standard Deviation across sample sizes should be caused by the relative increases or decreases in efficiency of the imputation algorithm across sample sizes, not by the changing sample sizes themselves. Though mean squared error (MSE) is the measure that is usually employed when comparing the efficacy of differently constructed multiple imputation procedures in simulation studies (e.g., Asparouhov & Muthén, 2010), that measure is only readily available in the particular case of simulation studies, because it measures both imprecision and bias. Standard Deviation is actually an effective stand-in for MSE in non-simulation studies, because it is mathematically similar to MSE. In an unbiased estimator, square root of MSE is equal to Standard Deviation (Wackerly, Mendenhall, & Scheaffer, 2008). In other words, Standard Deviation is very similar to the proper (but unavailable) measure.

The imputation algorithm actually artificially inflates the Standard Error as a means of correcting for the uncertainty of replacing unobserved values, but because that is mathematically tied to the Standard Deviation, the Standard Deviation becomes an ideal way to isolate the robustness of the imputation procedure on point estimates at various sample sizes. Ultimately, while Standard Error is the best comparison between

current and past real data (when there is access to the original data), and mean square error is the best comparison in simulation studies, neither of those are available in a standalone replication without access to the raw data from the original study. Therefore, I compared Standard Deviations to measure the change in the degree of precision with which the multiple imputation procedure completed the data set at each of the varying sample sizes, including the full data set. Differences in Standard Deviation between sample sizes were compared with repeated measures ANOVA tests for each scale. Finally, because SPSS does not conduct post-hoc tests for repeated measures ANOVA tests that have only within-subjects effects, but no between-subjects effects (as was the case here), a series of paired-samples t-tests, which are appropriate in situations when tests are performed on matched or paired samples (Warner, 2008), served as stand-in post hoc tests to probe for pairwise differences beyond the omnibus F that will be reported by the repeated measures ANOVA. Appropriate corrections to alpha were made to correct for the possibility of familywise error.

Results

The goal of this study was to illustrate the utility and efficacy of planned missing data designs for communication research by replicating Infante and Wigley's (1986) Verbal Aggressiveness scale validation study. In order to accomplish this, a study-specific hypothesis was posed in this chapter, in addition to a series of PMD-specific goals that were posed in chapter 2. In this chapter, each hypothesis and goal will be restated, followed by the results of the statistical test conducted for each analysis.

Study-Specific Results

Hypothesis 4 posed a series of relationships between Verbal Aggressiveness and each of the comparison measures, such that Verbal Aggressiveness will be positively associated with assault, verbal hostility, and communication apprehension, and it will be negatively associated with social desirability and inadequacy. In the original study, the analysis consisted of a series of Pearson zero-order correlations between Verbal Aggressiveness and seven other measures (six of which were used in this replication) designed to establish the concurrent validity of the Verbal Aggressiveness measure. The results of this study's replication of that series of Pearson correlations are presented in Table 7. Also presented in Table 7 are the Pearson correlations for the series of randomly generated subsamples from the full data at intervals between N=100 and N=800. For comparison, the original set of Pearson correlations from Infante and Wigley (1986) is also included in the table.

Table 7
Correlations between Verbal Aggressiveness and Comparison Variables

	Relationship in original (Pearson *r*) *N* = 104	*N* = 100	*N* = 200	*N* = 400	*N* = 600	*N* = 800	*N* = 1020 (full sample)
Argumentativeness	-.04	.30*	.34*	.30*	.26*	.27*	.28*
Assault	.32*	.40*	.43*	.33*	.40*	.34*	.35*
Verbal Hostility	.43*	.36*	.32*	.34*	.37*	.35*	.36*
Social Desirability	-.39*	-.23*	-.15*	-.18*	-.21*	-.22*	-.23*
Inadequacy	-.15	-.15	-.06	-.04	-.07	-.09*	-.07*
Communication Apprehension	.25*	.10	.16*	.18*	.14*	.19*	.16*

Notes: * $p < .05$.

Full Sample

Results from this study indicated that the associations between Verbal Aggressiveness and four of the six comparison measures were consistent with the hypothesis presented previously, and therefore the results of the original study. Verbal Aggressiveness was positively associated with verbal assault which was the same direction of relationship, and a similar magnitude of relationship to the original study. Verbal Aggressiveness was positively associated with verbal hostility, which was the same direction of relationship, and a slightly lower magnitude of relationship to the original study. Verbal Aggressiveness was positively associated with communication apprehension, which was the same direction of relationship, and a moderately lower magnitude of relationship to the original study. Verbal Aggressiveness was negatively

associated with social desirability, which was the same direction of relationship, and a moderately lower magnitude of relationship to the original study.

Two results in this study demonstrated inconsistencies with the original results. The first was the association between Verbal Aggressiveness and Argumentativeness. In this study, there was a significant relationship. However, in the original study, the predicted relationship was null, and the results in the original study supported that prediction (Infante & Wigley, 1986). Despite the fact that Verbal Aggressiveness and Argumentativeness were conceptualized as orthogonal (Infante & Wigley, 1986; Infante & Rancer, 1982), and that conceptualization was supported by the findings of Infante and Wigley (1986), more recent research suggests that that it was the results of this study which were in line with the extant body of literature, and it was the original validation study that has become increasingly isolated in its results (Levine, Kotowski, Beatty, & Van Kelegom, 2012). Levine, et al. (2012) performed a meta-analysis of 11 studies that measured the relationship between Verbal Aggressiveness and Argumentativeness, comprising a total of 2,349 participants, and found a weighted mean correlation of .45 between the scales. In light of their findings, this study appears to have properly replicated the relationship between these two scales, and it is the original Infante and Wigley (1986) study that misidentified the nature of the relationship between Verbal Aggressiveness and Argumentativeness.

Additionally, the correlation between Verbal Aggressiveness and Inadequacy was also significant in this study, which was consistent with the hypothesized relationship from the original study, though inconsistent with the results from the original study, as their results found no significant relationship between the variables. Clearly, this

disparity is an artifact of the much larger sample size employed in this study, as the magnitude of the relationship was actually smaller in this study (r = -.07) than it was in the original (r = -.15). Finally, In addition to the correlations between Verbal Aggressiveness and the six comparison measures, the original study also reported a significant correlation between the Buss-Durkee Hostility Scale (Buss and Durkee, 1957) and the Argumentativeness scale (Infante & Rancer, 1982) r = .25, N = 104, p < .05, to confirm expectations about how those two comparison scales should perform relative to each other. That correlation was exactly replicated in this study, r = .25, N = 1020, p < .001.

Subsamples

Analysis of the subsamples indicates that at the N = 800, N = 600, and N = 400 levels, all of the hypothesized relationships between Verbal Aggressiveness and the comparison measures maintained significance at the p < .001 level, with the exception of the relationship between verbal aggressiveness and feelings of inadequacy, which was non-significant at the p = .05 level for all sample sizes below N = 800. The non-significance of the relationship in the subsamples was not caused by the decreasing sample size. In fact, exactly the opposite situation occurred, such that the dramatically larger sample size of the full sample caused there to be a significant correlation without there being any stronger magnitude of a relationship. Verbal Aggressiveness and inadequacy have the same magnitude of relationship at both the N = 1020 and N = 600 levels, but the relationship is significant at the p = .05 level in the full N = 1020 sample, while it is non-significant at the p = .05 level in the N = 600 sample. This suggests that the significant correlation in the full sample is, in fact, an artifact of the greatly increased

sample size, and that this study has successfully reproduced the magnitude of the relationship between Verbal Aggressiveness and the feelings of inadequacy scale from the original study.

Planned Missing Data Results

This study also tested two of the overarching dissertation goals related to Planned Missing Data Designs. Specifically, this study tested whether the data collection process from Infante and Wigley's (1986) Verbal Aggressiveness validation study could be conducted in dramatically reduced time through the use of an efficiency-type Planned Missing Data Design, and whether the it could be conducted without a problematic decrease in the precision of parameter point estimates due to the multiple imputation process.

Time Reduction

With regards to the first goal, results indicated that the mean time that participants spent completing the web-based questionnaire was 22.34 minutes (SD = 69.85, N =1020). However, because Qualtrics only measures the time difference between the time that a new session is opened and the time that session is closed, there is no guarantee the participant is actually completing the questionnaire the whole time. Indeed, one participant took 974 minutes, or about 16 hours, between starting and finishing the questionnaire, and I highly doubt he/she was entering data non-stop. Further investigation of results indicate a median time between start and finish of M =12.00 minutes, 5% trimmed mean time between start and finish of M = 12.90 minutes, and distributional data indicating a positively skewed (= 9.143) and highly kurtodic (= 92.423) distribution of scores. Though Infante and Wigley (1986) did not specifically note how long it took their

participants to complete the questionnaire, their design did require three separate data collections for the single questionnaire because the instrument would have been too burdensome for one collection. This indicates that this dissertation's goal of a significant reduction in the time necessary to administer the data collection process was met.

Estimate Precision

With regards to the second goal, previous research suggests that an imputation model of the size and scope of the one implemented in this study, using the type of model used in this study, and the number of imputations employed in this study, should be operating at or near peak efficiency (Asparouhov &Muthén, 2010; Graham, Olchowski, & Gilreath, 2007). This indicates that there should be very little power lost relative to a fully observed study, although, again, *actual* comparisons can only be made in a simulation study.

Mplus reports two types of fit statistics to assess the model fit of a multiple imputation model: The Posterior Predictive Check, which, used in this context, measures the ability of the model to generate data that is similar to the observed data (Rubin, 1984), and the Kolmogorov-Smirnov test, which compares the distributions of the observed and imputed data for each variable and provides a significance test for the difference (Abayomi, Gelman, & Levy, 2008). In the case of this study, the two tests gave confounding results. The PPC reported a value $p < .001$, which indicated that the imputation model generated data significantly different from the observed data. However, the K-S test, which reports a *p*-value for each parameter, reported that in none of the parameters did the imputed data differ significantly from the observed data.

Another way of assessing the functioning of the multiple imputation procedure in this study was to inspect the parameter estimates across the various subsamples, to establish whether or not those parameters were estimated with greater precision at larger sample sizes. I conducted repeated measures ANOVAs for each scale, and these tests indicated that as sample size increased, the precision with which parameters were estimated increased significantly. The results from these tests are detailed in Table 8.

Table 8

Repeated Measures ANOVA Results Table for Scale Variables

	Wilk's Lambda	F( )	F =	*p*	Partial η^2
Verbal Aggressiveness	.003	(5,15)	1049.60	< .001	.997
Argumentativeness	.010	(5,15)	310.75	< .001	.990
Assault	.013	(5,5)	73.42	< .001	.987
Verbal Hostility	.005	(5,8)	346.21	< .001	.995
Social Desirability	.007	(5,28)	780.59	< .001	.993
Inadequacy	.018	(5,18)	196.60	< .001	.982
Communication Apprehension	.007	(5,19)	521.00	< .001	.993

Note: Table measures change in Standard Deviation between N = 100 subsample and N = 1020 full sample

Because the subsamples were drawn from the full sample, the results from the repeated measures ANOVAs indicate that the only difference between the various sample sizes is the relative performance of the multiple imputation procedure. Paired-samples t-tests with alpha set to .01 to reflect the 5x chance of familywise type I error served as post-hoc tests for each repeated measures ANOVA. These tests indicated that with a single exception, each scale was measured with significantly more precision at each successive sample size than at the next lowest sample size. The results from these paired-samples t-tests are detailed in Table 9.

Table 9

Change in Standard Deviation as Sample Size Increases.

	N = 100-200	*N* = 200-400	*N* = 400-600	*N* = 600-800	*N* = 800-1020
Verbal Aggressiveness	-.027	-.019	-.009	-.004	-.005
	12.08***	13.76***	11.46***	5.01***	7.68***
Argumentativeness	-.033	-.026	-.012	-.010	-.004
	9.50***	17.32***	8.89***	10.94***	4.08***
Assault	-.020	-.012	-.005	-.004	-.002
	9.03***	6.62***	5.36***	4.92***	3.28**
Verbal Hostility	-.023	-.012	-.006	-.004	-.002
	11.45***	10.19***	7.94***	6.33***	2.80(ns)
Social Desirability	-.020	-.013	-.006	-.003	-.003
	18.0***	23.79***	13.5***	7.38***	7.87***
Inadequacy	-.040	-.029	-.011	-.008	-.005
	10.58***	12.55***	8.89***	6.61***	4.27***
Communication	-.040	-.027	-.011	-.005	-.006
Apprehension	11.56***	17.61***	10.30***	5.57***	8.73***

Notes: For t-statistics: $**p < .01$, $***p < .001$..*Includes Paired-Samples t Statistics on Second Line.*

In sum, these results show that the greatest impact of additional participants came in the jump from *N*=200 to *N*=400 participants. The average effect size (Cohen's *d*) for

that increase in sample size (averaged across all of the scales) was 3.61, compared to 3.14 for the *N*=100 to *N*=200 increase, 2.25 for the *N*=400 to *N*=600 increase, 1.61 for the *N*=600 to *N*=800 increase, and 1.25 for the *N*=800 to *N*=1020 increase. This indicates that while additional participants did have a positive effect on the precision of parameter estimates at all sample sizes, each individual person had the greatest impact between *N*=200 to *N*=400. Beyond that point, the marginal value of each participant decreased monotonically.

Discussion

Introduction

This study had one main goal, which was to test the efficacy of an efficiency type planned missing data design in a communication context. To do that, this study employed the replication of an established study, in this case the validation study for a communication research measure (Infante & Wigley, 1986). Taken as a whole, the results of this study strongly support the argument that planned missing data designs can be of significant value to communication researchers. This study replicated the results of the original study, and it did so while putting a greatly reduced data collection burden on participants.

Summary of Findings

The main finding from this study is that an efficiency type planned missing data design can be effective at dramatically reducing the length of the questionnaire employed in communication research. This benefit can be realized with only minor negative implications for the study's ability to realize its internal goals, or, in some circumstances, no negative implications whatsoever.

This study presents and supports the argument that an efficiency-type planned missing data design is a viable, valid method for conducting quantitative research in communication through three findings. The first finding is that data collection can take place with dramatically reduced burden on participants using an efficiency-type planned missing data design. The second main finding is that the multiple imputation procedure functioned effectively. Finally, the third main finding is that this study was able to largely replicate the relationships between variables found by Infante and Wigley (1986) in their Verbal Aggressiveness scale validation process through the use of a planned missing data design. In the place where this study's results diverged from the original study's results, further research in the intervening 27 years supports the findings of *this* study, not the original (Levine, et al., 2012). Though various forms of efficiency-type planned missing data designs have been used in other fields (ex. Raghunathan & Grizzle, 1995; Adigüzel & Wedel, 2008; Graham, Olchowski, & Gilreath, 2007), they have not yet been tested in a communication research context, and these findings lend substantial support to the argument that they should be used in this context.

Reduction in Participant Burden

One goal of this study was to establish whether a planned missing data design could reduce the time necessary to complete a questionnaire, and thereby reduce the burden to the participant completing it. The first finding of this study is that this assertion is supported by the data in this study, and therefore the goal was accomplished. In the original study, Infante and Wigley (1986) describe their process of data collection as follows:

> the instruments were administered to respondents during three sessions, with two-week intervals between sessions, two or three instruments per session, by three different investigators. This procedure of making the study appear to be three separate and independent studies was followed to reduce demand characteristics." (p.65)

This was a convoluted design, necessitated by the very large questionnaire that was used. It involved multiple people, multiple time points, and multiple trips to a research site for every participant, in an effort to prevent the total measurement instrument from imposing a too-great demand on participants if presented all at once.

In the current study, each participant spent an average of 12.90 minutes completing the online questionnaire, between the time their web browser initially loaded the page, and the time they either finished and submitted the questionnaire or closed the browser without finishing. By any measure of comparison, from three separate collections over the course of four weeks to an average of 12.90 minutes is a dramatic difference in time commitment on the part of the participant. This reduction in time commitment and participant burden is due in large part to the planned missing data design employed in this study.

Previous research suggests that once questionnaires reach 100 items in length, there is a potential loss of validity due to the length of the instrument (Andrews, 1984), and that in general, the longer a questionnaire becomes, the lower response rates become, the less variation there is in those responses, and the more response rate decreases (Galesic & Bosnjak, 2009). All of these contribute negatively to validity. By being able

to reduce the length of the questionnaire to fewer than 100 items, this study was able to reduce the effect of such design problems.

There were other factors that contributed to the ability of this study to collect data in fewer than three measurement points. Most importantly, the use of an online questionnaire meant that participants could complete the questionnaire from a location of their convenience at a time of their convenience. Additionally, the deletion of the Cognitive Complexity scale (Crockett, 1965) shortened the process of data collection, because that measure requires 5 minutes of prompted free-writing by the participant. However, even taking the use of the online questionnaire and the deletion of a scale into account, in light of the recommendations of Andrews (1984) about the decline in instrument validity after 100 items, this study would not have been replicable in one data collection without the use of a planned missing data design. If this study had been conducted with a fully-observed design, each participant would have been required to complete 147 items, enough for there to be a risk to validity. In sum, while there are other technological factors that have been invented in the intervening 27 years that have streamlined the data collection process, it is only the use of a planned missing data design that allowed data to be collected at a single time point, dramatically reducing the time commitment and demand on study participants.

Multiple Imputation Process

Another goal of this study was to conduct the replication without a significant inflation of standard error relative to the original. Making a direct case for the accomplishment of this goal is difficult in this study, absent the original Infante and Wigley (1986) data for comparison, but the results from this study make a strong indirect

case for the performance of the multiple imputation procedure. Circumstantial evidence to the efficacy of the imputation process was available in the form of Standard Deviations for every point estimate made by Mplus. Those results, when averaged across scales and viewed across sample sizes, were able to be measured and tested against each other at each successive sample size, and present a compelling case that the multiple imputation procedure functioned exactly as it was expected to in this study.

The variation between estimates generated by the multiple imputation procedure should fall as sample sizes increase (Schafer & Graham, 2002). In small sample sizes, the imputation is based largely on the Bayesian prior, which expresses uncertainty before the data are taken into account. However, as sample sizes increase, the likelihood function increasingly takes over. As this happens, the data are drawn from the posterior distribution, which takes into account both uncertainty, but also the information contained in the data, leading to less variation within the imputed values as uncertainty is replaced by information (Gelman, et al., 2004). This reduction in variation (as measured by Standard Deviation) is exactly what is expressed in the data in this study, demonstrating that the Multiple Imputation procedure was functioning as expected.

Replication Findings

The third finding from this study is that the efficiency type planned missing data design was capable of recreating the original results, and that almost 30 years later, the relationships between Verbal Aggressiveness and the comparison measures are largely unchanged. In this study, the relationships between Verbal Aggressiveness and five of the six comparison measures (the only exception was Argumentativeness) were consistent with the results from the original study. Verbal Aggressiveness was positively associated

with Assault, Verbal Hostility, and Communication Apprehension, and negatively associated with Social Desirability. There was a statistically significant negative association with Feelings of Inadequacy, but this was an artifact of the increased sample size, as the magnitude of the association was approximately equivalent to the magnitude in the original study. Verbal Aggressiveness was positively associated with Argumentativeness in this study, which was in disagreement with the original results, but is in accordance with the last 27 years worth of research involving the two constructs (Levine, et al., 2012).

The magnitude of the relationship between Verbal Aggressiveness and Communication Apprehension, and Verbal Aggressiveness and Social Desirability were both less than the magnitude of the respective relationships in the original study. However, the fact that the relationship between Verbal Aggressiveness and Verbal Assault was greater than in the original, and relationship between Verbal Aggressiveness and Argumentativeness was significantly greater, suggests that this reduced magnitude is not due to any problems with the methodology. Because this methodology is capable of measuring increased relationships, we know that any reductions in observed magnitude are due to other factors, not some muting factor inherent to the design hiding the relationships.

Summary

In summary, what this study does is provide support for the argument that the efficiency-type planned missing data designs are valuable methods for use in communication research. It does this by establishing that these designs function in a theoretically expected manner in a communication research context, and accomplishes

the goal of dramatically reducing the time required to complete the questionnaire in this study and the corresponding level of participant burden. Most importantly, this study gives warrant for the conducting of further original research using planned missing data designs in communication..

Limitations

Despite the potential this study has for advancing knowledge in communication research, there are several limitations that must be acknowledged. First, the instrumentation used in this study was not ideal. Both the Buss and Durkee subscales, and the Janis-Field Feelings of Inadequacy scale were considered in the original study (and therefore in this study) to be well-constructed unidimensional measures. However, subsequent research has shown this not to be the case. The Buss and Durkee subscales were designed using 1950's procedures, and as a result, do not hold up as completely valid, regardless of how many times they had been used in the intervening years (Buss & Perry, 1992).

An additional limitation is generalizability. Though much communication research is done on large, Midwestern university campuses, and is of limited ability to be generalized to the broader public, this study added two points, one positive, one negative, to that initial level of generalizability: Making this study's results more generalizable is the large sample size. Few studies in communication are larger than N= 1000, as this study was. Conversely, making this study's results less generalizable is the fact that the sample drawn for this study was not exactly representative of the University. It was too first- and second-year student heavy, and it was also over-represented with women compared to the university at large.

Also, in highlighting the strengths of the efficiency-type planned missing data design, this study has also demonstrated a weakness. This study was not able to completely replicate the original study, because the design of this study prevented it from including the Cognitive Complexity measure (Crockett, 1965). Because the Cognitive Complexity measure is free-writing based, it was too labor intensive to include in this study.

Finally, there are two limitations to address with regard to the multiple imputation procedure. First, random subdraws from the full N=1020 sample are not the same thing as bootstrapped subsamples. Bootstrapped subsamples would more precisely approximate independently created samples drawn from the same population as the full sample. Given sufficient time and training to generate those results, bootstrapped subsamples likely would have resulted in less jitter in the parameter estimates, especially in the smallest subsamples (seen in the correlations between scales).

The second limitation with regard to multiple imputation is that in the absence of simulation data to compare the imputation results against, there is no way to assert the quality of the imputation process against an external metric. I would like in the future to conduct such Monte Carlo simulations to establish exactly how close to the theoretical ideal efficiency each imputation model has performed.

Conclusion

The goal of this study was to present a method for reducing participant burden by shortening the measurement instrument through the use of an efficiency type planned missing data design. The results of this study show that this method can be effective at achieving this goal without affecting the quality of information gathered.

This study was also specifically designed to test the limits of the efficiency type planned missing data design. Because of that, this study used as long a questionnaire as was feasible, as high a degree of missingness as was feasible, and yet results indicate that the methodology was still successful in providing significant advantages relative to the status quo methodology for conducting this type of study.

Furthermore, additional analyses from the subsamples indicate that this study still would have been able to replicate the series of relationships between the variables with fewer than half of the total participants that this study included, so the results are not simply a case of benefitting from a particularly generous sample size. This suggests that future researchers will have an even greater degree of flexibility to use this methodology as a powerful tool for streamlining and improving the quality of their research than has been shown in this study, and would not be limited to circumstances where they have access to samples as large as this one. In other words, future researchers will be able to use this method to model complex interactions with many fewer participants than were used in this study. This study, therefore, demonstrates efficiency type planned missing data designs as an effective tool that should be added to the methodological toolbox of communication research.

CHAPTER 5: STUDY 3 - ACCELERATED LONGITUDINAL COMMUNICATION COMPETENCE

Introduction

Study three was an accelerated longitudinal replication of Rubin, Graham, and Mignerey's (1990) study of college students' communication competence. By replicating the study using an accelerated longitudinal design, this study attempted to demonstrate the advantages conferred by planned missing data designs in terms of time saved and reduced participant attrition.

In the original study, the authors followed students from their entry into college to the fall of their senior year, surveying them every fall. The study collected data about how the students' self-reported (and other-observed) levels of communication competence changed over the course of their four-year college career. The study encountered one of the main problems that a longitudinal study can encounter – attrition. The initial *N* was 50, but by the last (4^{th}) measurement point, that *N* had dropped to 20. Additionally, the study required three years to collect data (Fall 1984 – Fall 1987). I replicated Rubin et al's study using a cohort-sequential longitudinal design, collecting data in 13 months. The accelerated design allowed for the faster completion of longitudinal data collection, which in turn reduces the likelihood of participant attrition over time.

Hypotheses

In addition to the hypotheses about planned missing data designs that were presented in chapter two, this study also tests study-specific hypotheses that are derived from the results found by Rubin, et al (1990). The original study did not include

hypotheses or research questions, but a prior study, working with a preliminary version of the same data set (Rubin & Graham, 1988), did present the variable relationships that were to be tested with the data. The hypotheses here are based on this information.

H_5: Communication competence is positively associated with Grade Point Average.

H_6: Communication-related experiences in high school will be positively associated with GPA.

H_7: Communication-related experiences in high school will be positively associated with communication competence.

H_8: Change in communication competence over a college career will be best explained by a positive linear trend.

Study Design

Sample Size

As was the case for study one, before conducting a power analysis, I established potential upper and lower bounds for degrees of freedom for the structural models that were to be used to analyze the data. To create these bounds, I followed the guidelines set forth in Rigdon (1994), which led me to an upper-bound of approximately 10,000 degrees of freedom for a completely un-parceled model including each scale item as an indicator in a structural model, although that number would have changed very slightly depending on the final arrangement of the structural paths in the model. A parceled model, including exactly three indicators for each of the seven latent constructs in this study, would have established an approximate lower bound at 168 degrees of freedom, again with some uncertainty based on the final arrangement of structural paths. This lower bound established a conservative estimate for the power analysis.

After establishing degrees of freedom, the next step was to use that information to conduct a power analysis, beginning by estimating null and alternative RMSEAs. MacCallum, Browne, and Sugawara (1996) suggest that the appropriate construction of hypotheses for testing model fit is to test a null hypothesis of not-good-fit against an alternate of good fit. Accordingly, the power analysis for this study employed a null hypothesis value of RMSEA = .08, and an alternative hypothesis value of RMSEA = .01, based on the more conservative definition for the boundary of acceptable fit and poor fit (Kline, 2010; Browne & Cudeck, 1993). I prefer the more conservative RMSEA = .08 boundary instead of the more liberal RMSEA = .10 boundary of acceptable fit and poor fit (Bollen & Curran, 2006) because I tried to consistently be conservative in making estimates during this power analysis (and all others) so as to avoid underestimating the eventual necessary sample size. I conducted the power analysis based on assumptions of alpha = .05, power = .80, degrees of freedom = 168, null RMSEA = .08, alternative RMSEA = .01, and it produced a suggested N of 50. This provided a conservative lower bound for the sample size needed for good model fit. From that power analysis, I determined that the ideal initial recruitment N would be 200, equally split across year in school. This N allowed for large attrition at each time point without compromising analytical power. For example, 20% attrition at each time point would have resulted in a cumulative attrition of 48.8% and a final N of 102, still more than double the minimum recommended N.

Recruitment

Initial recruitment for this study took place in February 2012, using the COMS research pool. Unfortunately, the research pool was not able to provide enough juniors

and seniors for the study. Of the 107 participants recruited only xx% were juniors and xx% (or n) were seniors. To augment my sample size, I initiated a secondary recruitment, using an email sent to all COMS majors who were not in a research pool-attached course during that term, which yielded an additional 47 participants, for a total initial N of 149. Although this N fell short of the planned sample size of 200, with efforts taken to combat participant attrition throughout the study, and subsequent recruitment efforts, I hoped that the final N would be sufficient.

Because far fewer participants than expected returned to the COMS research pool at measurement point two, I used the secondary recruitment as an ongoing recruitment method for any participant who had previously been in the study, but was not in the COMS research pool at that measurement point. I compared a complete list of previous participants against the list of research pool participants, and any previous participants who were in the research pool were returned to my study. Any previous participants who were not in the research pool were contacted outside of the research pool. This ensured that each participant, whether they were in a research-pool attached class or not, would be able to return to participate in subsequent measurement points.

Participants in the research pool received 1.5% course credit. Participants recruited through the secondary external recruitment received the chance to win a $10 gift card. If a participant was in a research pool-attached COMS course, they were entered into this study through the research pool, not through the external recruitment, even if they had previously been recruited through the secondary recruitment procedure.

Participants

A total of 530 unique participants contributed data to this study at least once. Only one participant provided data at all four time points, 19 participants provided data at three time points, and 79 provided data at two time points. The remaining 431 only provided data at a single time point. 470 participants (88.6%) were COMS majors enrolled in classes during the 2011-12 and 2012-13 school years; the remaining 60 (11.4%) were COMS minors.

This sample, which includes a significant number of COMS majors, had more females than the general university population, which is about 52% female (Ohio University Office of Institutional Research, 2012). In total, this sample was 68% female. Nearly all participants (n = 98%) were between the ages of 18 and 22. There were a total of 124 participants in the first-year cohort, 125 participants in the sophomore cohort, 127 participants in the junior cohort, 30 participants in the senior cohort, and 126 additional participants in the incoming first-year cohort at time point 3.

Cohort-Sequential Design

Cohort-sequential designs are based in a different approach to measuring change than a traditional longitudinal design. In a traditional longitudinal design, each participant is measured across the entire span of the research timeframe, from the beginning to the end, and the study takes an amount of time to conduct equivalent to the research timeframe, because change happens in real-time. In a cohort-sequential design, participants are grouped into cohorts based on a common time-related factor (e.g., age, number of years working in an organization, grade level). This time-related factor is then used to organize the cohorts for entry into the study at different points along the research

timeframe. In the case of this study, instead of measuring one group of students as they change over four grade years, this study measured four cohorts of students, each exactly one grade-level apart, as they changed over one year. All four years' worth of change was captured in 13 months instead of four school years.

Cohort-sequential designs require the existence of measurement points that can be equated between cohorts, called overlaps. In a cohort-sequential design, longitudinal change is measured in the form of line segments that are measured simultaneously, instead of one continuous line, the way it is measured in a traditional longitudinal design. The overlap points are necessary because to reconstruct these line segments into a continuous line to model change over the measured timeframe, researchers need to be able to measure how each cohort relates to the next cohort. In the case of this study, after measuring each grade-level cohort simultaneously, the next step was to figure out how the first-year cohort related to the sophomore cohort, how the sophomore cohort related to the junior cohort, etc. Accomplishing that goal is the purpose of the overlap points.

In this case, since the cohort was defined as a class year, the overlap was the February measurement point, which happened at the same time in two successive class years, at time points one and four. Participants in the first-year cohort, measured at time point four, were actually being measured at the same point in their college career as the sophomore cohort measured at time point one. Though two or more overlaps would be ideal, it is possible to fit models that use one overlap as this study did (Little, et al., 2007; Little, 2013).

Procedures

I collected data at four time points: February 2012, June 2012, October 2012, and February 2013. At each time point, I contacted four separate cohorts of students, representing each class year in school (i.e., first years, sophomores, juniors, and seniors). At time point three, a new cohort of incoming first-years entered the study (and, conversely, outgoing seniors left the study after time point two). This resulted in a cohort-sequential design with one overlapping collection point (measurements one and four).

After recruitment, participants completed the questionnaire online using the survey system Qualtrics. Before starting the questionnaire, participants read an informed consent document and either provided their consent to participate, which took them into the questionnaire, or did not provide their consent, which took them to a message of thanks for their time. To minimize potential order effects, I programmed Qualtrics to randomly present the items within each scale. Also, I randomized the order of the scales except for the demographic section which always was presented last.

Measures

This study used all but one of the measures from the original study (Rubin, et al., 1990). Thus, the measures included the Personal Report of Communication Apprehension (PRCA-24) (McCroskey, 1982), the Interaction Involvement Scale (IIS) (Cegala, 1981), and the Communication Competence Self-Report (CCSR) (Rubin, 1985). This study did not use the Communication Competence Assessment Instrument because that measure is a participant observation measure, involving a panel of trained observers evaluating a student's demonstrated skill in presenting a speech (Rubin, 1982a; Rubin, 1982b). In much the same way that Cognitive Complexity (Crockett, 1965) was too unwieldy to

include in study two, the CCAI would have overly taxed the resources available for this particular dissertation. The three scales in this study are described below.

Personal Report of Communication Apprehension

The PRCA-24 is a measure of communication apprehension, with very high predictive validity (McCroskey, 1982). It contains 24 Likert-type items, with five response options from strongly agree to strongly disagree. It measures communication apprehension in four contexts: public speaking, dyadic interaction, small groups, and large groups. Example items include "I feel relaxed while giving a speech" and "I face the prospect of giving a speech with confidence." Table 10 details the alpha, mean, and standard deviation scores for the four measurement points in this study.

Table 10

Alpha, Mean, and Standard Deviation scores for the PRCA

	Wave 1	Wave 2	Wave 3	Wave 4
Alpha	.93	.93	.94	.95
Mean	3.52	3.43	3.41	3.34
Standard Deviation	.59	.64	.64	.68

Interaction Involvement Scale

The IIS (Cegala, 1981) is a measure of the tendency to demonstrate both attentiveness and perceptiveness in interactions. It is an 18-item scale, with seven response options from "not at all like me" to "very much like me", It has previously

demonstrated Kuder-Richardson-8 reliabilities in the high .80's for its three factors (Cegala, 1981), and appears to be a valid measure of the cognitive dimension of communication competence. The scale measures three dimensions of interaction involvement: attentiveness, perceptiveness, and other-oriented perceptiveness. Example items include "I am keenly aware of how others perceive me during my conversations" and "Often in conversations I'm not sure how I'm expected to respond." Table 11 details the alpha, mean, and standard deviation scores for the four measurement points in this study.

Table 11

Alpha, Mean, and Standard Deviation scores for the IIS

	Wave 1	Wave 2	Wave 3	Wave 4
Alpha	.88	.88	.88	.89
Mean	5.00	4.96	4.99	5.05
Standard Deviation	.76	.79	.76	.80

Communication Competence Self-Report

The CCSR (Rubin, 1985) is a measure of communication competence designed to be a self-report version of the previously-established Communication Competence Assessment Instrument (Rubin, 1982; Rubin, 1985). It is a 19-item scale, with five response options from always to never. It is internally reliable (alpha coefficient .87), demonstrates construct validity (Rubin, 1985), and correlates as expected with known

measures, such as the PRCA-24 (McCroskey, 1982). The items on the CCSR mirror the competencies present in the CCAI. Example items include "When giving a speech, I speak clearly and distinctly" and "I have to answer a question several times before others seem satisfied with my answer." Table 12 details the alpha, mean, and standard deviation scores for the four measurement points in this study.

Table 12

Alpha, Mean, and Standard Deviation scores for the CCSR

	Wave 1	Wave 2	Wave 3	Wave 4
Alpha	.87	.88	.85	.86
Mean	3.67	3.63	3.57	3.56
Standard Deviation	.53	.49	.44	.48

*Previous Experiences Scale***.**

This scale is a measure of previous communication-related activities. It is a 16-item scale, with five response options ranging from "no experience" to "extensive experience". This scale is based on a list of high school communication-related activities established previously (Rubin & Graham, 1988; Rubin, et al., 1990), but where previous research asked respondents to list whether they had any experience with each activity, this study has adjusted the response options to aid in employing this scale as a formative construct in a structural equation model (Edwards & Bagozzi, 2000). No prior reliability

measures were available for this scale. Table 13 details the alpha, mean, and standard deviation scores for the four measurement points in this study.

Table 13

Alpha, Mean, and Standard Deviation scores for the Previous Experience Scale

	Wave 1	Wave 2	Wave 3	Wave 4
Alpha	.85	.87	.86	.88
Mean	2.77	2.75	2.78	2.69
Standard Deviation	.71	.74	.74	.76

Planned Missingness

In this study, the missingness took a different form from the previous two studies. In the previous two studies, the missing data was almost tangible – it was data that could have been collected, but was not. In this study, the missingness came in the form of time. Because this was an accelerated longitudinal design, the missingness was time saved: time that was not spent waiting between measurement waves, time that did not pass between the earliest cohort to which a study can generalize and the latest cohort. By completing what was a 3 calendar year design in 13 months, time was gained. In this type of a planned missing design, the missingness did not appear on the SPSS data file, but it appeared in the study design because it did not require three years to collect data to draw conclusions about change over three years. Because this was planned to be a fully-

observed study, multiple imputation was not expected play a substantial role in this analysis, except to fix minor, unplanned missing data. There was no plan to conduct imputation of large amounts of planned missing data as there was in either of the previous studies.

Data Analysis

The analysis phase of this study was to take place in Mplus 7.0 **(**Muthén & Muthén, 2012), which would have enabled me to build a longitudinal structural equation model (Little, et al., 2007; Little, 2013) to model the change in levels of the latent constructs and the inter-individual changes and relationships between constructs over the course of measurement points. Unfortunately, due to problems in the design, recruitment, and data collection phases, the ability to conduct the analysis relied far more heavily on an in-depth multiple imputation procedure than had been planned and this procedure could not be completed successfully with the available data. As a result, the analysis was halted.

Results

The original design of this study anticipated the recruitment of approximately 200 participants, distributed equally across four class levels. At the time of initial recruitment before measurement point one, the recruitment was planned to take place entirely through the research pool. When it appeared that the research pool could not provide enough participants, a secondary, external recruitment was conducted to supplement the initial participant numbers, and to ensure that, at later measurement points, participants who had previously participated, but were not then in a research-pool attached class, could still participate again.

When initially designing this study, I judged the research pool to be a robust enough recruitment method to ensure that participants would be returning in subsequent quarters. Previous experience, including the other studies in this dissertation, suggested that studies in the research pool regularly achieve greater than 90% response rate, and I assumed that this historical success at participant retention would translate longitudinally. Even in the (assumed) unlikely case that the COMS major-only participants would not be a part of the research pool in subsequent terms, the secondary recruitment for participants who were not in a research pool-attached class should have been enough of a draw to ensure that participants returned for later time points. As it turned out, these were both flawed assumptions, and this study experienced extraordinarily high attrition.

In the case of the research pool, the design was such that upper-level classes were not required to participate in the pool, and it was at the discretion of each instructor whether these classes participated. A few instructors opted in, but far more than anticipated opted out. As a result, senior-level participants were under-represented from the very beginning, though this did not represent a fatal design flaw initially, had most of the initial participants returned at subsequent time points. However, it later became clear that the research pool would not return the original participants in nearly sufficient numbers, leaving this study in a position of having to rely on the secondary recruitment, which initially was conceptualized as serving a supporting, not primary role. Because the secondary recruitment did not attract enough participants to the study at later measurement points, either, it quickly became necessary to recruit fresh participants through the research pool at measurement points two, three, and four. This had the

eventual effect of bolstering the total N, but increasing the number of participants who only participated in one, or maybe two measurement points.

An additional complication was the fact that Ohio University was still using the quarter system in the 2011-2012 school year. As a result of this scheduling, measurements one and two were in separate academic terms. Measurement one (in February) was during winter quarter, and measurement two (in June) was during spring quarter. This meant that for a participant to be assigned to this study through the COMS research pool at all four measurement points, they had to be enrolled in a participating COMS class in each of four successive academic terms. Had this study been started in February 2013, running through February 2014, in comparison, participants would only have had to be enrolled in a participating COMS class in *three* successive academic terms, because the first two measurement points would have been conducted at the beginning and end of the same term (Spring 2013). As a result of still being on the quarter system, combined with low participation rate from upper-level COMS classes, the needed number of participants was neither initially recruited nor retained.

What compounded this failure to retain participants, however, was the failure of the secondary recruitment to attract any large degree of interest from potential participants. The return rate from the secondary recruitment, which was as low as eight percent at the third measurement point, was surprising. Based on social exchange principles (Dillman, Smyth, & Christian, 2008), the secondary recruitment should have been far more effective than it was: Cost for participation was low because the questionnaire was web-based and could be completed at any time, there was an incentive prize for taking the questionnaire, and the contact list for the secondary recruitment was

based on prior participation lists, suggesting that participants were people who were already inclined towards participation.

The combination of recruitment failures led to a sample that barely resembled the intended design. Instead of a nearly completely-observed design in which ~ 200 participants were measured at all four measurement points (and multiple imputation was used only in the more traditional, data triage sense), the data set included a total of 530 unique participants. Table 14 details the breakdown of participants by grade year at each measurement point.

Table 14

Number of Participants by Measurement Wave and Grade Year

	Wave 1	Wave 2	Wave 3	Wave 4
First-Year	19	20	59	67
Sophomore	39	56	45	68
Junior	48	47	44	48
Senior	21	16	26	33
Total	127	119	174	216

Even with all of these procedures in place, of the 530 participants, only one person provided data at all four measurement points. Because so few participants provided data at multiple points, making any intra-individual change claims from the data

relied almost completely on the multiple imputation procedure. More importantly, the lack of repeat participants drove down the covariance coverage, which is the measure of how much information a variable has in common with any other variable. If a variable is 50% missing, and another variable is also 50% missing, then their covariance coverage would be 25%. Mplus' multiple imputation procedure functions most effectively at covariance coverages above 50%, and returns an error if any covariance coverage in the model is below 10% (Brown, 2006; Muthén & Muthén, 2012). To run this analysis, I had to override the default settings in Mplus to allow for covariance coverages below 10%.

Overall, the data set had 69% missing data. This level of missingness is almost identical to that of the other studies in this dissertation, but in this case the missingness was unanticipated. In the previous two studies, this level of missingess was both anticipated and within the ability of the multiple imputation procedure to complete the data set. In the case of this study, the number of variables was too great, and the total N too low, to multiply impute such a high amount of missing information.

As was done in the two previous studies, an additional correction to the planned study N could have been made to account for the presence of missing data, but was not, because I judged that the corrections already made beyond the minimum N (to account for attrition at each time point) would be sufficient to allow for the multiple imputation of any minor unplanned missing data. Because the missingess in this study came in the form of time that did not have to be spent during the data collection process, and not in the form of data not collected, multiple imputation was not supposed to play a significant role in the analysis phase of this study.

Analysis

The analysis phase of an accelerated longitudinal PMD study can be conducted from either of two distinct approaches. It can be done using structural equation modeling (e.g. Little, et al., 2007; Little, 2013) or it can be done through growth curve modeling (e.g., Duncan, Duncan, & Strycker, 2005; Duncan, Duncan, & Strycker, 2006). Each of these presents challenges specific to that analytical approach, but each allow the researcher to prioritize particular analyses from the data. The circumstances of this study were such that neither analytical approach was a viable option, and as a result, no analysis could be performed on the data. This results section presents the data that was collected in this study, the approaches available for analysis of the data, including the relative strengths and weaknesses of each approach, and explains how those strengths and weaknesses intersected with the circumstances of this study to prevent any analysis from going forward.

Given the research goals of this particular study which include both inter-individual, correlational hypotheses (e.g., Communication-related experiences in high school will be positively associated with communication competence) and an intra-individual univariate hypothesis (Change in communication competence over a college career will be best explained by a positive linear trend), a strong case can be made for either analytical approach. Structural Equation Modeling might be most appropriate on the basis that it more readily models complex variable interrelationships. Although Growth Curve Modeling is capable of modeling these interrelationships, the basic form of a growth curve is a univariate time-sequence model, and any multi-level or multivariate extensions of this basic model test the analytical skills of the researcher in a

more significant fashion than does SEM (Geiser, 2012; Duncan, Duncan, & Strycker, 2006). Conversely, GCM is capable of measuring intra-individual change over time, which SEM cannot do, and GCM can more easily measure group-level change over time than can SEM. The slope parameter provided in the GCM model provides a single value to represent direction of change, which can be tested for its difference from zero. In the SEM framework, in the absence of a single slope parameter, the researcher must calculate parameter differences across each time point, which, though possible, requires stacking hypothesis tests. However, along with advantages presented by the two potential analytical options, there are also disadvantages to each option as well.

To begin, despite what the design intentions of this study were before beginning data collection, during data collection it became readily apparent that either analytical approach would require multiple imputation playing a much greater role in the analytical phase of this study than had been previously anticipated. Conducting an SEM-based analysis would require that the data be imputed at the parcel level, using the duplicate-scale imputation technique used in the longitudinal public speaking anxiety study (Rubin, 2010). This would be necessary for an SEM-based analysis because for the model to be locally identified, multiple indicators per construct (ideally three) (Little, 2013) are required. In the particular circumstances of this data, that variable structure results in an imputation model with 3654 parameters (as reported by Mplus). When I attempted to multiply impute the data so that the accelerated longitudinal structural equation model could be fitted, the MCMC sequence that precedes the imputing of the data took almost four days trying to converge before failing to produce imputed data sets. Because the rate of convergence in the MCMC chain is related to the number of parameters to be

estimated (Asparouhov & Muthén, 2010), the sample N, fraction of missing information, and total number of parameters in the data become critical in determining the number of parameters that have to be estimated. Although the sample N and fraction of missing information cannot be changed at the analytical stage, it was possible to reduce the number of parameters that have to be estimated by changing the shape of the variable structure within the data, by switching to a CGM-based analysis

Although SEM functions best with three indicators per construct, the same data could be analyzed through growth curve modeling, which in its most basic form only requires one indicator per construct per measurement point (Duncan, Duncan, & Strycker, 2006). Attempting to multiply impute the data from this study under the assumption that it will be analyzed using GCM, which is effectively a true duplicate-scale imputation (Enders, 2010) leads to an 84% reduction in the number of parameters that the multiple imputation procedure needs to estimate (from 3654 in an SEM format to 560 in a GCM format). The resulting multiple imputation procedure converged successfully and created the imputed datasets in approximately 4 hours.

Unfortunately, constructing an accelerated longitudinal analysis in a growth curve context proved to be dramatically more challenging than constructing the equivalent model in a SEM framework, and was beyond the scope of this dissertation. As a result, this study sits in a state of limbo, without a final word on what, if any, results will be found. It sits stuck between one analytical method that the data will not support (SEM), and another that the researcher cannot yet support (GCM). There may yet be fascinating results to speak of in this data, but they cannot be extracted at this point in time. Currently, the value of this study is as an example of how a boundary condition of the

accelerated longitudinal design can be encountered if sufficient attention is not paid early in the design process, or if there are unanticipated problems with recruitment that prevent participants from returning to the study.

Discussion

Insofar as this dissertation has been about providing an argument for the implementation of planned missing data designs in communication research, this study adds to that argument despite failing to accomplish its study-specific goals. Because the goal of this dissertation has always been modeling how to implement PMD designs in a robust fashion across different contexts in communication research, the studies in this dissertation were all designed to be test cases, to demonstrate the full extent of what the method is capable of doing. As a result of choosing ambitiously designed test cases, it has always been a possibility that one of the studies in this dissertation might not be successful. That possibility happens to have been realized in this study.

Fortunately, the failure of this study to accomplish its internal goals does not warrant any conclusions being drawn about the efficacy of planned missing data designs for longitudinal studies. This study was flawed by a facet of the design unrelated to, and implemented in advance of, planned missing data designs. What this failure does do, however, is illuminate one potential way that PMD studies can fail and highlights a limiting boundary condition of the design – in this case, the large study N requirement relative to non-PMD designs.

Though the cause of the failure of this study occurred earlier in the design, the effect of the failure was to encounter a boundary condition of planned missing designs. Planned missing designs are limited by the ability of the imputation procedure to function

effectively, and this is governed by (broadly speaking) the number of parameters that have to be imputed, the fraction of missing information, and the study N. If the researcher fails to properly conduct a power analysis by anticipating and correcting for the presence of missing data (Enders, 2010), or there is an unusually high rate of unplanned missing data (as happened in this study) the end result could be a study N that is too low to support the multiple imputation procedure.

Even though the planned design for this study did not necessitate making an adjustment to the study N to account for a multiple imputation procedure (because originally, there was not supposed to be any imputation in this study), in hindsight, failing to do so was a mistake. Knowing that I was using an untested (longitudinally, at least) recruitment system should have signaled the potential for unanticipated attrition in the recruitment system itself. Had I paid sufficient attention to this potential problem from the start, I could have recruited greater numbers at each measurement point. Best practices for this type of design in the future should include making an adjustment for much higher unplanned missing data than expected, and the need for the use of multiple imputation.

CHAPTER 6: DISCUSSION

The primary goal of this dissertation has been to demonstrate the efficacy of planned missing data designs for ameliorating particular methodological challenges faced in communication research. In chapter one, I presented the complex nature of longitudinal designs and the demanding nature of overly long questionnaires as two particular methodological challenges that could potentially have measureable, negative impacts on the research done in communication. I demonstrated those negative impacts by comparing how frequently longitudinal research is being done in the field relative to how frequently longitudinal research is being stated as needed in the future. I posed the high cost of conducting longitudinal research, the extensive time commitment necessary for longitudinal studies, and the lack of training in longitudinal designs as three major causes for the lack of longitudinal research. I also demonstrated the negative impacts of overly long questionnaires by presenting relevant literature which argues for a relationship between increased questionnaire length and decreased validity of measurement (e.g., Andrews, 1984; Galesic & Bosnjak, 2009; Herzog & Bachman, 1981).

Having posed the challenges, I then conducted three studies to establish whether PMDs would solve these challenges. Though all of the studies included in this dissertation were survey research-based, their findings are more broadly applicable to a wide-range of communication methodologies, including experimental designs, quasi-experimental designs, and content analytical designs. The main findings from this dissertation indicate that PMDs are powerful methodological tools which are of considerable value to communication research, though they come at the cost of increased

methodological complexity which can increase a study's chance of failure. It is up to the researcher and his/her understanding of the particular research context to know whether the gains that PMDs can offer are worth the increased cost of design complexity.

The first main finding, from study one, is that PMDs are effective at helping researchers reduce participant burden in longitudinal research, which helps to minimize attrition over the course of the study design. The first study achieved 92% participant retention over a semester-long measurement design. The second main finding, originating from study two, is that PMDs are effective at allowing researchers to shorten the questionnaires used in research without affecting the relationships between the constructs measured by that questionnaire. The second study was able to successfully replicate Infante and Wigley's (1986) verbal aggressiveness scale validation study through the use of an efficiency-type PMD. The third main finding is that PMDs add complexity and points of failure to research designs. The experiences from the third study indicate that PMDs are complex unto themselves and add a layer of complexity beyond the equivalent completely-observed design. As a result, PMDs require the constant attention of a skilled researcher. Even otherwise familiar research designs cannot be conducted on auto-pilot when they are implemented in a PMD format. Assumptions that can be made in a non-PMD context cannot necessarily be taken as writ in a PMD context (as study three shows). If insufficient attention is paid to the ramifications of each and every step in the design phase, then the likelihood of failure, as happened in study three, greatly increases.

Finding One

The first main finding from this dissertation is that PMDs can be powerful tools for making longitudinal research more accessible to communication researchers. This is

particularly valuable, because as identified in chapter one, a significant challenge facing communication research is the disparity between the rhetoric and the reality of conducting longitudinal studies. As the review of top communication research journals I conducted indicated, scholars far more frequently suggest that someone in the future should do longitudinal research in a particular area (18% of articles) than actually conduct longitudinal research (11% of articles). The low rate of longitudinal research, and the disproportionately higher rate of calls for longitudinal research combine to form a particularly urgent challenge for a discipline based in the study of processes and interactions that inherently have temporal components.

Study one indicates that PMDs can be powerful tools for making longitudinal research more accessible to communication researchers. Through the use of a controlled enrollment-type design (Little, et al., 2007), the results from study one indicate that PMDs can reduce demand effects due to repeated measurement at multiple waves. By reducing the demand on each individual participant from five measurements (in a five-wave fully-observed study) to three, this study has lowered the effort for participation. Because the effort required to participate in a study is one of the most important barriers to participation (Dillman, et al., 2008), this study improved the perceived cost-benefit calculation for participants. Improving the cost-benefit calculation for participants makes many-wave study designs more practical for researchers, because it reduces the likelihood of attrition over the course of the study. In study one, 92% of recruited participants were retained, for a total loss rate of 8%, which is an improvement over previous longitudinal research in communication. For comparison, other semester-long,

classroom-based longitudinal studies in communication have encountered loss rates ranging from 23% (Duff, et al., 2007) to 67% (Frymier, 1994).

In comparison to the status quo, PMDs offer a considerable advantage in attrition reduction. By reducing the effort required to participate in longitudinal research, PMDs make it less likely that participants will drop out of the study. By lowering attrition, PMDs make it easier to do the types of research that the discipline consistently calls for but less consistently does – longitudinal research that can show causation through time ordering.

The results from study two can also be applied to longitudinal research in communication, even though it was a cross-sectional design. Through the use of an efficiency PMD in which fewer than every item on the questionnaire was presented to every participant, this study showed how PMDs could be used to reduce the time commitment participants are asked to make when recruiting them for a study. In study two, the use of an efficiency PMD allowed the questionnaire to be shortened from its original 147 items to 95. The total time commitment for each participant to complete the instrument went from (approximately) several hours spread over three measurements, to one measurement of length M = 12.9 minutes. Previous research indicates that the length of a questionnaire negatively impacts not only the number of participants who begin completing the instrument in the first place, but also the number who eventually complete it (Galesic & Bosnjak, 2009). The impact of the effect of instrument length is magnified for longitudinal research, where attrition is frequently monotonic – once participants decide to leave a study, they often never return.

The benefits to longitudinal research realized from study one can also be viewed from the perspective of cost reduction. Though the instruments used in this dissertation were all online questionnaires, and presented little marginal cost for additional participants or additional measurement waves, this is not the rule in communication research. There are researchers in communication who use instruments that are far more costly to deploy than a Qualtrics questionnaire, and PMDs offers them a way to field a longitudinal study at reduced cost. Using the longitudinal design from study one as a model, a researcher could conduct a five-wave design with the cost of a three-wave design, because each participant is only measured three times. Some researchers may be in a situation where the prospect of a potential 40% reduction in cost (three measurements instead of five) may enable them to do research they could not previously afford to do. Taken together, this dissertation presents a compelling argument that PMDs can be of substantial benefit to the communication researcher who is interested in conducting longitudinal research.

Finding Two

The second main finding from this dissertation is that PMDs are effective at allowing researchers to shorten questionnaires without affecting the nature of the relationships between the constructs being measured. Currently, when communication researchers are interested in conducting complex, multi-construct studies, they may be faced with a situation where their ideal questionnaire is too long to be effective. There is a body of literature which establishes that beyond a certain point (approximately 100 items), there is a negative impact of length on instrument validity (e.g., Andrews, 1984; Galesic & Bosnjak, 2009; Herzog & Bachman, 1981). Without PMDs, the choices of

what to do with overly long questionnaires are, unfortunately, not ideal. Ad-hoc questionnaire shortening (e.g., Lapinski & Orbe, 2007) is premised on the assumption that the remaining items had perfect item-total correlations to the previously validated scale. If this is not the case, then the previous validation cannot be assumed to transfer to the ad-hoc short form. Using the short form of the desired scale can also be problematic, because it can reduce reliability (McCroskey, 1978). Of course, this option assumes that a previously validated short form already exists. If there is no previously validated short form of the measure, then the researcher is put in the position of either ad-hoc short forming, or doing the entire scale validation process themselves, which is not an ideal choice.

Study two provides clear evidence that PMDs can provide a benefit to communication research by presenting researchers with another option for how to deal with overly long questionnaires. Efficiency PMDs allow researchers to reduce the length of the questionnaires employed in their studies in a more valid and less potentially time consuming manner than previous options. This provides communication researchers with a method to shorten the length of their questionnaire while still using every item in every scale. This is an improvement over existing options available to communication researchers, who may currently find themselves in the position of deciding which items to randomly throw out, or which scale to remove in order to shorten a too-long instrument.

Ultimately, the net effect of using an efficiency-type PMD is to give the researcher control over the relationship between the number of items in a questionnaire and the number of items presented to participants, which exist in a 1:1 relationship in most questionnaire designs. With an efficiency-type design, the researcher has control

over how many of the total items each participant sees. A researcher may choose to reduce the length of a too-long questionnaire, as discussed above, but they may also choose to add more measures to an instrument without increasing the length, thus measuring more constructs than they otherwise could. For example, in our School, we recently began pilot testing a public speaking lab, and one component of the lab experience is a pre- and post-test questionnaire. The questionnaire is designed to measure whether the lab is meeting its goals of improving students' presentational skills. One of the primary requirements of this questionnaire was that it had to take no longer than ~5 minutes to complete. There were a variety of constructs that would have been interesting to measure – apprehension, willingness to communicate, competence – but our ability to include all of those constructs was limited by the time constraint. Had we chosen to use an efficiency PMD, we could have shortened every scale on that questionnaire to between 60% - 75% of its original length, and could have measured more constructs without the questionnaire taking any longer to complete.

Finding Three

The third main finding of this dissertation is that PMDs add complexity and points of failure to a research design. This finding was not intentional, but was a byproduct of the unanticipated failures of study three. Though much of this dissertation has been dedicated to making the affirmative case for the implementation of PMDs, it must be acknowledged that they are more complex to implement than an equivalent completely-observed design. Even dedicated and attentive researchers are presented with more opportunities to make minor mistakes that can lead to major problems when

working with any type of PMD. Such major problems cause failures as serious as a complete loss of a study, as happened in study three.

The most obvious point of complexity introduced through the use of PMDs is multiple imputation (MI). MI is a complex mathematical procedure with many potential points of failure. And yet, even getting a study to the point of implementing multiple imputation requires managing additional complications. For example, in study one, implementing the efficiency design involved properly designing a randomization process through which each participant would be presented with fewer than all of the items on each measure. Web-based survey-research programs like Qualtrics make this process easier, but researchers still need to have familiarity with the Qualtrics interface, and a proper attention to detail. Delivering participants to study one's online questionnaire meant recruiting 500 participants, and then organizing them into 10 groups of 50, each with a different data collection schedule. It also meant scheduling 10 different reminder email schedules, because each group progressed through the study design at a different pace.

Most importantly, using multiple imputation adds complexity because it makes it necessary to attend to the number of model parameters early in the analytical process. If your analysis involves SEM (like study one, here), then you would have had to deal with the issue of parameters anyway, and multiple imputation simply makes you confront them earlier through a process such as duplicate-scale imputation (Enders, 2010) which might be necessary if you have an excessively large number of parameters. If, on the other hand, your eventual analysis is less complex, you may not have to be concerned with exactly how many parameters your data set contains. In this case multiple

imputation adds complexity, because you will have to pay attention to the number of parameters in the imputation phase, whether or not you are otherwise interested in them later in your analysis. Failing to attend to the number of parameters before the imputation phase can prevent the imputation from converging, or lead to dramatically slower convergence times (as seen in study three, in the SEM analysis attempt).

In study two, implementing the controlled-enrollment design involved the added complexity of managing a 10-group design, in which various groups proceeded through the measurement waves at different times. This group design dramatically complicated scheduling recruitment and reminder emails.

In study three, a series of flawed assumptions about recruitment methods and participants retention rates led to a situation in which multiple imputation became necessary even though it had not been part of the original plan for the study. This put study three in the particular situation of encountering a boundary condition for PMDs, a condition where this type of design would fail, but where a similar, non-PMD study might not have failed: A very large number of items (which increase parameters), few repeat participants (which drive down covariance coverage), SEM-based analysis (which necessitate multiple indicators for every construct), and a very large rate of unplanned missing data. This is not to say that unplanned missing data are necessarily fatal to a PMD. On the contrary, study one had 16% unplanned missing data, and this unplanned missing data did not have any negative impact on the outcomes of that study. Ultimately, it is a matter of degrees. Study three experienced a rate of unplanned missing data beyond the rate of planned missing data in studies one and two, and I had not put in place the necessary mechanism to attempt to cope with that level of unplanned missing data.

The silver lining of study three's failure is how quickly it became apparent. Because study three was attempting to replicate a study that took four academic years to complete, a failure of recruitment/retention in a traditional longitudinal version of this study would have meant the loss of that same four academic years. Because study three used an accelerated longitudinal design, data collection was completed after only 13 months. Even though the study was not successful, I could still restart study three at the beginning of the next academic year, and have it completed more than a full calendar year before a traditional longitudinal design would have finished.

This study clearly highlights the fact that planned missing data designs are not magic methods that can solve all problems. Rather, they are effective ways of making a wide variety of research less burdensome, cheaper, faster, or more valid. But they introduce complexities of their own that require a diligent attention to detail at all steps in the research process. More than anything else, conducting a planned missing data design study requires constant attention to the fact that the multiple imputation procedure that enables time-saving, burden-reducing, or instrument-shortening needs to be considered throughout the research process, not just in the analytical phase. If the researcher is not fully appreciative of these complexities at all times, the possibility exists that the study can break down in a way that a fully-observed study might not.

Strengths, Limitations, and Future Research

The conclusions I advance in this dissertation are contextualized by the strengths and limitations of my research. In offering ideas for future research and next steps in PMDs for communication research, it is important to acknowledge the limitations of this

research. The following section discusses both the strengths and limitations of this research and presents guidance for future research in this area.

Strengths

The strengths of this dissertation are both the depth of the evidence it provides in favor of the adoption of PMDs, and the breadth of contexts across which this evidence has been presented. This dissertation has provided strong evidence that PMDs can be effective additions to communication research by demonstrating their successful application in both original and replication research.

By employing PMDs in the context of communication research, this dissertation has provided evidence that PMDs can make many-wave longitudinal studies less burdensome on participants. It has also provided evidence that PMDs can be used to shorten the length of a questionnaire without removing any items from the scales contained in the instrument, or negatively affecting the results of the study. It has shown that the benefits of this method are as applicable to original research on public speaking apprehension today as they are to a replication of the validation study of a 27-year-old interpersonal communication theory. By demonstrating that this method is not limited to any particular context of communication research, the conclusions from this dissertation have much more import for the discipline.

Limitations

Though this dissertation has shown that PMDs can have an immediate impact on communication research across a wide variety of contexts, it has not been completely successful at accomplishing its own internal goals, and there are limitations that need to be acknowledged. The most important limitation to address is the failure to demonstrate

the efficacy of PMDs to shorten the timeframe of longitudinal research through accelerated longitudinal designs. There are also, however, additional limitations to discuss.

Study two demonstrated some of the advantages to be gained by using a PMD in a longitudinal context, which directly addressed the most pressing methodological challenge posed in the beginning of this dissertation. At the same time, I was not able to demonstrate a solution to the *most important* challenge related to longitudinal research, which is the extensive time commitment required to conduct longitudinal studies, due to the often long timeframe of longitudinal designs. In chapter one, I detailed the reasons for a lack of longitudinal research in communication, and the (often) multi-year commitment to a study, which is difficult at most parts of a research career, was highlighted as being of particular importance. Demonstrating a solution to this problem was the primary goal of the third study, which used an accelerated longitudinal design to compact four academic years' worth of change into one year of actual research time. Unfortunately, this study was not completed. It is still possible that a future analysis of the data, using a growth curve modeling approach, may lead to useful conclusions that support the goal of the dissertation, but those conclusions cannot be drawn at present.

Another limitation of this dissertation was the choice of constructs to include in the questionnaire in study one. In hindsight, the hypotheses posed in study one, and the scales used to test those hypotheses, were not the most in-depth investigations. Though the Personal Report of Public Speaking Anxiety (McCroskey, 1970) and the Self-Perceived Public Speaking Competency Scale (Ellis, 1995), were both effective, if I were

to do that study again, I would ask better questions and include more constructs in that instrument to allow for a more interesting analysis.

Future Research

Based on the work presented in this dissertation, there is a rich potential for future research using PMDs both as a continuation of work in this dissertation, and simply using this dissertation as a guide. The types of designs presented in this dissertation have applicability to a wide variety of research contexts, and there are even more types of PMDs that have not been presented here. Additionally, I see the possibility for application of PMDs to new methodological contexts.

The most logical future work to come out of this dissertation must be another attempt at study three, this time with a more effective recruitment and retention strategy. Knowing what I know now about the efficacy of longitudinal recruitment through a research pool, I would be able to more effectively retain participants throughout the research design. Beyond a more effective participant recruitment and retention strategy, I would also base my sample size requirements on an assumption that multiple imputation might be necessary if unforeseen problems arise, especially in participant retention.

One existing type of PMD that has not been presented in this dissertation, but which could also be of great service to communication research, is the "two-method measurement" type (Graham, et al., 2006, p. 324), sometimes referred to as the "gold-standard" type design (Rhemtulla & Little, 2012, p. 426). In this design type, a cheap measure of a construct (frequently a questionnaire measure) is used in tandem with a high quality but expensive measure of a construct. The cheap measure is collected from the full sample, while the gold-standard measure has missingness inserted for a percentage of

the sample. This has the effect of lowering the cost to conduct the study, or, viewed from another perspective, improving the statistical power per cost unit ratio (Graham, et al., 2006). This could have direct and immediate impact on any researcher whose agenda deals with expensive lab work. For example, the extensive work of Kory Floyd into biological responses to communicative stimuli (e.g. Floyd, Mikkelson, Hesse, & Pauley, 2007; Floyd, Boren, Hannawa, Hesse, McEwan, & Veksler, 2009) frequently involve such lab work, and future research like this could benefit from the application of a gold-standard type PMD.

Another potential avenue for advancement of the discipline through this method is in scale development. The process of developing a new measure involves a preliminary step of testing a large pool of items (Clark & Watson, 1995) to establish their performance with a pilot sample of participants. This initial item pool is, by design, much larger than then final measure will be, because some of the initial items will fail to perform as expected in some way. Items might not measure exactly the right construct, they might dual load onto multiple constructs, they might have high error variance, or they might be worded in a confusing or double-barreled way. Any number of minor quirks in the wording of an item can ruin its ability to measure the intended construct, and it is impossible to predict with certainty how participants will interpret items until they are actually asked to do so. The process of pilot testing a large initial pool of items is an important step in the scale development process. Unfortunately, it can involve asking participants to complete instruments that are well above the known upper limit for questionnaire length before causing potential validity problems due to participant burden, which is about 100 items (Andrews, 1984). For example, Ledbetter's (2009) development

of the Measure of Online Communication Attitude scale used an initial pool of 115 items. Even larger than that was the initial item pool of the Intercultural Development Inventory (Hammer, Bennett, & Wiseman, 2003). This began with a 239-item initial pool, which was administered to a pilot sample, followed by a "smaller pool of 145 items" (p. 429) administered to a full sample of 226 participants. This dissertation has already established the potential for efficiency-type PMDs to shorten questionnaires without affecting the structure of relationships between measured constructs. A logical follow-up context for applying this same principle is in scale development.

Another potential area of future methodological research might be in the impact of differing levels of missing data in a multi-group (or multi-condition if experimental research) research design. Because of the way the multiple imputation algorithm inflates standard error, groups with more missing data may have an artificially increased within-group variance than groups with less missing data. This could potentially affect any conclusions drawn from the F-value derived from the between-group/within-group ratio. These differing levels of missing data across groups may indicate the presence of a hidden problem in group assignment, such that the change in the rates of missing data across groups are associated with group membership. In an experimental design with random assignment to conditions, differing levels of missing data may indicate a problem with the random participant assignment process. In a non-experimental design, it may only indicate that the missing data is not MCAR. This area is ripe for future methodological research.

Future research should also extend the principles and methods of PMDs to other, non-survey research-based methodologies. Experimental research and quasi-experimental

research could immediately benefit from the cost savings offered by gold-standard PMDs. This cost savings could allow the researchers to increase the study N, or it might allow them to expand the study design to include more data collections, potentially increasing the explanatory and predictive power of their research. Content-analytic research, which can be time and labor intensive, might benefit from the application of an efficiency design. In this application, data sources could be intentionally under-sampled (i.e., left missing in the data set) and the missing data could be multiply imputed. PMDs also present additional applications in exploratory research. One interesting potential future application for controlled-enrollment PMD is in exploratory experimental studies where the optimal time lag is unknown. Identifying the correct time lag in any type of multi-wave study is a known methodological concern (Slater, 2007). By using a controlled-enrollment PMD, researchers can include more measurement waves than they otherwise would be able to include (given the same level of expected participant demand characteristics). They can then cluster these additional measurement points around the hypothesized ideal lag point to give confidence that they are correctly measuring the correct lag point even if the hypothesized measurement does not have the exactly correct lag time.

Finally, future research should, and will, continue to advance the statistical bases and effects of PMD through comparisons of different imputation approaches and the advancement of those methods, and the effects of non-normal data as suggested by Enders (2010). As an observer of the current topical discussions on, for example, the SEMNET LISTSERV, would discover missing data theory is still a rapidly evolving area of study with ongoing discussion from leading scholars (Hayduk, 2013). This

development of missing data theory could translate into new best practices for PMD in years to come. It is not improbable to think that of the missing data handling strategies which are considered relatively equivalent today, such as MI and FIML, one or the other may be preferred or disfavored in the future as a result of ongoing developments in missing data research.

Conclusion

The central conclusion of this dissertation is that PMDs are valuable tools for use in communication research. They are not, however, panaceae, and are neither necessary nor applicable in all research contexts. From the beginning of this project, I have been describing PMDs as methodological tools that more researchers should be trained to use. And just like any tools, they help solve problems in some, but not all situations. I hope that as a result of the work done in this dissertation, more communication researchers will be equipped to conduct PMD-based studies if they can be of benefit to the particular research situation.

Planned missing data designs are not separate types of research that should be considered alone from other research designs. Rather, they are designs that should be considered to be complimentary, or additive to, a wide variety of currently-employed research designs. Just like any other research design, whether they should be used for any particular study should be assessed by the individual researcher on the basis of the needs of the research situation.

As this dissertation has shown, the cost/benefit calculation for employing PMDs is clearly established. The cost to a PMD-based study is in the marginally lower statistical power and marginally inflated standard error of estimates. However, these are known and

understood quantities and can be accounted for (eg., Graham, Olchowski, & Gilreath, 2007). The benefits of PMDs come in the form of enabling research that might otherwise not be possible, or making more valid research that might otherwise be undermined due to participant demand concerns. On balance, this dissertation indicates that the benefits of PMDs exceed the costs in a wide range of research situations, and will be of great value to communication researchers when they are more widely adopted.

Finally, in the long term, I hope that the adoption of PMDs will not inhibit fully-observed longitudinal research from being conducted in communication. As has been mentioned in this dissertation, there are some types of questions that are not well suited to PMDs, and it is critical that communication researchers continue to conduct time-based inquiries with fully-observed longitudinal designs as well.

REFERENCES

Aasland, O. G., Olff, M., Falkum, E., Schweder, T., & Ursin, H. (1997). Health complaints and job stress in norwegian physicians: The use of an overlapping questionnaire design. *Social Science & Medicine, 45*(11), 1615-1629.

Adigüzel, F., & Wedel, M. (2008). Split questionnaire design for massive surveys. *Journal of Marketing Research (JMR), 45*(5), 608-617. doi:10.1509/jmkr.45.5.608

Andersen, P. A., Buller, D. B., Voeks, J. H., Walkosz, B. J., Scott, M. D., Cutter, G. R., & Dignan, M. B. (2008). Testing the long-term effects of the go sun smart worksite health communication campaign: A group-randomized experimental study. *Journal of Communication, 58*(3), 447-471. doi: 10.1111/j.1460-2466.2008.00394.x

Andersen, P. A., Buller, D. B., Walkosz, B. J., Maloy, J., Scott, M. D., Cutter, G. R., & Dignan, M. B. (2009). Testing a Theory-Based Health Communication Program: A Replication of Go Sun Smart in Outdoor Winter Recreation. *Journal Of Health Communication*, *14*(4), 346-365.

Andrews, F.M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *The Public Opinion Quarterly 48*(2), 409-442.

Arbuckle, J. L. (1996). Full information in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum.

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. Technical Report. Los Angeles: Muthén & Muthén.

Aubrey, J. S., & Taylor, L. D. (2009). The role of lad magazines in priming men?s chronic and temporary appearance-related schemata: An investigation of longitudinal and experimental findings. *Human Communication Research, 35*(1), 28-58. doi: 10.1111/j.1468-2958.2008.01337.x

Ayabomi, K., Gelman, A., & Levi, M. (2008). Diagnostics for multiple imputations. *Journal of the Royal Statistical Society (Applied Statistics) 57*(3) 273-291.

Backor, K., Golde, S., & Nie, N. (2007). Estimating Survey Fatigue in Time Use Study. Paper presented at the 2007 International Association for Time Use Research Conference, Washington, D.C.

Baer, J., & Schmitz, M. (2000). Latent growth curve modeling with a cohort sequential design. *Social Work Research, 24*(4), 243-248.

Ballard, D. I., & Seibold, D. R. (2000). Time orientation and temporal variation across work groups: Implications for group and organizational communication. *Western Journal of Communication 64*(2), 218-242.

Batra, R., Lenk, P., & Wedel, M. (2010). Brand extension strategy planning: Empirical estimation of brand-category personality fit and atypicality. *Journal of Marketing Research, 47*(2), 335-347.

Baxter, L. A., Henauw, C., Huisman, D., Livesay, C. B., Norwood, K., Hua, S., & ... Young, B. (2009). Lay Conceptions of "Family": A Replication and Extension. *Journal Of Family Communication*, *9*(3), 170-189.

Beatty, M. (1988). Situational and predispositional correlates of public speaking anxiety. *Communication Education, 37*(1), 28-39.

Behnke, R. R., & Sawyer, C. R. (1998). Conceptualizing speech anxiety as a dynamic trait. *Southern Communication Journal 63*(2), 160-168.

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24*, 145–152.

Bell, R. Q. (1954). An experimental test of the accelerated longitudinal approach. *Child Development, 25*, 281–286.

Brown, T.A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

Brumfiel, G. (2011, September 23). Particles break light-speed limit. *Nature*. Retrieved from http://www.nature.com/news/2011/110922/full/news.2011.554.html

Burgoon, J. (1976). Unwillingness-to-communicate scale - development and validation. *Communication Monographs, 43*(1), 60-69.

Burgoon, J.K., & Le Poire, B.A. (1999). Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs 66*, 105-124.

Buss, A.H. & Perry, M. (1992). The aggression questionnaire. *Journal of personality and social psychology 63*(3), 452-459.

Butler Ellis, J., & Smith, S. W. (2004). Memorable messages as guides to self-assessment of behavior: a replication and extension diary study. *Communication Monographs*, *71*(1), 97-119.

Cheney, G. (1983). On the various and changing meanings of organizational membership: A field study of organizational identification. *Communication Monographs, 50,* 342-362.

Chipperfield, J. O., & Steel, D. G. (2009). Design and estimation for split questionnaire surveys. *Journal of Official Statistics, 25*(2), 227-244.

Cho, A. (2011, September 22). Neutrinos Travel Faster Than Light, According to One Experiment. *Science*. Retrieved from http://news.sciencemag.org/sciencenow/2011/09/neutrinos-travel-faster-than-lig.html

Church, M.A., Truss, C.V., & Velicer, W.F. (1980). Structure of the Janis-Field feelings of inadequacy scale. *Perceptual and Motor Skills 50*(3) 935-9.

Cohen, J. (2008). What I watch and who I am: National pride and the viewing of local and foreign television in Israel. *Journal of Communication 58*, 149-167.

Clark, L. A., & Watson, D. (1995). Constructing Validity. *Psychological Assessment*, *7*(3), 309-319.

Covert, J. J., & Dixon, T. L. (2008). A changing view. *Communication Research, 35*(2), 232-256. doi: 10.1177/0093650207313166

Daly, J. (1978). Assessment of social-communicative anxiety via self-reports - comparison of measures. *Communication Monographs, 45*(3), 204-218.

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiple imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation*, *78*, 69–84.

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters, 15*(1), 21-36.

Dillman, D.A., Sinclair, M. D., & Clark, J.R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *The Public Opinion Quarterly, 57*(3), pp. 289-304.

Donovan, L. A., & MacIntyre, P. D. (2004). Age and sex differences in willingness to communicate, communication apprehension, and self-perceived competence. *Communication Research Reports 21*(4), 420-427.

Duff, D. C., Levine, T. R., Beatty, M. J., Woolbright, J., & Park, H. S. (2007). Testing public anxiety treatments against a credible placebo control. *Communication Education, 56*(1), 72-88.

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy, 35*(2), 333-363. doi:10.1016/S0005-7894(04)80042-X

Duncan, S., Duncan, T., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods, 1*(3), 236-248. doi:10.1037/1082-989X.1.3.236

Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2006). Alcohol use from ages 9 to 16: A cohort-sequential latent growth model. *Drug and Alcohol Dependence, 81*(1), 71-81. doi:10.1016/j.drugalcdep.2005.06.001

Fleming, J.S. & Watts, W.A. (1980). The dimensionality of self-esteem: Some results from a college sample. *Journal of Personality and Social Psychology 39*(5), 921-929.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods 5*(2), 155-174.

Ellis, D. G., & Maoz, I. (2007). Online argument between Israeli Jews and Palestinians. *Human Communication Research, 33*(3), 291-309. doi: 10.1111/j.1468-2958.2007.00300.x

Ellis, K. (1995). Apprehension, self-perceived competency, and teacher immediacy in the laboratory-supported public-speaking course: Trends and relationships. *Communication Education 44*, 64-78.

Enders, C. K. (2010). *Applied missing data analysis.* New York: Guilford.

Eveland, W. P., & Hively, M. H. (2009). Political discussion frequency, network size, and ‘‘heterogeneity’’ of discussion as predictors of political knowledge and participation. *Journal of Communication 59*, 205-224.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Floyd, K., Boren, J. P., Hannawa, A. F., Hesse, C., McEwan, B., & Veksler, A. E. (2009). Kissing in marital and cohabitating relationships: Effects on blood lipids, stress, and relational satisfaction. *Western Journal of Communication 73*(2), 113-133.

Floyd, K., Mikkelson, A. C., Hesse, C. & Pauley, P. M. (2007), Affectionate Writing Reduces Total Cholesterol: Two Randomized, Controlled Trials. Human *Communication Research*, 33: 119–142.

Frymier, A. (1994). A model of immediacy in the classroom. Communication Quarterly, 42, 133-144.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349-360. doi: 10.1093/poq/nfp031

Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. Survey Research Methods, 2, 21-32.

Geiser, C. (2012). *Data Analysis with Mplus*. New York: Guilford.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). Bayesian Data Analysis. Second Edition. Boca Raton: Chapman & Hall.

Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457-511.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31(2),* 197–218.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in the analysis of change. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 335–353). Washington, DC: American Psychological Association.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science 8*, 206−213.

Graham, J. W., Taylor, B. J, Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods 11*(4), 323-343.

Grimm, K. J., & Ram, N. (2011). Growth curve modeling from an SEM perspective. In T. Little, B. Laursen, & N. Card (Eds.), *Handbook of developmental research methods* (pp. 411-431). New York: Guilford Publications.

Hancock, G. R., & Mueller, R. 0. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), Structural Equation Modeling: Present and Future — A Festschrift in honor of Karl Jöreskog. Lincolnwood, IL: Scientific Software International, Inc.

Harel, O., Zimmerman, R., & Dekhtyar, O. (2008). Approaches to the handling of missing data in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The SAGE sourcebook of advanced data analysis methods for communication research* (pp. 349–371). Los Angeles, CA: Sage Publications.

Hargie, O., Dickson, D., Mallett, J., & Stringer, M. (2008). Communicating social identity. *Communication Research, 35*(6), 792-821. doi: 10.1177/0093650208324270

Hayduk, L. (2013, March 26). Re: Listwise/pairwise (was: Some substance (previously: Re: Recent "good" texts.) [Online forum comment]. Retrieved from https://listserv.ua.edu/cgi-bin/wa?A2=SEMNET;184eef34.1303

Hecht, M., Graham, J.W., & Elek, E. (2006). The drug resistant strategies intervention: Program effects on substance use. *Health Communication 20*(3), 267-276.

Hendrick, C. (1991). Replications, Strict Replications, and Conceptual Replications: Are They Important?. In J. W. Neuliep (Ed.), *Replication in the social sciences* (pp. 41-49). Newbury Park, CA: Sage.

Henningsen, D. D., & Henningsen, M. L. M. (2007). Do groups know what they don't know? dealing with missing information in decision-making groups. *Communication Research, 34*(5), 507-525. doi: 10.1177/0093650207305594

Hensley, W. E., & Batty, P. (1974). The measurement of communication anxiety among students in public speaking courses. *Indiana Speech Journal 8*, 7-10.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly, 45*(4), 549.

von Hippel, P. T. (2013). Should a normal imputation model be modified to imputed skewed variables? *Sociological Methods and Research 42*(1), 105-138.

Hodis, G. A., Bardhan, N. R., & Hodis, F. A. (2010). Patterns of change in willingness to communicate in public speaking contexts: A latent growth modeling analysis. *Journal of Applied Communication Research 38*(3), 248-267.

Horn, J. L. & McArdle, J. J. (1992). A practical guide to measurement invariance in aging research. *Experimental Aging Research,18(*3), 117-144.

Hullman, G. A., & Daily, M. (2008). Evaluating Physician Communication Competence Scales: A Replication and Extension. *Communication Research Reports*, *25*(4), 316-322.

Infante, D. A., & Wigley, C. J. (1986). Verbal aggressiveness: An interpersonal model and measure. *Communication Monographs 53*, 61-69.

Ishii, K. (2006). Implications of mobility: The uses of personal communication media in everyday life. *Journal of Communication 56*, 346-365.

Jöreskog, K.G. & Sörbom, D. (2006). LISREL 8.8 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Kam, J. A. (2011). The Effects of Language Brokering Frequency and Feelings on Mexican-Heritage Youth's Mental Health and Risky Behaviors. *Journal of Communication 61*, 455-475.

Kelly, L., & Keaten, J. A. (2007). Development of the affect for communication channels scale. *Journal of Communication 57*, 349-365.

Kerkhof, P., Finkenauer, C., & Muusses, L. D. (2011). Relational consequences of compulsive internet use: A longitudinal study among newlyweds. *Human Communication Research 37*, 147-173.

King, G., Honaker, J., Joseph, A., & Scheve, K. (1998). Listwise deletion is evil: What to do about missing data in political science (Unpublished manuscript). St. Louis, MO: Washington University in St. Louis. Retrieved from http://polmeth.wustl.edu/media/Paper/king98e.pdf

King, P., & Witt, P. (2009). Teacher immediacy, confidence, testing, and the measurement, of cognitive learning. *Communication Education 58*(1), 110-123.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Knobloch, L.K., & Theiss, J.A. (2011). Relational uncertainty and relationship talk within courtship: A longitudinal actor-partner interdependence model. *Communication Monographs 78*(1), 3-26.

Kraut, A. I., Wolfson, A. D., & Rothenberg, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology, 60*(6), 774-776. doi:10.1037/0021-9010.60.6.774

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

Lamal, P. A. (1991). On the importance of replication. In J. W. Neuliep (Ed.), *Replication in the social sciences* (pp. 31-36). Newbury Park, CA: Sage.

Ledbetter, A. M. (2009). Family communication patterns and relational maintenance behavior: Direct and mediated associations with friendship closeness. *Human Communication Research, 35*(1), 130-147. doi: 10.1111/j.1468-2958.2008.01341.x

Ledbetter, A. M., & Kuznekoff, J. H. (2012). More than a game. *Communication Research, 39*(2), 269-290. doi: 10.1177/0093650210397042

Ledbetter, A. M., Mazer, J. P., DeGroot, J. M., Meyer, K. R., Yuping Mao, & Swafford, B. (2011). Attitudes toward online social connection and self-disclosure as predictors of facebook communication and relational closeness. *Communication Research, 38*(1), 27-53. doi: 10.1177/0093650210365537

Lee, H. E., & Park, H. S. (2011). Why Koreans are more likely to favor "apology," while Americans are more likely to favor "thank you" *Human Communication Research, 37*(1), 125-146. doi: 10.1111/j.1468-2958.2010.01396.x

Levine, T.R., Kotowski, M.R., Beatty, M.J., & Van Kelegom, M.J. (2012). A meta-analysis of trait-behavior correlations in argumentativeness and verbal aggression. *Journal of Language and Social Psychology 3(*1) 95-111.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association 83*(404), 1198-1202.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guilford press.

Little, T. D. (2011, June). *Foundations of Longitudinal Modeling* [PowerPoint Slides]. Presented at The University of Kansas Summer Institute in Statistics, Lawrence, KS.

Little, T. D., Bovaird, J. A., & Card, N. A. (Eds.) (2007). *Modeling contextual effects in longitudinal studies.* Mahwah, NJ: LEA

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling 9*(2), 151-173.

Little, T.D., Preacher, K. J., Selig, J. P., & Card, N. A., (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development 31(*4), 357-365).

Liu, J. S. (1994) Fraction of missing information and convergence rate of data augmentation. In *Computationally Intensive Statistical Methods: Proceedings of the 26th Symposium Interface* (J. Sall and A. Lehmann, eds), 490–497. Interface Foundation of North American, Fairfax Station, VA.

Lohr, J.M., Rea, R. G., Porter, B. B., & Hamberger, L.K. (1980). Communication fear: A correlational study of trait generality. *Human Communication Research 6(*3) 280-284.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods 1*(2), 130-149.

Maloney, P., Grawitch, M. J., & Barber, L. K. (2011) Strategic item selection to reduce survey length: Reduction in validity? *Consulting Psychology Journal: Practice and Research 63*(3), 162-175.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the spielberger state?trait anxiety inventory (STAI). *British Journal of Clinical Psychology, 31*(3), 301-306. doi: 10.1111/j.2044-8260.1992.tb00997.x

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication methods and measures 2(*4), 260-293.

McComas, K. A., Trumbo, C. W., & Besley, J. C. (2007). Public meetings about suspected cancer clusters: The impact of voice, interactional justice, and risk perception on attendees' attitudes in six communities. *Journal of Health Communication 12*, 527-549.

McComas, K. A., Zheng, Y., Gay, G. K., Leonard, J. P., Dannenberg, A. J., & Dillon, H. (2010). Individuals' willingness to talk to their doctors about clinical trial enrollment. *Journal of Health Communication 15*, 189-204.

McCroskey, J. C. (1978). Validation of the PRCA as an index of oral communication apprehension. *Communication Monographs 45(*3), 192-203.

McCroskey, J.C. (1982). *An Introduction to rhetorical communication (4th Ed)*. Englewood Cliffs, NJ: Prentice Hall.

McCroskey, J.C. (2007). Personal report of communication apprehension. Retreived from http://www.jamescmccroskey.com/measures/prca24.htm

Meerkerk, G. J., Van den Eijnden, R. J. J. M., Vermulst, A. A., & Garretsen, H. F. L. (2009). Internetverslaving. In R. Kerssemakers, R. Van Meerten, E. Noorlander,

& H. Vervaeke (Eds.), *Drugs en alcohol. Gebruik, misbruik en verslaving* (pp. 456–470). Houten, The Netherlands: Bohn Stafleu van Loghum.

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

Miller, M.D. (1987). The relationship of communication reticence and negative expectations. *Communication Education, 36,* 228-235.

Miller, S. M., Fleisher, L., Roussi, P., Buzaglo, J. S., Schnoll, R., Slater, E., Rasor, S., & Popa-Mabe, M. (2005). Facilitating informed decision making about breast cancer risk and genetic counseling among women calling the NCI's cancer information service. *Journal of Health Communication, 10*, 119-136. doi:10.1080/07366290500265335

Mitchell, K., & Beard, F. (2010). Measuring Internet Dependence Among College Students: A Replication and Confirmatory Analysis. *Southwestern Mass Communication Journal*, *25*(2), 15-28.

Miyazaki, Y., & Raudenbush, S.W. (2000). Tests for linkages of multiple cohorts in an accelerated longitudinal design. *Psychological Methods 5*(1), 44-63.

Moerbeek, M. (2011). The effects of the number of cohorts, degree of overlap among cohorts, and frequency of observation on power in accelerated longitudinal designs. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 7*(1), 11-24. doi:10.1027/1614-2241/a000019

Montgomery, J. M., & Cutler, J. (2013) Computerized adaptive testing for public opinion surveys. *Political Analysis 21*, 172-192.

Moore, J. C., Stinson, L. L., & Welniak Jr., E. J. (1999). Income measurement error in surveys: A review. *Journal of Official Statistics 16*(4), 331-361.

Morreale, S. (1990). *The competent speaker.* Speech Communication Association Conference on Assessment of Communication Competency, Denver, CO.

Mueller, R. O., & Hancock, G. R**.** (2010). Structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 371-383) New York: Routledge.

Myers, T. (2011): Goodbye, Listwise Deletion: Presenting Hot Deck Imputation as an Easy and Effective Tool for Handling Missing Data*. Communication Methods and Measures 5*(4), 297-310.

Nussbaum, J. F. (2007). Life span communication and quality of life. *Journal of Communication, 57*(1), 1-7. doi: 10.1111/j.1460-2466.2006.00325.x

Ohio University Office of Institutional Research (2012, July). Ohio University Factbook. Retrieved from: http://www.ohio.edu/instres/factbook.pdf

O'Keefe, D. J., & Sypher, H. E. (1981). Cognitive complexity measures and the relationship of cognitive complexity to communication. *Human Communication Research 8*(1), 72-92.

Paek, H. (2008a). Mechanisms through which adolescents attend and respond to antismoking media campaigns. *Journal of Communication, 58*(1), 84-105. doi: 10.1111/j.1460-2466.2007.00375.x

Paek, H. (2008b). Moderating roles of primary social influences in the relationship between adolescent self-reported exposure to antismoking messages and smoking

intention. *Health Communication, 23*(6), 526-537. doi: 10.1080/10410230802460259

Paek, H. (2009). Differential effects of different peers: Further evidence of the peer proximity thesis in perceived peer influence on college students' smoking. *Journal of Communication, 59*(3), 434-455. doi: 10.1111/j.1460-2466.2009.01423.x

Paek, H., & Gunther, A. C. (2007). How peer proximity moderates indirect media influence on adolescent smoking. *Communication Research, 34*(4), 407-432. doi: 10.1177/0093650207302785

Pearson, J. C., Child, J. T., Carmon, A. F., & Miller, A. N. (2009). The influence of intimacy rituals and biological sex on relational quality and intimacy among dating couples. *Communication Research Reports 26*(4), 297-310.

Peter, J., & Valkenburg, P.M. (2009). Adolescents' exposure to sexually explicit internet materials and sexual satisfaction: A longitudinal study. *Human Communication Research 35*, 171-194.

Peugh, J. L., & Enders, C. K. (2004). Missing data in education research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74* (4), 525–556.

Porter, S. R. (2004). Multiple surveys of students and survey fatigue In Porter, S. R. (ed.), *Overcoming Survey Research Problems* (63-74). San Francisco: Jossey-Bass.

Price, V., Nir, L., & Cappella, J. N. (2006). Normative and informational influences in online political discussions. *Communication Theory 16*, 47-74.

Prinzie, P., Onghena, P., & Hellinckx, W. (2006). A cohort-sequential multivariate latent growth curve analysis of normative CBCL aggressive and delinquent problem

behavior: Associations with harsh discipline and gender. *International journal of business development 30*(5), 444-459.

Qi, W., Fink, E. L., & Cai, D. A. (2008). Loneliness, gender, and parasocial interaction: A uses and gratifications approach. *Communication Quarterly 56*(1), 87-109.

Raghunathan, T., & Grizzle, J. (1995). A split questionnaire survey design. *Journal of the American Statistical Association, 90,* 54–63.

Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, 13, 425-438.

Richmond, V. P., & McCroskey, .J. C. (1989). *Communication: Apprehension, avoidance, and effectiveness.* Scottsdale, AZ: Gorsuch Scarisbrick.

Rigdon, E. E. (1994). Calculating degrees of freedom for a structural equation model. *Structural Equation Modeling 1*(3), 274-278.

Roberto, A. J., & Goodall, C. E. (2009). Using the extended parallel process model to explain physicians' decisions to test their patients for kidney disease. *Journal of Health Communication 14*, 400-412.

Rosenthal, R. (1991). Replication in behavioral research. In J. W. Neuliep (Ed.), Replication in the social sciences (pp. 1-30). Newbury Park, CA: Sage.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika, 63*(2), 581-592.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics 12*(4), 1151-1172.

Rubin, R. B. (1982a). Assessing speaking and listening competence at the college level: The Communication Competency Assessment Instrument. *Communication Education, 31,* 19-32.

Rubin, R. B. (1982b). *Communication Competency Assessment Instrument.* Annandale, VA: Speech Communication Association.

Rubin, R. B. (1985). The validity of the communication competency assessment instrument. *Communication Monographs 52*, 173-185.

Rubin, R. B., Graham, E. E., & Mignerey, J. T. (1990). A longitudinal study of college students' communication competence. *Communication Education 39*, 1-14.

Rubin, R. B., Rubin, A. M., & Jordan, F. F. (1997). Effects of instruction on communication apprehension and communication competence. *Communication Education, 46*(2), 104-114.

Sallinen-Kuparinen, A., McCroskey, J. C., & Richmond, V. R. (1991). Willingness to communicate, communication apprehension, introversion, and self-reported communication competence: Finnish and American comparisons. *Communication Research Reports* 8, 55-64.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods 7*(2), 147-177.

Schaie, K.W. (1986). Beyond calendar definitions of age, time, and cohort: The general developmental model revisited. *Developmental Review 6*, 252-277.

Schrodt, P., & Ledbetter, A. M. (2007). Communication processes that mediate family communication patterns and mental well-being: A mean and covariance structures analysis of young adults from divorced and nondivorced families. *Human Communication Research, 33*(3), 330-356. doi: 10.1111/j.1468-2958.2007.00302.x

Scott, C., & Rockwell, S. (1997). The effect of communication, writing, and technology apprehension on likelihood to use new communication technologies. *Communication Education, 46*(1), 44-62.

Sinharay, S., Stern, H. S, & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods 6*(4), 317-329.

Smits, N., & Vorst, H. C. M. (2007). Reducing the length of questionnaires through structurally incomplete designs: An illustration. *Learning and Individual Differences, 17*(1), 25-34. doi:10.1016/j.lindif.2006.12.005

Slater, M. D. (2007). Reinforcing spirals: The Mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory 17*(3), 281-303.

Strahan, R., & Gerbasi, K. C. (1972). Short homogenous version of the Marlowe Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28(2), 191–193.

The National Youth Anti-Drug Media Campaign: Hearing Before the Subcommittee on Criminal Justice, Dug Policy, and Human Resources of the Committee on Government Reform of the House of Representatives. 106th Cong. 1 (1999) (Testimony of S. Shayam Sundar).

Valkenburg, P. M., & Peter, J. (2009). The effects of instant messaging on the quality of adolescents? existing friendships: A longitudinal study. *Journal of Communication, 59*(1), 79-97. doi: 10.1111/j.1460-2466.2008.01405.x

Valentino, N. A., Traugott, M. W., & Hutchings, V. L. (2002). Group Cues and Ideological Constraint: A Replication of Political Advertising Effects Studies in the Lab and in the Field. *Political Communication*, *19*(1), 29-48.

Velicer, W.F., Govia, J.M., Cherico, N.P., & Corriveau, D.P. (1985). Item format and the structure of the Buss-Durkee Hostility Inventory. *Aggressive Behavior 11*, 65-82.

Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2008). *Mathematical Statistics with Applications* (7th ed.). Belmont, CA: Thompson Higher Education.

Wardenaar, K. J., van Veen, T., Giltay, E. J., de Beurs, E., Penninx, B. W. J. H., & Zitman, F. G. (2010). Development and validation of a 30-item short adaptation of the mood and anxiety symptoms questionnaire (MASQ). *Psychiatry Research, 179*(1), 101-106. doi:10.1016/j.psychres.2009.03.005

Warren, J. R., Hecht, M. L., Jung, E., Kvasny, L., & Henderson, M. G. (2010). African American ethnic and class-based identities on the world wide web: Moderating the effects of self-perceived information Seeking/Finding and web self-efficacy. *Communication Research, 37*(5), 674-702. doi: 10.1177/0093650210374005

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Williams, D., Caplan, S., & Xiong, L. (2007). Can you hear me now? the impact of voice in an online gaming community. *Human Communication Research 33*(4), 427-449. doi: 10.1111/j.1468-2958.2007.00306.x

Witt, P. L., & Behnke, R. R. (2006). Anticipatory speech anxiety as a function of public speaking assignment type. *Communication Education 55*(2), 167-177.

Wrench, J. S. (2005). Development and validity testing of the homonegativity short form. *Journal of Intercultural Communication Research 34*(3), 152-165.

Zhang, Y., Dixon, T. L., & Conrad, K. (2009). Rap music videos and African American women's body image: The moderating role of ethnic identity. *Journal of Communication, 59*(2), 262-278. doi: 10.1111/j.1460-2466.2009.01415.x

APPENDIX A – STUDY 1 - RECRUITMENT EMAILS

Recruitment Email for Participants at First Data Collection

Subject: Your COMS 1030 Research Participation Requirement - Study Assignment

Hi,

As a student enrolled in COMS 1030, you are required to participate in the COMS Research Participation System. You have been assigned to my study, which will require you to complete a survey at three separate time points. This is the first of those surveys. It will take you about 15 minutes to complete the survey, and **it is really important that you complete it at some point before you present your Informative speech**.

Here is the link to take the survey:

https://ohioscripps.us2.qualtrics.com/SE/?SID=SV_8ixh1QnIbYacL3f

I will email you again before each of your upcoming speeches in COMS 1030 asking you to complete this survey two additional times.

Thank you for your participation in my research, and best of luck on your COMS 1030 presentations.

-Mike Parsons

Recruitment Email for Participants at Second Data Collection

Subject: Your COMS 1030 Research Participation Requirement – Second Survey

Hi,

As a student enrolled in COMS 1030, you are required to participate in the COMS Research Participation System. You have been assigned to my study, which will require you to complete a survey at three separate time points. This is the Second of those surveys.

It will take you about 15 minutes to complete the survey, and **it is really important that you complete it at some point before you present your Persuasive speech**.

Here is the link to take the survey:

https://ohioscripps.us2.qualtrics.com/SE/?SID=SV_5t1PWqQT2XTPNIx

I will email you again before your upcoming speech in COMS 1030 asking you to complete this survey one final time.

Thank you for your participation in my research, and best of luck on your COMS 1030 presentations.

-Mike Parsons

Recruitment Email for Participants at Third Data Collection

Subject: Your COMS 1030 Research Participation Requirement – Final Survey

Hi,

As a student enrolled in COMS 1030, you are required to participate in the COMS Research Participation System. You have been assigned to my study, which will require you to complete a survey at three separate time points. You have already completed it twice, and I am emailing you to request that you complete my survey a third and final time.

It will take you about 15 minutes to complete the survey, and **it is really important that you complete it at some point before you present your Persuasive speech**.

Here is the link to take the survey:

https://ohioscripps.us2.qualtrics.com/SE/?SID=SV_5t1PWqQT2XTPNIx

Again, **completion of this survey will complete your research participation requirement for COMS 1030**.

Thank you for your participation in my research, and best of luck on your COMS 1030 presentations.

-Mike Parsons

APPENDIX B: STUDY 1 - CONSENT FORM

Ohio University Consent Form

Title of Research: COMS 1030 Assessment

Researcher: Michael Parsons, School of Communication Studies

You are being asked to participate in research. For you to be able to decide whether you want to participate in this project, you should understand what the project is about, as well as the possible risks and benefits in order to make an informed decision. This process is known as informed consent. This form describes the purpose, procedures, possible benefits, and risks. It also explains how your personal information will be used and protected. Once you have read this form and your questions about the study are answered, you will be asked to participate in this study. You should receive a copy of this document to take with you.

Explanation of Study

The purpose of this study is to better understand how communication apprehension changes over the course of a semester-long public speaking class. This will be measured by a series of short survey instruments which will ask you about how you experience public speaking anxiety and apprehension, communicate with others with regards to public speaking events.

If you agree to participate, you will be asked to complete a short online survey 2-5 times during the course of Fall semester 2012.

You may discontinue participation in this study at any time, and there are no negative consequences for discontinuing participation.

Risks and Discomforts

No risks or discomforts are anticipated.

Benefits

There are no direct benefits involved with your participation in the study.

Confidentiality and Records

All surveys will be confidential, which means that identifying information (in this study, email addresses) will be seen only by the researcher. Additionally, at the end of data collection, all identifiable information will be removed.

Additionally, while every effort will be made to keep your study-related information confidential, there may be circumstances where this information must be shared with:

* Federal agencies, for example the Office of Human Research Protections, whose responsibility is to protect human subjects in research;

* Representatives of Ohio University (OU), including the Institutional Review Board, a committee that oversees the research at OU;

Contact Information

If you have any questions regarding this study, please contact Michael Parsons, the primary investigator at (740) 249-9383 or mp164309@ohio.edu. You may also contact the faculty advisor, Dr. Amy Chadwick, at (740) 593-4821 or Chadwick@ohio.edu.

If you have any questions regarding your rights as a research participant, please contact Jo Ellen Sherow, Director of Research Compliance, Ohio University, (740) 593-0664.

By agreeing to participate in this study, you are agreeing that:

- You have read this consent form (or it has been read to you) and have been given the opportunity to ask questions and have them answered
- You have been informed of potential risks and they have been explained to your satisfaction.
- You understand Ohio University has no funds set aside for any injuries you might receive as a result of participating in this research protocol
- You are 18 years of age or older
- Your participation in this research is completely voluntarily
- You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you and you will not lose any benefits to which you may otherwise be entitled.

APPENDIX C: STUDY 1 – MEASURES

Personal Report of Public Speaking Anxiety

(McCroskey, 1970)

Directions: Below are 34 statements that people sometimes make about themselves. Please indicate whether or not you believe each statement applies to you by marking whether you:

Strongly Disagree = 1; Disagree = 2; Neutral = 3; Agree = 4; Strongly Agree = 5.

1. While preparing for giving a speech, I feel tense and nervous.

2. I feel tense when I see the words "speech" and "public speech" on a course outline when studying.

3. My thoughts become confused and jumbled when I am giving a speech.

4. Right after giving a speech I feel that I have had a pleasant experience.

5. I get anxious when I think about a speech coming up.

6. I have no fear of giving a speech.

7. Although I am nervous just before starting a speech, I soon settle down after starting and feel calm and comfortable.

8. I look forward to giving a speech.

9. When the instructor announces a speaking assignment in class, I can feel myself getting tense.

10. My hands tremble when I am giving a speech.

11. I feel relaxed while giving a speech.

12. I enjoy preparing for a speech.

13. I am in constant fear of forgetting what I prepared to say.

14. I get anxious if someone asks me something about my topic that I don't know.

15. I face the prospect of giving a speech with confidence.

16. I feel that I am in complete possession of myself while giving a speech.

17. My mind is clear when giving a speech.

18. I do not dread giving a speech.

19. I perspire just before starting a speech.

20. My heart beats very fast just as I start a speech.

21. I experience considerable anxiety while sitting in the room just before my speech starts.

22. Certain parts of my body feel very tense and rigid while giving a speech.

23. Realizing that only a little time remains in a speech makes me very tense and anxious.

24. While giving a speech, I know I can control my feelings of tension and stress.

25. I breathe faster just before starting a speech.

26. I feel comfortable and relaxed in the hour or so just before giving a speech.

27. I do poorer on speeches because I am anxious.

28. I feel anxious when the teacher announces the date of a speaking assignment.

29. When I make a mistake while giving a speech, I find it hard to concentrate on the parts that follow.

30. During an important speech I experience a feeling of helplessness building up inside me.

31. I have trouble falling asleep the night before a speech.

32.My heart beats very fast while I present a speech.

33. I feel anxious while waiting to give my speech.

34. While giving a speech, I get so nervous I forget facts I really know.

Self-Perceived Public Speaking Competency Scale

(Ellis, 1995)

Directions: Please indicate the frequency with which you engage in each of the following behaviors, from 0=Never through 4=Very often.

1. I choose a topic that is appropriate for the audience.

2. I have excellent posture when giving a speech.

3. I have difficulty using appropriate gestures.

4. Generally, I move smoothly from idea to idea within my speech.

5. I choose a topic that is appropriate for the occasion.

6. Generally, giving an effective introduction is a problem for me.

7. I use appropriate facial expressions.

8. Generally, the body of my speech is logically organized.

9. I use a variety of supporting material (e.g., examples, expert opinions, statistics, research findings, illustrations) to enhance my speech.

l 0. I use variety in pitch to enhance my message.

11. Maintaining eye contact is a problem for me.

12. Generally, my conclusion clearly reflects the content of my speech.

13. I use language that is extremely clear.

14. Some audience members have difficulty hearing me

15. I use variety in my rate of speech.

16. I have trouble articulating my words clearly

17. I dress to enhance my credibility.

18. Using high quality supporting material is often problematic for me.

19. I make very few, if any, pronunciation errors.

APPENDIX D: STUDY 2 – CONSENT FORM

Ohio University Consent Form

Title of Research: Scale Validation

Researcher: Michael Parsons, School of Communication Studies

You are being asked to participate in research. For you to be able to decide whether you want to participate in this project, you should understand what the project is about, as well as the possible risks and benefits in order to make an informed decision. This process is known as informed consent. This form describes the purpose, procedures, possible benefits, and risks. It also explains how your personal information will be used and protected. Once you have read this form and your questions about the study are answered, you will be asked to participate in this study. You should receive a copy of this document to take with you.

Explanation of Study

The purpose of this study is to better understand how verbal aggressiveness relates to other communicative concepts such as apprehension, guilt, and argumentativeness. This will be measured by a short survey instrument which will ask you about how you experience argumentativeness, communication apprehension, guilt, and verbal aggressiveness.

If you agree to participate, you will be asked to complete a short online survey.

You may discontinue participation in this study at any time, and there are no negative consequences for discontinuing participation.

Risks and Discomforts

No risks or discomforts are anticipated.

Benefits

There are no direct benefits involved with your participation in the study.

Confidentiality and Records

All surveys will be anonymous, which means that no identifiable information will be collected.

Additionally, while every effort will be made to keep your study-related information confidential, there may be circumstances where this information must be shared with:

* Federal agencies, for example the Office of Human Research Protections, whose responsibility is to protect human subjects in research;

* Representatives of Ohio University (OU), including the Institutional Review Board, a committee that oversees the research at OU;

Contact Information

If you have any questions regarding this study, please contact Michael Parsons, the primary investigator at (740) 249-9383 or mp164309@ohio.edu. You may also contact the faculty advisor, Dr. Amy Chadwick, at (740) 593-4821 or Chadwick@ohio.edu.

If you have any questions regarding your rights as a research participant, please contact Jo Ellen Sherow, Director of Research Compliance, Ohio University, (740) 593-0664.

By agreeing to participate in this study, you are agreeing that:

- You have read this consent form (or it has been read to you) and have been given the opportunity to ask questions and have them answered
- You have been informed of potential risks and they have been explained to your satisfaction.
- You understand Ohio University has no funds set aside for any injuries you might receive as a result of participating in this research protocol
- You are 18 years of age or older
- Your participation in this research is completely voluntarily
- You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you and you will not lose any benefits to which you may otherwise be entitled.

Version Date: *8/30/12*

APPENDIX E: STUDY 2 - MEASURES

Verbal Aggressiveness Scale

(Infante & Wigley, 1986)

This survey is concerned with how we try to get people to comply with our wishes. Indicate how often each statement is true for you personally when you try to influence other persons. Use the following scale:

1- almost never true 2 - rarely true 3 - occasionally true

4 - often true 5- almost always true

1. I am extremely careful to avoid attacking individuals' intelligence when I attack their ideas.

2. When individuals are very stubborn, I use insults to soften the stubbornness.

3. I try very hard to avoid having other people feel bad about themselves when I try to influence them.

4. When people refuse to do a task I know is important, without good reason, I tell them they are unreasonable.

5. When others do things I regard as stupid, I try to be extremely gentle with them.

6. If individuals I am trying to influence really deserve it, I attack their character.

7.When people behave in ways that are in very poor taste, I insult them in order to shock them into proper behavior.

8. I try to make people feel good about themselves even when their ideas are stupid.

9. When people simply will not budge on a matter of importance I lose my temper and say rather strong things to them.

10. When people criticize my shortcomings, I take it in good humor and do not try to get back at them.

11. When individuals insult me, I get a lot of pleasure out of really telling them off.

12. When I dislike individuals greatly, I try not to show it in what I say or how I say it.

13. I like poking fun at people who do things which are very stupid in order to stimulate their intelligence.

14. When I attack persons' ideas, I try not to damage their self-concepts.

15. When I try to influence people, I make a great effort not to offend them.

16. When people do things which are mean or cruel, I attack their character in order to help correct their behavior.

17. I refuse to participate in arguments when they involve personal attacks.

18. When nothing seems to work in trying to influence others, I yell and scream in order to get some movement from them.

19. When I am not able to refute others' positions, I try to make them feel defensive in order to weaken their positions.

20. When an argument shifts to personal attacks, I try very hard to change the subject.

Argumentativeness Scale

(Infante & Rancer, 1982)

This questionnaire contains questions about arguing controversial issues. Indicate how often each statement is true for you personally when you try to influence other persons. Use the following scale:

1- almost never true 2 - rarely true 3 - occasionally true

4 - often true 5- almost always true

1. While in an argument, I worry that the person I am arguing with will form a negative impression of me.

2.Arguing over controversial issues improves my intelligence..

3. I enjoy avoiding arguments.

4. l am energetic and enthusiastic when I argue.

5. Once I finish an argument I promise myself that I will not get into another.

6. Arguing with a person creates more problems for me than it solves.

7. I have a pleasant, good feeling when I win a point in an argument.

8. When I finish arguing with someone I feel nervous and upset.

9. I enjoy a good argument over a controversial issue.

10. I get an unpleasant feeling when I realize I am about to get into an argument.

11. I enjoy defending my point of view on an issue.

12. I am happy when I keep an argument from happening.

13. I do not like to miss the opportunity to argue a controversial issue.

14. I prefer being with people who rarely disagree with me.

15. I consider an argument an exciting intellectual challenge.

16. I find myself unable to think of effective points during an argument.

17. I feel refreshed and satisfied after an argument on a controversial issue.

18. I have the ability to do well in an argument.

19. I try to avoid getting into arguments.

20. I feel excitement when I expect that a conversation I am in is leading to an argument.

Subscales of the Hostility-Guilt Inventory

(Buss & Durkee, 1957)

For each item, please indicate whether the item is true or false

(Assault Subscale)

1. Once in a while, I cannot control my urge to harm others

2. I can think of no good reason for ever hitting anyone

3. If somebody hits me first, I let him have it.

4. Whoever insults me or my family is asking for a fight.

5. People who continually pester you are asking for a punch in the nose.

6. I seldom strike back, even if someone hits me first.

7. When I really lose my temper, I am capable of slapping someone.

8. I get into fights about as often as the next person

9. If I have to resort to physical violence to defend my rights, I will.

10. I have known people who pushed me so far that we came to blows.

(Verbal Hostility Subscale)

11. When I disapprove of my friends' behavior, I let them know it.

12. I often find myself disagreeing with people.

13. I can't help getting into arguments when people disagree with me.

14. I demand that people respect my rights.

15. Even when my anger is aroused, I don't use strong language.

16. If somebody annoys me, I am apt to tell him what I think of him.

17.When people yell at me, I yell back.

18. When I get mad, I say nasty things.

19. I could not put someone in his place, even if he needed it.

20. I often make threats I don't really mean to carry out.

21.When arguing, I tend to raise my voice.

22. I generally cover up my poor opinion of others.

23. I would rather concede a point than get into an argument about it.

Social Desirability Index

(Crowne & Marlowe, 1960)

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is *true* or *false* as it pertains to you personally.

1. Before voting I thoroughly investigate the qualifications of all the candidates.

2. I never hesitate to go out of my way to help someone in trouble.

3. It is sometimes hard for me to go on with my work if I am not encouraged.

4. I have never intensely disliked anyone.

5. On occasion I have had doubts about my ability to succeed in life.

6. I sometimes feel resentful when I don't get my way.

7. I am always careful about my manner of dress.

8. My table manners at home are as good as when I eat out in a restaurant.

9. If I could get into a movie without paying and be sure I was not seen I would probably do it.

10. On a few occasions, I have given up doing something because I thought too little of my ability.

11. I like to gossip at times.

12. There have been times when I felt like rebelling against people in authority even though I knew they were right.

13. No matter who I'm talking to, I'm always a good listener.

14. I can remember "playing sick" to get out of something.

15. There have been occasions when I took advantage of someone.

16. I'm always willing to admit it when I make a mistake.

17. I always try to practice what I preach.

18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people.

19. I sometimes try to get even rather than forgive and forget.

20. When I don't know something I don't at all mind admitting it.

21. I am always courteous, even to people who are disagreeable.

22. At times I have really insisted on having things my own way.

23. There have been occasions when I felt like smashing things.

24. I would never think of letting someone else be punished for my wrong- doings.

25. I never resent being asked to return a favor.

26. I have never been irked when people expressed ideas very different from my own.

27. I never make a long trip without checking the safety of my car.

28. There have been times when I was quite jealous of the good fortune of others.

29. I have almost never felt the urge to tell someone off.

30. I am sometimes irritated by people who ask favors of me.

31. I have never felt that I was punished without cause.

32. I sometimes think when people have a misfortune they only got what they deserved.

33. I have never deliberately said something that hurt someone's feelings.

The Janis & Field Personality Questionnaire

Feelings of Inadequacy subscale.

(Hovland & Janis, 1959)

All questions beginning with the phrase "how often do you…? And "Do you ever…?" had the following checklist of five answer categories: Very often, Fairly often, sometimes, once in a great while, practically never.

Questions with the wording "[…]how_____ do you feel….?" Had the answer options: Very, fairly, slightly, not very, not at all.

1. How often do you feel inferior to most of the people you know?

2. Do you ever think that you are a worthless individual?

3. How confident do you feel that some day the people you know will look up to you and respect you?

4. How often do you feel to blame for your mistakes?

5. Do you ever feel so discouraged with yourself that you wonder whether anything is worthwhile?

6. How often do you feel that you dislike yourself?

7. In general, how confident do you feel about your abilities?

8. How often do you have the feeling that there is *nothing* you can do well?

9. How much do you worry about how well you get along with other people?

10.How often do you worry about criticisms that might be made of your work by whoever is responsible for checking up on your work?

11.Do you ever feel afraid or anxious when you are going into a room by yourself where other people have already gathered and are talking?

12.How often do you feel self-conscious?

13. When you have to talk in front of a class or a group of people your own age, how afraid or worried do you usually feel?

14. When you are trying to win in a game or sport and you know that other people are watching you, how rattled or flustered do you usually get?

15. How much do you worry about whether other people will regard you as a success or a failure in your job or career?

16. When in a group of people, do you have trouble thinking of the right things to talk about?

17. When you have made an embarrassing mistake or have done something that makes you look foolish, how long to you usually keep on worrying about it?

18. Do you find it hard to make talk when you meet new people?

19. How often do you worry about whether other people like to be with you?

20. How often are you troubled with shyness?

21. When you are trying to convince other people who disagree with your ideas, how worried do you usually feel about the impressions you are making?

22. When you think about the possibility that some of your friends or acquaintances might *not* have a good opinion of you, how concerned or worried do you feel about it?

23. How often do you feel worried or bothered about what other people think of you?

Personal Report of Communication Apprehension

(McCroskey, 1978)

This instrument is composed of twenty-four statements concerning feelings about communicating with others. Please indicate the degree to which each statement applies to you by marking whether you: **Strongly Disagree = 1; Disagree = 2; are Neutral = 3; Agree = 4; Strongly Agree = 5**

1. I dislike participating in group discussions.

2. Generally, I am comfortable while participating in group discussions.

3. I am tense and nervous while participating in group discussions.

4. I like to get involved in group discussions.

5. Engaging in a group discussion with new people makes me tense and nervous.

6. I am calm and relaxed while participating in group discussions.

7. Generally, I am nervous when I have to participate in a meeting.

8. Usually, I am comfortable when I have to participate in a meeting.

9. I am very calm and relaxed when I am called upon to express an opinion at a meeting.

10. I am afraid to express myself at meetings.

11. Communicating at meetings usually makes me uncomfortable.

12. I am very relaxed when answering questions at a meeting.

13. While participating in a conversation with a new acquaintance, I feel very nervous.

14. I have no fear of speaking up in conversations.

15. Ordinarily I am very tense and nervous in conversations.

16. Ordinarily I am very calm and relaxed in conversations.

17. While conversing with a new acquaintance, I feel very relaxed.

18. I'm afraid to speak up in conversations.

19. I have no fear of giving a speech.

20. Certain parts of my body feel very tense and rigid while giving a speech.

21. I feel relaxed while giving a speech.

22. My thoughts become confused and jumbled when I am giving a speech.

23. I face the prospect of giving a speech with confidence.

24. While giving a speech, I get so nervous I forget facts I really know.

APPENDIX F: STUDY 3 – CONSENT FORM

Ohio University Consent Form

Title of Research: Communication Competence

Researcher: Michael Parsons, School of Communication Studies

You are being asked to participate in research. For you to be able to decide whether you want to participate in this project, you should understand what the project is about, as well as the possible risks and benefits in order to make an informed decision. This process is known as informed consent. This form describes the purpose, procedures, possible benefits, and risks. It also explains how your personal information will be used and protected. Once you have read this form and your questions about the study are answered, you will be asked to participate in this study. You should receive a copy of this document to take with you.

Explanation of Study

The purpose of my study is to better understand how communication competence changes over the course of a college career. This will be measured by a series of short survey instruments which will ask you about how you communicate with others, and what kinds of experiences you had in high school that involved your communicative skills.

If you agree to participate, you will be asked to complete a short online survey. You will also be contacted and asked to complete the survey again at 3 later points in time, as follows:

1) June, 2012
2) September, 2012
3) January, 2013

You may discontinue participation in this study at any time, and there are no negative consequences for discontinuing participation.

Risks and Discomforts

No risks or discomforts are anticipated.

Benefits

There are no direct benefits involved with your participation in the study.

Confidentiality and Records

All surveys will be confidential, which means your identity will be unknown, even to the researcher. Additionally, all surveys will be anonymous, which means that no one will be able to connect you to your answers.

Additionally, while every effort will be made to keep your study-related information confidential, there may be circumstances where this information must be shared with:

* Federal agencies, for example the Office of Human Research Protections, whose responsibility is to protect human subjects in research;

* Representatives of Ohio University (OU), including the Institutional Review Board, a committee that oversees the research at OU;

Contact Information

If you have any questions regarding this study, please contact Michael Parsons, the primary investigator at (740) 249-9383 or mp164309@ohio.edu. You may also contact the faculty advisor, Dr. Scott Titsworth, at (740) 593-9160 or titswort@ohio.edu.

If you have any questions regarding your rights as a research participant, please contact Jo Ellen Sherow, Director of Research Compliance, Ohio University, (740) 593-0664.

By agreeing to participate in this study, you are agreeing that:

- You have read this consent form (or it has been read to you) and have been given the opportunity to ask questions and have them answered
- You have been informed of potential risks and they have been explained to your satisfaction.
- You understand Ohio University has no funds set aside for any injuries you might receive as a result of participating in this research protocol

- You are 18 years of age or older
- Your participation in this research is completely voluntarily
- You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you and you will not lose any benefits to which you may otherwise be entitled.

Version Date: *11/09/11*

APPENDIX G: STUDY 3 – RECRUITMENT MESSAGES

Initial Recruitment Tool

(Text Example Only)

Hello,

Are you a current college student? If so, I'd like you to take my survey.

I'm looking for research participants to help me better understand how college students' communication competence changes over the course of a college career. If you're willing, I'd like you to take a short survey that asks you a few questions about how you communicate with others. Since this is a longitudinal study, and I am interested in change over time, it is my hope that you will also be willing to take the same survey 3 more times over the course of the next year.

Surveys will be conducted online, via http://ohio-coms.sona-systems.com/, and will ask questions about how you communicate with others, how you feel about communicating, and what activities you participated in during high school that involved being a competent communicator. Surveys are completely anonymous, and the only requirement for participation is that you must be a current college student.

If you agree to participate, the online research system used for this study will contact you and ask you to take the survey 3 more times over the course of the next year.

For more information, you can email me at mp164309@ohio.edu, or call me at (740) 249-9383.

Thanks for your help!

Mike Parsons

School of Communication Studies

Ohio University

Pre-Time Point #3 Recruitment Tool.

Hello,

Are you a current first-year college student? If so, I'd like you to take my survey.

I'm looking for research participants to help me better understand how college students' communication competence changes over the course of a college career. If you're willing, I'd like you to take a short survey that asks you a few questions about how you communicate with others. Since this is a longitudinal study, and I am interested in change over time, it is my hope that you will also be willing to take the same survey one more time in January.

Surveys will be conducted online, via http://ohio-coms.sona-systems.com/, and will ask questions about how you communicate with others, how you feel about communicating, and what activities you participated in during high school that involved being a competent communicator. Surveys are completely anonymous, and the only requirement for participation is that you must be a current first-year college student.

If you agree to participate, the online research system used for this study will contact you and ask you to take the survey again in January.

For more information, you can email me at mp164309@ohio.edu, or call me at (740) 249-9383.

Thanks for your help!

Mike Parsons

School of Communication Studies

Ohio University

APPENDIX H: STUDY 3 – MEASURES

Interaction Involvement Scale

(Cegala, 1981)

For the following statements, please rate how much each statement sounds like it describes you. Please rate on a scale of 1-7, with 1 being "not at all like me" and 7 being "very much like me"

I. I am keenly aware of how others perceive me during my conversations.

2. My mind wanders during conversations and I often miss parts of what is going on.

3. Often in conversations I'm not sure what to say, I can't seem to find the appropriate lines.

4. I carefully observe how others respond to me during my conversations.

5. Often I will pretend to be listening to someone when in fact I'm thinking about something else.

6. Often in conversations I'm not sure what my role is; that is, I'm not sure how I'm expected to relate to others.

7. I listen carefully to others during a conversation.

8. Often I am preoccupied in my conversations and do not pay complete attention to the others.

9. Often in conversations I'm not sure what the other is *really* saying.
10. Often in conversations I am not sure what others' needs (e.g., reassurance, a compliment, etc.) are until it is too late to respond appropriately.
11. During conversations I am sensitive to others' subtle or hidden meanings.
12. I am very observant during my conversations with others.

13. In conversations I pay close attention to what others say and do and try to obtain as much information as I can.
14. Often I feel sort of "unplugged" from the social situation of which I am part; that is, I'm uncertain of my role, others' motives, and what's happening.
15. In my conversations I really know what's going on; that is, I have a "handle on the situation."
16. In my conversations I can accurately perceive others' intentions quite well.
17. Often in conversations I'm not sure how I'm expected to respond.
18. In conversations I am responsive to the meaning of others' behavior in relation to myself and the situation.

Personal Report of Communication Apprehension-24

(McCroskey, 1978)

For the following twenty-four statements, please indicate the degree to which each statement applies to you on a scale of 1-5, by marking whether you

(1) strongly agree, (2) agree, (3) are undecided, (4) disagree, or (5) strongly disagree.

Work quickly; record your first impression.

1. I dislike participating in group discussions.
2. Generally, I am comfortable while participating in group discussions.
3. I am tense and nervous while participating in group discussions.
4. I like to get involved in group discussions.
5. Engaging in a group discussion with new people makes me tense and nervous.

6. I am calm and relaxed while participating in group discussions.
7. Generally, I am nervous when I have to participate in a meeting.
8. Usually I am calm and relaxed while participating in meetings.
9. I am very calm and relaxed when I am called upon to express an opinion at a meeting.
10. I am afraid to express myself at meetings.
11. Communicating at meetings usually makes me uncomfortable.
12. I am very relaxed when answering questions at a meeting.
13. While participating in a conversation with a new acquaintance, I feel very nervous.
14. I have no fear of speaking up in conversations.
15. Ordinarily I am very tense and nervous in conversations.
16. Ordinarily I am very calm and relaxed in conversations.
17. While conversing with a new acquaintance, I feel very relaxed.
18. I'm afraid to speak up in conversations.
19. I have no fear of giving a speech.
20. Certain parts of my body feel very tense and rigid while giving a speech.
21. I feel relaxed while giving a speech.
22. My thoughts become confused and jumbled when I am giving a speech.
23. I face the prospect of giving a speech with confidence.
24. While giving a speech, I get so nervous I forget facts I really know.

Communication Competency Self-Report Questionnaire

(Rubin, 1985)

For the following items, please indicate how each statement reflects your own communication behavior on a scale of 1-5 by marking if it applies to you:

(1) ALWAYS (2) USUALLY (3) SOMETIMES (4) SELDOM (5) NEVER

1.I mispronounce a lot of words.

2.When speaking with someone, the words I use say one thing while my face and tone of voice say something different.

3.When giving a speech, I speak clearly and distinctly.

4.When giving a speech, I can be persuasive when I want to be.

5.When I speak with others, my ideas are clearly and concisely presented.

6.When giving a speech, I thoroughly express and fully defend my positions on issues.

7.I am unable to tell whether or not someone has understood what I have said.

8.I know when I'm hearing a fact and when I'm hearing someone's personal opinion.

9.When professors make suggestions in class on how I can improve, I understand the suggestions.

10.I understand the assignments that are given orally in class.

11.When I tell others about a class lecture I've heard, my version leaves out some important items.

12.When I have to introduce myself in a class, I am able to fully and concisely describe my interests and let others know who I am.

13.When speaking with others, I have to ask a question several times, in several ways, to get the information I want.

14.I have to answer a question several times before others seem satisfied with my answer.

15.I find it difficult to express my satisfaction or dissatisfaction about a course to the professor.

16.When I explain something to someone, it tends to be disorganized.

17.When I give directions to another person, the directions are accurate.

18.When I try to describe someone else's point of view, I have trouble getting it right.

19.I am able to give a balanced explanation of differing opinions.

Previous Experiences Scale

Please rate your experience during High School with the following activities on a scale 1-5, ranging from 1, "no experience, to 5, "extensive experience."

1.Selling

2.Working with people in office

3.Serving Others

4.Working as a receptionist

5.Advising/Counseling

6.Supervising/Managing/Organizing

7.Coaching

8.Directing

9.Producing

10.Performing/Acting/Singing

11.Lecturing/Speaking

12.TV or Radio broadcasting

13.Debating

14.Conducting Meetings

15.Helping Others

16.Ministering

APPENDIX I: SPSS RECODING SCRIPTS

Study 1 – Parcel and Augment Creation Script

*Creates Parcels**

```
COMPUTE P1a=(Q1+Q6+Q7+Q12+Q13+Q18+Q19+Q24+Q25+Q30+Q31)/11.
COMPUTE P1b=(Q2+Q5+Q8+Q11+Q14+Q17+Q20+Q23+Q26+Q29+Q32)/11.
COMPUTE P1c=(Q3+Q4+Q9+Q10+Q15+Q16+Q21+Q22+Q27+Q28+Q33)/11.
COMPUTE P2a=(QB1+QB6+QB7+QB12+QB13+QB18+QB19+QB24+QB25+QB30+QB31)/11.
COMPUTE P2b=(QB2+QB5+QB8+QB11+QB14+QB17+QB20+QB23+QB26+QB29+QB32)/11.
COMPUTE P2c=(QB3+QB4+QB9+QB10+QB15+QB16+QB21+QB22+QB27+QB28+QB33)/11.
COMPUTE P3a=(QC1+QC6+QC7+QC12+QC13+QC18+QC19+QC24+QC25+QC30+QC31)/11.
COMPUTE P3b=(QC2+QC5+QC8+QC11+QC14+QC17+QC20+QC23+QC26+QC29+QC32)/11.
COMPUTE P3c=(QC3+QC4+QC9+QC10+QC15+QC16+QC21+QC22+QC27+QC28+QC33)/11.
COMPUTE P4a=(QD1+QD6+QD7+QD12+QD13+QD18+QD19+QD24+QD25+QD30+QD31)/11.
COMPUTE P4b=(QD2+QD5+QD8+QD11+QD14+QD17+QD20+QD23+QD26+QD29+QD32)/11.
COMPUTE P4c=(QD3+QD4+QD9+QD10+QD15+QD16+QD21+QD22+QD27+QD28+QD33)/11.
COMPUTE P5a=(QE1+QE6+QE7+QE12+QE13+QE18+QE19+QE24+QE25+QE30+QE31)/11.
COMPUTE P5b=(QE2+QE5+QE8+QE11+QE14+QE17+QE20+QE23+QE26+QE29+QE32)/11.
COMPUTE P5c=(QE3+QE4+QE9+QE10+QE15+QE16+QE21+QE22+QE27+QE28+QE33)/11.

COMPUTE S1a=(Q34+Q39+Q40+Q45+Q46+Q51+Q52)/7.
COMPUTE S1b=(Q35+Q38+Q41+Q44+Q47+Q50)/6.
COMPUTE S1c=(Q36+Q37+Q42+Q43+Q48+Q49)/6.
COMPUTE S2a=(QB34+QB39+QB40+QB45+QB46+QB51+QB52)/7.
COMPUTE S2b=(QB35+QB38+QB41+QB44+QB47+QB50)/6.
COMPUTE S2c=(QB36+QB37+QB42+QB43+QB48+QB49)/6.
COMPUTE S3a=(QC34+QC39+QC40+QC45+QC46+QC51+QC52)/7.
COMPUTE S3b=(QC35+QC38+QC41+QC44+QC47+QC50)/6.
COMPUTE S3c=(QC36+QC37+QC42+QC43+QC48+QC49)/6.
COMPUTE S4a=(QD34+QD39+QD40+QD45+QD46+QD51+QD52)/7.
COMPUTE S4b=(QD35+QD38+QD41+QD44+QD47+QD50)/6.
COMPUTE S4c=(QD36+QD37+QD42+QD43+QD48+QD49)/6.
COMPUTE S5a=(QE34+QE39+QE40+QE45+QE46+QE51+QE52)/7.
COMPUTE S5b=(QE35+QE38+QE41+QE44+QE47+QE50)/6.
COMPUTE S5c=(QE36+QE37+QE42+QE43+QE48+QE49)/6.
EXECUTE.
```

Creates Augment scores for the parcels

```
COMPUTE P1augA=MEAN(Q1,Q6,Q7,Q12,Q13,Q18,Q19,Q24,Q25,Q30,Q31).
COMPUTE P1augB=MEAN(Q2,Q5,Q8,Q11,Q14,Q17,Q20,Q23,Q26,Q29,Q32).
COMPUTE P1augC=MEAN(Q3,Q4,Q9,Q10,Q15,Q16,Q21,Q22,Q27,Q28,Q33).
COMPUTE P2augA=MEAN(QB1,QB6,QB7,QB12,QB13,QB18,QB19,QB24,QB25,QB30,QB31).
COMPUTE P2augB=MEAN(QB2,QB5,QB8,QB11,QB14,QB17,QB20,QB23,QB26,QB29,QB32).
COMPUTE P2augC=MEAN(QB3,QB4,QB9,QB10,QB15,QB16,QB21,QB22,QB27,QB28,QB33).
COMPUTE P3augA=MEAN(QC1,QC6,QC7,QC12,QC13,QC18,QC19,QC24,QC25,QC30,QC31).
COMPUTE P3augB=MEAN(QC2,QC5,QC8,QC11,QC14,QC17,QC20,QC23,QC26,QC29,QC32).
COMPUTE P3augC=MEAN(QC3,QC4,QC9,QC10,QC15,QC16,QC21,QC22,QC27,QC28,QC33).
COMPUTE P4augA=MEAN(QD1,QD6,QD7,QD12,QD13,QD18,QD19,QD24,QD25,QD30,QD31).
COMPUTE P4augB=MEAN(QD2,QD5,QD8,QD11,QD14,QD17,QD20,QD23,QD26,QD29,QD32).
COMPUTE P4augC=MEAN(QD3,QD4,QD9,QD10,QD15,QD16,QD21,QD22,QD27,QD28,QD33).
COMPUTE P5augA=MEAN(QE1,QE6,QE7,QE12,QE13,QE18,QE19,QE24,QE25,QE30,QE31).
```

```
COMPUTE P5augB=MEAN(QE2,QE5,QE8,QE11,QE14,QE17,QE20,QE23,QE26,QE29,QE32).
COMPUTE P5augC=MEAN(QE3,QE4,QE9,QE10,QE15,QE16,QE21,QE22,QE27,QE28,QE33).

COMPUTE S1augA=MEAN(Q34,Q39,Q40,Q45,Q46,Q51,Q52).
COMPUTE S1augB=MEAN(Q35,Q38,Q41,Q44,Q47,Q50).
COMPUTE S1augC=MEAN(Q36,Q37,Q42,Q43,Q48,Q49).
COMPUTE S2augA=MEAN(QB34,QB39,QB40,QB45,QB46,QB51,QB52).
COMPUTE S2augB=MEAN(QB35,QB38,QB41,QB44,QB47,QB50).
COMPUTE S2augC=MEAN(QB36,QB37,QB42,QB43,QB48,QB49).
COMPUTE S3augA=MEAN(QC34,QC39,QC40,QC45,QC46,QC51,QC52).
COMPUTE S3augB=MEAN(QC35,QC38,QC41,QC44,QC47,QC50).
COMPUTE S3augC= MEAN(QC36,QC37,QC42,QC43,QC48,QC49).
COMPUTE S4augA=MEAN(QD34,QD39,QD40,QD45,QD46,QD51,QD52).
COMPUTE S4augB=MEAN(QD35,QD38,QD41,QD44,QD47,QD50).
COMPUTE S4augC=MEAN(QD36,QD37,QD42,QD43,QD48,QD49).
COMPUTE S5augA=MEAN(QE34,QE39,QE40,QE45,QE46,QE51,QE52).
COMPUTE S5augB=MEAN(QE35,QE38,QE41,QE44,QE47,QE50).
COMPUTE S5augC=MEAN(QE36,QE37,QE42,QE43,QE48,QE49).
EXECUTE.
```

Study 2 – Item Recoding and Scale Creation

```
*Recodes for the VA Scale.
RECODE V1 V3 V5 V8 V10 V12 V14 V15 V17 V20 (1=5) (5=1) (2=4) (4=2).

*recodes for the Assaut & Verbal hostility scales.
RECODE V42 V46 (1=2) (2=1).
EXECUTE.
RECODE V55 V59 V62 V63 (1=2) (2=1).
EXECUTE.

*recodes for the SDI.
RECODE V66 V68 V69 V72 V73 V74 V75 V77 V78 V82 V85 V86 V91 V93 V95 (1=2) (2=1).

*Recode for the FoI.
RECODE V99 V103 (1=5) (5=1) (2=4) (4=2).
EXECUTE.

*Recodes for the PRCA.
RECODE V121 V123 V125 V127 V128 V131 V133 V135 V136 V138 V140 V142 (1=5) (5=1) (2=4)
(4=2).
EXECUTE.


*Master recode puts ALL T/F items on a True is GREATER directionality.
RECODE V41 V42 V43 V44 V45 V46 V47 V48 V49 V50 V51 V52 V53 V54 V55 V56 V57 V58 V59
V60 V61 V62 V63 V64 V65 V66 V67 V68 V69 V70 V71 V72 V73 V74 V75 V76 V77 V78 V79 V80 V81
V82 V83 V84 V85 V86 V87 V88 V89 V90 V91 V92 V93 V94 V95 V96 (1=2) (2=1).

*The following commands create the composite scales - Assuming ALL recodes have been done!.
COMPUTE
V147=SUM(V1,V2,V3,V4,V5,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20).
```

```
EXECUTE.
COMPUTE V148=SUM((V22+V24+V27+V29+V31+V33+V35+V37+V38+V40)-
(V21+V23+V25+V26+V28+V30+V32+V34+V36+V39)).
EXECUTE.
COMPUTE V149=SUM(V41,V42,V43,V44,V45,V46,V47,V48,V49,V50).
EXECUTE.
COMPUTE V150=SUM(V51,V52,V53,V54,V55,V56,V57,V58,V59,V60,V61,V62,V63).
EXECUTE.
COMPUTE
V151=SUM(V64,V65,V66,V67,V68,V69,V70,V71,V72,V73,V74,V75,V76,V77,V78,V79,V80,V81,V82,
V83,V84,V85,V86,V87,V88,V89,V90,V91,V92,V93,V94,V95,V96).
EXECUTE.
COMPUTE
V152=SUM(V97,V98,V99,V100,V101,V102,V103,V104,V105,V106,V107,V108,V109,V110,V111,V112
,V113,V114,V115,V116,V117,V118,V119).
EXECUTE.
COMPUTE
V153=SUM(V120,V121,V122,V123,V124,V125,V126,V127,V128,V129,V130,V131,V132,V133,V134,V
135,V136,V137,V138,V139,V140,V141,V142,V143).
EXECUTE.

SAVE TRANSLATE OUTFILE='ScaleValidations.DAT' /TYPE=TAB.
EXECUTE.
```

Study 3 – Parcel and Augments Creation Script

```
*Creates Parcels**
**IIS***

COMPUTE II1a=(Q1+Q6+Q7+Q12+Q13+Q18)/6.
COMPUTE II1b=(Q2+Q5+Q8+Q11+Q14+Q17)/6.
COMPUTE II1c=(Q3+Q4+Q9+Q10+Q15+Q16)/6.
COMPUTE II2a=(QB1+QB6+QB7+QB12+QB13+QB18)/6.
COMPUTE II2b=(QB2+QB5+QB8+QB11+QB14+QB17)/6.
COMPUTE II2c=(QB3+QB4+QB9+QB10+QB15+QB16)/6.
COMPUTE II3a=(QC1+QC6+QC7+QC12+QC13+QC18)/6.
COMPUTE II3b=(QC2+QC5+QC8+QC11+QC14+QC17)/6.
COMPUTE II3c=(QC3+QC4+QC9+QC10+QC15+QC16)/6.
COMPUTE II4a=(QD1+QD6+QD7+QD12+QD13+QD18)/6.
COMPUTE II4b=(QD2+QD5+QD8+QD11+QD14+QD17)/6.
COMPUTE II4c=(QD3+QD4+QD9+QD10+QD15+QD16)/6.

**PRCA***

COMPUTE PR1a=(Q19+Q24+Q25+Q30+Q31+Q36+Q37+Q42)/8.
COMPUTE PR1b=(Q20+Q23+Q26+Q29+Q32+Q35+Q38+Q41)/8.
COMPUTE PR1c=(Q21+Q22+Q27+Q28+Q33+Q34+Q39+Q40)/8.
COMPUTE PR2a=(QB19+QB24+QB25+QB30+QB31+QB36+QB37+QB42)/8.
COMPUTE PR2b=(QB20+QB23+QB26+QB29+QB32+QB35+QB38+QB41)/8.
COMPUTE PR2c=(QB21+QB22+QB27+QB28+QB33+QB34+QB39+QB40)/8.
COMPUTE PR3a=(QC19+QC24+QC25+QC30+QC31+QC36+QC37+QC42)/8.
COMPUTE PR3b=(QC20+QC23+QC26+QC29+QC32+QC35+QC38+QC41)/8.
```

```
COMPUTE PR3c=(QC21+QC22+QC27+QC28+QC33+QC34+QC39+QC40)/8.
COMPUTE PR4a=(QD19+QD24+QD25+QD30+QD31+QD36+QD37+QD42)/8.
COMPUTE PR4b=(QD20+QD23+QD26+QD29+QD32+QD35+QD38+QD41)/8.
COMPUTE PR4c=(QD21+QD22+QD27+QD28+QD33+QD34+QD39+QD40)/8.

*CCSR*

COMPUTE CC1a=(Q43+Q48+Q49+Q54+Q55+Q60+Q61)/7.
COMPUTE CC1b=(Q44+Q47+Q50+Q53+Q56+Q59)/6.
COMPUTE CC1c=(Q45+Q46+Q51+Q52+Q57+Q58)/6.
COMPUTE CC2a=(QB43+QB48+QB49+QB54+QB55+QB60+QB61)/7.
COMPUTE CC2b=(QB44+QB47+QB50+QB53+QB56+QB59)/6.
COMPUTE CC2c=(QB45+QB46+QB51+QB52+QB57+QB58)/6.
COMPUTE CC3a=(QC43+QC48+QC49+QC54+QC55+QC60+QC61)/7.
COMPUTE CC3b=(QC44+QC47+QC50+QC53+QC56+QC59)/6.
COMPUTE CC3c=(QC45+QC46+QC51+QC52+QC57+QC58)/6.
COMPUTE CC4a=(QD43+QD48+QD49+QD54+QD55+QD60+QD61)/7.
COMPUTE CC4b=(QD44+QD47+QD50+QD53+QD56+QD59)/6.
COMPUTE CC4c=(QD45+QD46+QD51+QD52+QD57+QD58)/6.

**Constructed Experience Scale**

COMPUTE EX1a=(Q62+Q67+Q68+Q73+Q74)/5.
COMPUTE EX1b=(Q63+Q66+Q69+Q72+Q75)/5.
COMPUTE EX1c=(Q64+Q65+Q70+Q71+Q76+Q77)/6.
COMPUTE EX2a=(QB62+QB67+QB68+QB73+QB74)/5.
COMPUTE EX2b=(QB63+QB66+QB69+QB72+QB75)/5.
COMPUTE EX2c=(QB64+QB65+QB70+QB71+QB76+QB77)/6.
COMPUTE EX3a=(QC62+QC67+QC68+QC73+QC74)/5.
COMPUTE EX3b=(QC63+QC66+QC69+QC72+QC75)/5.
COMPUTE EX3c=(QC64+QC65+QC70+QC71+QC76+QC77)/6.
COMPUTE EX4a=(QD62+QD67+QD68+QD73+QD74)/5.
COMPUTE EX4b=(QD63+QD66+QD69+QD72+QD75)/5.
COMPUTE EX4c=(QD64+QD65+QD70+QD71+QD76+QD77)/6.




EXECUTE.

**Creates Augment scores for the parcels**

COMPUTE II1AugA=MEAN(Q1,Q6,Q7,Q12,Q13,Q18).
COMPUTE II1AugB=MEAN(Q2,Q5,Q8,Q11,Q14,Q17).
COMPUTE II1AugC=MEAN(Q3,Q4,Q9,Q10,Q15,Q16).
COMPUTE II2AugA=MEAN(QB1,QB6,QB7,QB12,QB13,QB18).
COMPUTE II2AugB=MEAN(QB2,QB5,QB8,QB11,QB14,QB17).
COMPUTE II2AugC=MEAN(QB3,QB4,QB9,QB10,QB15,QB16).
COMPUTE II3AugA=MEAN(QC1,QC6,QC7,QC12,QC13,QC18).
COMPUTE II3AugB=MEAN(QC2,QC5,QC8,QC11,QC14,QC17).
COMPUTE II3AugC=MEAN(QC3,QC4,QC9,QC10,QC15,QC16).
COMPUTE II4AugA=MEAN(QD1,QD6,QD7,QD12,QD13,QD18).
COMPUTE II4AugB=MEAN(QD2,QD5,QD8,QD11,QD14,QD17).
COMPUTE II4AugC=MEAN(QD3,QD4,QD9,QD10,QD15,QD16).
```

```
COMPUTE PR1AugA=MEAN(Q19,Q24,Q25,Q30,Q31,Q36,Q37,Q42).
COMPUTE PR1AugB=MEAN(Q20,Q23,Q26,Q29,Q32,Q35,Q38,Q41).
COMPUTE PR1AugC=MEAN(Q21,Q22,Q27,Q28,Q33,Q34,Q39,Q40).
COMPUTE PR2AugA=MEAN(QB19,QB24,QB25,QB30,QB31,QB36,QB37,QB42).
COMPUTE PR2AugB=MEAN(QB20,QB23,QB26,QB29,QB32,QB35,QB38,QB41).
COMPUTE PR2AugC=MEAN(QB21,QB22,QB27,QB28,QB33,QB34,QB39,QB40).
COMPUTE PR3AugA=MEAN(QC19,QC24,QC25,QC30,QC31,QC36,QC37,QC42).
COMPUTE PR3AugB=MEAN(QC20,QC23,QC26,QC29,QC32,QC35,QC38,QC41).
COMPUTE PR3AugC=MEAN(QC21,QC22,QC27,QC28,QC33,QC34,QC39,QC40).
COMPUTE PR4AugA=MEAN(QD19,QD24,QD25,QD30,QD31,QD36,QD37,QD42).
COMPUTE PR4AugB=MEAN(QD20,QD23,QD26,QD29,QD32,QD35,QD38,QD41).
COMPUTE PR4AugC=MEAN(QD21,QD22,QD27,QD28,QD33,QD34,QD39,QD40).

COMPUTE CC1AugA=MEAN(Q43,Q48,Q49,Q54,Q55,Q60,Q61).
COMPUTE CC1AugB=MEAN(Q44,Q47,Q50,Q53,Q56,Q59).
COMPUTE CC1AugC=MEAN(Q45,Q46,Q51,Q52,Q57,Q58).
COMPUTE CC2AugA=MEAN(QB43,QB48,QB49,QB54,QB55,QB60,QB61).
COMPUTE CC2AugB=MEAN(QB44,QB47,QB50,QB53,QB56,QB59).
COMPUTE CC2AugC=MEAN(QB45,QB46,QB51,QB52,QB57,QB58).
COMPUTE CC3AugA=MEAN(QC43,QC48,QC49,QC54,QC55,QC60,QC61).
COMPUTE CC3AugB=MEAN(QC44,QC47,QC50,QC53,QC56,QC59).
COMPUTE CC3AugC=MEAN(QC45,QC46,QC51,QC52,QC57,QC58).
COMPUTE CC4AugA=MEAN(QD43,QD48,QD49,QD54,QD55,QD60,QD61).
COMPUTE CC4AugB=MEAN(QD44,QD47,QD50,QD53,QD56,QD59).
COMPUTE CC4AugC=MEAN(QD45,QD46,QD51,QD52,QD57,QD58).

COMPUTE EX1AugA=MEAN(Q62,Q67,Q68,Q73,Q74).
COMPUTE EX1AugB=MEAN(Q63,Q66,Q69,Q72,Q75).
COMPUTE EX1AugC=MEAN(Q64,Q65,Q70,Q71,Q76,Q77).
COMPUTE EX2AugA=MEAN(QB62,QB67,QB68,QB73,QB74).
COMPUTE EX2AugB=MEAN(QB63,QB66,QB69,QB72,QB75).
COMPUTE EX2AugC=MEAN(QB64,QB65,QB70,QB71,QB76,QB77).
COMPUTE EX3AugA=MEAN(QC62,QC67,QC68,QC73,QC74).
COMPUTE EX3AugB=MEAN(QC63,QC66,QC69,QC72,QC75).
COMPUTE EX3AugC=MEAN(QC64,QC65,QC70,QC71,QC76,QC77).
COMPUTE EX4AugA=MEAN(QD62,QD67,QD68,QD73,QD74).
COMPUTE EX4AugB=MEAN(QD63,QD66,QD69,QD72,QD75).
COMPUTE EX4AugC=MEAN(QD64,QD65,QD70,QD71,QD76,QD77).

EXECUTE.
```

APPENDIX J: MPLUS SCRIPTS

Study 1 – Imputation Syntax

```
INPUT INSTRUCTIONS

  TITLE: 1030 Assessment Imputation Input File
    DATA: FILE = AssessmentFinal.dat;
   VARIABLE: NAMES = Q1-Q66;
              USEVARIABLES = Q1-Q66;
                   MISSING = ALL (99999);

          ANALYSIS:
                   TYPE = BASIC;
                   BSEED = 59765;
                   BCONVERGENCE = .01;
                   BITERATIONS = (4000);
                   PROCESSORS = 2;
                   THIN = 1000;
          DATA IMPUTATION:

          IMPUTE = Q1-Q30 Q61-Q66;

          NDATASETS = 40;
          SAVE = 1030PRPSAPlusAug*.dat;

          OUTPUT:

          TECH8;
```

Study 1 – Analysis Syntax

INPUT INSTRUCTIONS

```
TITLE: 1030 Assessment Analysis File Structural Paths
 DATA: FILE = 1030PRPSAPlusAuglist.dat;
    TYPE = IMPUTATION;
 VARIABLE: NAMES = Q1-Q66;
     USEVARIABLES =  Q1-Q30 Q66;
   MODEL:  PRPSA1 BY Q1*(1)
                Q2(2)
                Q3(3);
           PRPSA2 BY Q4*(1)
                Q5(2)
                Q6(3);
           PRPSA3 BY Q7*(1)
                Q8(2)
                Q9(3);
           PRPSA4 BY Q10*(1)
                Q11(2)
                Q12(3);
           PRPSA5 BY Q13*(1)
                Q14(2)
                Q15(3);
           SPPSC1 BY Q16*(4)
                Q17(5)
                Q18(6);
           SPPSC2 BY Q19*(4)
                Q20(5)
                Q21(6);
           SPPSC3 BY Q22*(4)
                Q23(5)
                Q24(6);
           SPPSC4 BY Q25*(4)
                Q26(5)
                Q27(6);
           SPPSC5 BY Q28*(4)
                Q29(5)
                Q30(6);
         PRPSA1 ON Q66; ! Regress average speech experience on PRPSA at wave 1
         SPPSC1 ON Q66; ! Regress average speech experience on SPPSC at wave 1
         SPPSC2 ON Q66;
         PRPSA1@1.0;! PRPSA2@1.0 PRPSA3@1.0 PRPSA4@1.0 PRPSA5@1.0;
         SPPSC1@1.0;! SPPSC2@1.0 SPPSC3@1.0 SPPSC4@1.0 SPPSC5@1.0;


         !PRPSA Auto-Regressive Paths
         PRPSA2 ON PRPSA1;
         PRPSA3 ON PRPSA2;
         PRPSA4 ON PRPSA3;
         PRPSA5 ON PRPSA4;

         !SPPSC Auto-Regressive-Paths
         SPPSC2 ON SPPSC1;
```

```
SPPSC3 ON SPPSC2;
SPPSC4 ON SPPSC3;
SPPSC5 ON SPPSC4;

!SPPSC regressed on PRPSA
!PRPSA1 ON SPPSC1;
!PRPSA2 ON SPPSC2;
!PRPSA3 ON SPPSC3;
!PRPSA4 ON SPPSC4;
!PRPSA5 ON SPPSC5;

!SPPSC Cross-Lag one timepoint regressed on PRPSA
PRPSA2 ON SPPSC1;
!PRPSA3 ON SPPSC2;
PRPSA4 ON SPPSC3;
PRPSA5 ON SPPSC4;

SPPSC2 ON PRPSA1;
!SPPSC3 ON PRPSA2;
!SPPSC4 ON PRPSA3;
!SPPSC5 ON PRPSA4;



[PRPSA2 PRPSA3 PRPSA4 PRPSA5 SPPSC2 SPPSC3 SPPSC4 SPPSC5];
 !WITHIN TIME COVARIANCES
 PRPSA1 WITH SPPSC1;
 !PRPSA2 WITH SPPSC2;
 PRPSA3 WITH SPPSC3;
 PRPSA4 WITH SPPSC4;
 PRPSA5 WITH SPPSC5@0;

[Q1 Q4 Q7 Q10 Q13] (7); ! This section estimates intercepts
[Q2 Q5 Q8 Q11 Q14] (8); ! But assumes measurement invariance
[Q3 Q6 Q9 Q12 Q15] (9); ! Which has been previously been established
[Q16 Q19 Q22 Q25 Q28] (10);
[Q17 Q20 Q23 Q26 Q29] (11);
[Q18 Q21 Q24 Q27 Q30] (12);

    !Correlated Residuals for PRPSA Indicators
 Q1 WITH Q4 Q7 Q10 Q13;
 Q4 WITH Q7 Q10 Q13;
 Q7 WITH Q10 Q13;
 Q10 WITH Q13;
 Q2 WITH Q5 Q8 Q11 Q14;
 Q5 WITH Q8 Q11 Q14;
 Q8 WITH Q11 Q14;
 Q11 WITH Q14;
 Q3 WITH Q6 Q9 Q12 Q15;
 Q6 WITH Q9 Q12 Q15;
 Q9 WITH Q12 Q15;
 Q12 WITH Q15;

          !Correlated Residuals for SPPSC Indicators
          Q16 WITH Q19 Q22 Q25 Q28;
```

```
Q19 WITH Q22 Q25 Q28;
Q22 WITH Q25 Q28;
Q25 WITH Q28;
Q17 WITH Q20 Q23 Q26 Q29;
Q20 WITH Q23 Q26 Q29;
Q23 WITH Q26 Q29;
Q26 WITH Q29;
Q18 WITH Q21 Q24 Q27 Q30;
Q21 WITH Q24 Q27 Q30;
Q24 WITH Q27 Q30;
Q27 WITH Q30;


OUTPUT: STDYX tech1 tech4;
```

Study 2 – Imputation Syntax

```
INPUT INSTRUCTIONS

  TITLE: Scale Validation Modeling Test File
   DATA: FILE = ScaleFinal.dat;
   VARIABLE: NAMES = Q1-Q143;
     !CATEGORICAL = Q41-Q96;
     MISSING = ALL (99999);
  DATA IMPUTATION:IMPUTE = Q1-Q40 Q41-Q96 Q97-Q143;
            NDATASETS = 40;
            SAVE = ScaleImp*.dat;
            VALUES = Q1-Q40(1-5) Q41-Q96(1-2) Q97-Q143(1-5);
            !ROUNDING = Q1-Q40(0) Q97-Q143(0);
  ANALYSIS:ESTIMATOR = BAYES;
         ALGORITHM = GIBBS (PX1);
         !PROCESSORS = 2;
         BITERATIONS = (4000);
  MODEL:
           [Q1-Q143]; !Means
           Q1-Q143; !Variances
           Q1-Q143 WITH Q1-Q143; !Covariances

    OUTPUT: TECH8 TECH4;
```

Study 2 – Analysis Syntax

```
TITLE: Scale Validation Modeling Test File
  DATA: FILE = ScaleImpEditlist.dat;
       TYPE = IMPUTATION;
  VARIABLE: NAMES = Q1-Q153;
          USEVARIABLES = Q147-153;
MODEL: Q147 ON Q148-Q153;
output: SAMPSTAT stdyx;
```

Study 3 – Imputation Syntax

```
TITLE: Accelerated Longitudinal Imputation File;
   DATA: FILE = Rubin_Master_ParcelsAugments.dat;
  VARIABLE: NAMES = Q1-Q91;
              USEVARIABLES = Q8-Q91;
                    MISSING = ALL (99999);

          ANALYSIS:
                    TYPE = BASIC;
                    BSEED = 59765;
                    BCONVERGENCE = .01;
                    BITERATIONS = (4000);
                    PROCESSORS = 2;
                    THIN = 1000;
          DATA IMPUTATION:

          IMPUTE = Q8-Q43;

          NDATASETS = 40;
          SAVE = Rubin_Master*.dat;

          OUTPUT:

          TECH8;
```

Thesis and Dissertation Services