Summer 8-2022

Longitudinal Measurement Invariance of Communication Apprehension Scores Among First-Year Undergraduates

Tonya Calloway
Louisiana Tech University

LONGITUDINAL MEASUREMENT INVARIANCE OF COMMUNICATION APPREHENSION SCORES AMONG FIRST-YEAR UNDERGRADUATES

by

Tonya Calloway, B.S., M.S., M.A.

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education, Educational Leadership

COLLEGE OF EDUCATION
LOUISIANA TECH UNIVERSITY

August 2022

LOUISIANA TECH UNIVERSITY

GRADUATE SCHOOL

May 6, 2022

Date of dissertation defense

We hereby recommend that the dissertation prepared by

Tonya Calloway, B.S., M.S., M.A.

entitled **Longitudinal Measurement Invariance of Communication Apprehension Scores Among First-Year Undergraduates**

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Education, Organizational Leadership Concentration

Marita Apter-Desselles
Supervisor of Dissertation Research

Dustin Hebert
Head of Curriculum, Instruction, and Leadership

Doctoral Committee Members:
Bryan McCoy
Dawn Basinger

Approved:

Don Schillinger
Dean of Education

Approved:

Ramu Ramachandran
Dean of the Graduate School

GS Form 13a
(01/20)

ABSTRACT

Testing measurement invariance (MI) is the most practical question to address in any analysis that involves multiple time points and/or groups. MI is a prerequisite to evaluate if an observed true change over time has occurred after an intervention. Communication Apprehension is one of the most widely studied constructs in the field of communication but has not been analyzed for MI. Leaders governing higher education institutions have implemented Quality Enhancement Plans (QEP) that strategically address workforce needs such intervention programs to improve student communication competency. Despite intervention efforts, industry still indicates a high demand for a workforce with communication competency. This study explores this issue by applying a tripartite model of change to assess the presence of alpha, beta, and gamma change in communication participants. The present study also conducted a secondary analysis using archival data from a communication intervention using college freshman. Factorial invariance was examined through the evaluation of three hierarchical levels of MI: configural, weak, and strong invariance. Results supported all three levels of MI; MI was upheld, and alpha change was determined to have occurred.

APPROVAL FOR SCHOLARLY DISSEMINATION

DEDICATION

This dissertation is dedicated to in loving memory of my grandmother, (Momo) Gladys Bell Calloway Gilbert, who gave me the strongest foundation and will to succeed. I would also like to dedicate my work to all my family and friends for providing love, encouragement, and unwavering support throughout my doctoral program. Mama, you were my first teacher and instilled confidence and a love of learning in me that has carried me through all my educational accomplishments. I still remember the Jack and Jill books.

Caleb, this has been a long journey, and I want you to know you were my motivation. May my inspirations be your aspirations.

TABLE OF CONTENTS

ABSTRACT.. iii

APPROVAL FOR SCHOLARLY DISSEMINATION iv

DEDICATION...v

LIST OF TABLES...x

LIST OF FIGURES xi

ACKNOWLEDGMENTS xii

CHAPTER 1 INTRODUCTION...1

Methodological Issues in Intervention Evaluation 6

Design Issues in Longitudinal Studies..6

Complexities Designing Longitudinal Studies 7

Structural Equation Modeling a Viable Remedy to Internal Validity Threats 9

Research Designs and Measurement in Evaluation Studies 11

Alpha, Beta, and Gamma Change and Assessment 13

History of Alpha, Beta, and Gamma Change 13

Assessing Change using Factor Analysis 14

Using the Measurement Model to Test Invariance 16

Measurement Invariance...16

First and Second-Order Measurement Invariance 17

Configural Invariance 18

Weak Factorial Invariance 18
Strong Factorial Invariance 19
Strict Invariance 19
Second-Order Invariance 20
Testing Measurement Invariance Models for Data Fit 21
Longitudinal Null Model 21
Statistical Rationale and Modeling Rationale 22
Scale Setting 24
Identification 25
Power 26
Sample Size 27
Relative Fit Measures 27
Research on Communication Apprehension 28
Communication Apprehension Construct 28
Measuring Communication Apprehension 30
Assessing the Quality of Communication Apprehension Interventions 34
Applying the Alpha Beta Gamma Change Model to Communication Interventions 36
Hypotheses 39
CHAPTER 2 METHOD 41
Participants and Measure 41
Data Analytic Approach 44
Power, Sample Size, and Degrees of Freedom 44
Parcels 46

Analysis Software 46
Missing Data 46
Model Fit Test Indices 47
Scale Setting and Identification 48
Hypothesis Testing 48
Hypothesis 1 48
Hypothesis 1 Rationale 51
Hypothesis 2 52
Hypothesis 2 Rationale 52
Hypothesis 3 53
Hypothesis 3 Rationale 54
Hypothesis 4 54
Hypothesis 4 Rationale 54
Hypothesis 5 54
Hypothesis 5 Rationale 54
CHAPTER 3 RESULTS 56
Pre-Analysis 56
A Case for Parcels 57
Parcels Improve Baseline Model 60
Lower-Order Factor Invariance Testing 62
Higher-Factor Invariance Testing 68
CHAPTER 4 DISCUSSION 70
Principal Findings 70

Limitations and Future Directions 72

Recommendations for Educational Leaders 75

Conclusion 76

REFERENCES 78

APPENDIX A HUMAN USE EXEMPTION LETTER 89

LIST OF TABLES

Table 1 *Indicators of the PRCA-24* 49

Table 2 *Hypothesis Testing for MI and Parameter Constraints for One Group Repeated Second-Order Measure* 50

Table 3 *Correlations, Means, and Standard Deviations Among Parceled Indicators for First-Order Factors* 63

Table 4 *Model Fit Statistics for the Tests of Invariance for First-Order and Second-Order Factors for Communication Apprehension* 64

LIST OF FIGURES

Figure 1 *Over-Identified Second-Order Measurement Model No Parcels* 58

Figure 2 *Just-Identified Longitudinal Second-Order Measurement Model with Parcels* 61

Figure 3 *Configural (Baseline) First-Order Model* 65

Figure 4 *Second-Order CA Partial Strong Invariant Model* 69

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Mitzi Apter-Desselles, for your guidance throughout the doctoral program. I would also like to acknowledge my committee members, Dr. Dawn Basinger and Dr. Bryan McCoy, for your support. I am honored to have the opportunity to learn and grow as a professional amongst some of the very best. Thank you for sharing your knowledge with me during this process.

CHAPTER 1

INTRODUCTION

Over the last few decades, many higher education institutions have undergone transformations to become more globally competitive, maintain accreditation, attract, and retain diverse student bodies, and prepare students for the burgeoning workforce (Gagliardi et al., 2018; Webber & Zheng, 2020). Leaders governing these institutions have implemented intervention programs to enhance the quality of education. They increasingly rely on evidence-based practices and data to improve institutional performance through the direct assessment of student interventions (Gagliardi et al., 2018; Webber & Zheng, 2020).

A major accrediting body, the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC), is committed to advancing the quality of higher education, and the Quality Enhancement Plan (QEP) is an important part of the accreditation reaffirmation (SACSCOC, 2020). The QEP is an opportunity for an institution to implement strategies for improving students' learning and fostering students' skills focused on employability (SACSCOC, 2020). Also, QEPs are developed and implemented based on the universities' strategic plans. Generally, QEPs incorporate institutional research that generates data to be processed and analyzed for use in both academic and administrative pursuits to foster student learning and successes. Effectively processed and analyzed data from institutional research is fundamental for leaders to make data-informed decisions (Gagliardi et al., 2018). Moreover, Gagliardi et al.(2018) states, “data from institutional research should be based on accuracy, timelines, relevance, integration and security” (p. 3). These are important analytics for effective data-driven decision making. However, the central analytic focus of this study is evaluating the accuracy of analyzed data used in institutional research to evaluate an intervention program. Specifically, leaders are demanding more insights and accuracy from institutional research data and traditional analyses (i.e. multiple regression, ANOVA) may not provide accurate results based on limiting assumptions about the data (Gagliardi et al., 2018; Little, 2013). For example, interventions implemented in QEPs foster student development that become a fundamental component of the academic experience; when students learn, gain new skills, and develop new interests and attitudes, they experience change. These changes should be measured accurately and appropriately to determine students’ progress (Hodis et al., 2010) and the effectiveness of educational systems (Boyas et al., 2012; Garside, 2010).

Communication competency is considered a highly valued soft skill to prospective employers (Rios et al., 2020). As a result, many institutions of higher education strategic plans have included the addition of a basic communication course requirement, sometimes accompanied by communication intervention programs (Du-Babcock, 2006; Morreale & Pearson, 2008). Additionally, “the Association of American Colleges and Universities (AAC&U) now includes communication as a recommended intellectual and practical skill in its description of essential learning outcomes for college students” (Morreale et al., 2014, p. 351). Furthermore, communication skills are considered more important than other competencies across occupations (Becker &

Eckdom, 1980; Du-Babcock, 2006; Morreale & Pearson, 2008; Shanahan, 2013; Winsor et al., 1997). Moreover, a study published in *Educational Researcher* reported, "when employers demanded multiple skills, the most in-demand pairing was oral and written communication with this pairing being demanded 180% more than the second most popular pairing" (Rios et al., 2020, p. 83). Employers want students who are effective communicators (Cavanagh et al., 2006).

Despite intervention efforts by higher education, industry leaders still report a gap in communication competence of newly hired graduates (Cavanagh et al., 2006; Gaff, 1981; Rios et al., 2020). Moreover, some empirical studies assessing communication skills in the 21st century indicate skills gaps remain and have widened in some occupations (Mitchell et al., 2010; Rios et al., 2020). Closing the gap for communication skills to meet workforce expectations will require high-quality communication intervention programs as well as robust approaches in how the impact of such programs is measured. Much attention has been focused on developing the content and pedagogy of communication interventions programs and less on the assessment of these programs (Hsu, 2009; Pribyl et al., 1998). How one measures change may be hindering efforts to understand whether and how our communication skills training is benefiting students. When institutions fully and validly assess the impact of communication interventions, the feedback loop between intervention, assessment, and redesign becomes stronger and holds the potential for greater insights and, ultimately, greater effectiveness. Specifically, data used to evaluate communication interventions rely heavily on student self-reports. Self-report data has a long history of being criticized as an outcome measure (Goldstein & Ford, 2002; McCroskey, 2009). Additionally, data are often analyzed using statistical procedures that compare the means between pre- and post-intervention scores and rely on several limiting assumptions (e.g., homogeneity of variance, measurement invariant). For example, reliance on these pre-posttest comparisons may render conclusions invalid because self-report data are assumed to be relatively stable dimensions when they are not (Howard et al., 1979).

The Golembiewski et al. (1976) tripartite model of change may be particularly useful for communication intervention studies that are focused on improving students' communication skills and measuring the impact via self-report. These authors differentiated three types of change: alpha, beta, and gamma (ABG) change. In defining each type, they also highlighted several measurement issues arising from the use of self-report data and repeated measure designs. Their definitions of the three types of change are described as follows:

1. Alpha change involves a variation in the level of some existential state, given a constantly calibrated measuring instrument related to a constant conceptual domain.
2. Beta change involves a variation in the level of some existential state, complicated by the fact that some intervals of the measurement continuum associated with a constant conceptual domain have been recalibrated.
3. Gamma change involves a redefinition or reconceptualization of some domain, a major change in the perspective or frame-of-reference within which phenomena are perceived and classified in what is taken to be relevant in some slice of reality (Golembiewski et al., 1976).

Alpha change has been the predominant type of change assumed by those who examine raw score differences (Ericson & Gardner, 1992**;** McCroskey et al., 1989; Vandenberg & Self, 1993). If gamma or beta change has occurred, explanation of results assuming only alpha change may lead to inaccurate conclusions because the scaling of the instrument and/or the underlying latent construct have changed.

Golembiewski et al.'s (1976) typology of change spurred researchers to examine factor structures across time and groups when evaluating the impact of an intervention, especially when data are self-reported. Today, analyzing factor structures across time and groups is commonly referred to as measurement invariance (MI) or factorial invariance. MI indicates the level of degree that the same underlying construct is being measured across time or groups (Little, 2013; Vandenberg & Lance, 2000). Conceptually, MI is the statistical evaluation to determine if a measure has the same psychometric properties across times or groups. Testing MI is the most practical question to address in any analysis that involves multiple time points and/or groups (Little, 2013; Meredith, 1993; Widaman & Reise, 1997). Since the evaluation of most social and behavioral designs are complex and dynamic, MI is a prerequisite to evaluate if an observed true change over time has occurred after an intervention (Little, 2013; Vandenberg & Lance, 2000). Otherwise, reporting any observed change would be biased and deleterious (Chen, 2007; Little, 2013; Vandenberg & Lance, 2000).

The study aims to conduct a secondary analysis using archival data to evaluate if an observed true change over time has occurred after a university-wide communication intervention using college freshman. Theoretically, the purpose of this study is to apply the ABG literature and measurement invariance methods to evaluate the impact of a communication skills intervention for university freshmen. In the communication literature, communication apprehension (CA) is considered one of the most useful and practical constructs to determine students' communication competence after a communication intervention program **(**Hsu, 2009; Kelly & Keaten, 1992; Levine & McCroskey, 1990; Pribyl et al., 1998; Rubin et al., 1990, 1997). "CA is defined as the fear or anxiety associated with real or anticipated communication with others" (McCroskey, 1982, p. 279). The construct is primarily measured using McCroskey's (1982) self-report Personal Report of Communication Apprehension-24 (PRCA-24). However, the measurement properties of the PRCA-24 have been problematic at times (e.g., different factor structures, overestimated effect sizes). It may be that the construct of CA is changed because of communication skills intervention.

Methodological Issues in Intervention Evaluation

Design Issues in Longitudinal Studies

Social and behavioral studies that investigate the effects of an intervention are best conducted using longitudinal designs (Little, 2013). The goal of evaluation research is to discern whether and how participants in an intervention have changed over time. Longitudinal designs involve the dynamic interplay of the context with cohort, age, and time-of-measurement effects. Interpretation and analyzing longitudinal studies are often challenges when the researcher fails to account for these possible confounding effects within the design (Little, 2013). Implementing an optimal longitudinal study design relies on the skill and knowledge of the researcher. A researcher who is guided by strong theory "will guide their thinking through most of these design/statistical conundrums" (Little, 2013, p. 43) to disentangle the confounds that affect longitudinal designs. Even still, designing a flawless longitudinal design may not be achievable. Often, the ideal longitudinal design is extremely time and resource consuming. As a result, “many longitudinal studies with less-than-ideal designs are often insufficiently exploited. Often only cross-sectional analyses are conducted on the different data points, and threats to internal validity of the studies are not examined even when data is available” (Schale, 1998, p. 10).

Cross-sectional designs have sometimes been used in intervention assessments; these designs do not use a repeated measurement approach. As a result, cross-sectional designs have limited ability to describe change over time or groups and are also confounded with “age cohort differences and strongly influenced by between-group sampling variability” (Little, 2013, p. 39). Cross-sectional designs are best suited to address measurement validity issues across groups. Specifically, cross-sectional designs may be employed to assess factorial invariance of constructs across groups before engaging in a longitudinal study (Little, 2013). Longitudinal designs are time consuming and prone to internal validity threats; however, these designs still offer higher levels of validity than cross-sectional designs.

Complexities Designing Longitudinal Studies

Implementing true longitudinal designs requires sufficient time to allow growth and change to emerge. Some developmental changes and growth arising from educational interventions may take many years to emerge, both of which require a long commitment from the researcher as well the participant. “Causes take time to exert their effects and the ability to detect effects depends on the time interval between measurement” (Little, 2013, p. 47). Timing issues may also become confounded with age-related effects. If studies take significant amounts of time, a researcher will have to consider historical changes with age-related effects. Little (2013) recommends indexing time to model change to identify patterns of age-related effects. He identifies two processes for representing change as a function of time: episodic time and experiential time. Episodic time is indexed based on a key developmental episode of interest (e.g., puberty) or event and may not be centered on age. Experiential time is focused on chronological age and "how long participants have experienced a state or process" (Little, 2013, p. 52).

An additional challenge in longitudinal studies is the large sample size required to obtain meaningful results about the patterns and relationships that develop over time. Unfortunately, it is often challenging to find individuals who are willing to participate for the duration of a longer study. Participants dropping out part-way through a study may lead to high longitudinal attrition and missing data. "In longitudinal datasets, the amount of missing data often approaches levels that make even make quantitatively minded scholars nervous" (Little, 2013, p. 58). However, the missing data problem may be remedied with relatively recent statistical approaches (e.g., data imputation), which have supplanted traditional approaches in which researchers simply discard observations with incomplete data.

Little (2013) does not recommend the commonly used classical technique of listwise or pairwise deletion to handle missing data in longitudinal designs. He describes modern missing data approaches as reconstructive surgery in contrast to classical approaches which he likens to surgical removal. Little (2013) argues that modern missing data approaches attempt to regain power from missing data, but classical approaches do not. He recommends addressing the issue of missing data via imputation such as full-information maximum likelihood (FIML) or data-based expectation maximization (EM). The available evidence indicates the two methods, one model-based (i.e., FIML) and the other data-based (i.e., EM), tend to produce essentially identical results (Little, 2013). Three basic mechanisms give rise to missing data in longitudinal studies (Little, 2013), and the recommended approach to data imputation is influenced by which mechanism is operating. The mechanisms for missing data are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Most missing data in longitudinal studies may be described as MAR in which missing data occur from attrition and not from an association with an unmeasured variable. A full discussion of the pros and cons of various data imputation techniques is outside the scope of this dissertation. Little's (2013) coverage of the issues concludes with the recommendation that data-based approaches such as EM are sufficient to handle data that is MAR, while FIML, a model-based approach, is more appropriate for MNAR and MCAR. However, he also cautions that the trustworthiness of missing data imputations by any means depends on the amount of missing data. For example, 60% of missing data from a sample of 100 is different than 60% of missing data from a sample of 1,000. In the first example (n = 40), generalizability would be lower when compared to second example (n = 400) (Little, 2013). In sum, data imputation is a viable approach in dealing with the challenge of missing data when conducting longitudinal research.

Structural Equation Modeling a Viable Remedy to Internal Validity Threats

Longitudinal designs are also fraught with other threats to internal validity such as practice effects (retest effects), regression to the mean, and instrumentation effects (factorial invariance) (Little, 2013; Schale, 1998). Retest effects are a function of the repeated practice of a measure; a remedy for this threat is often difficult to accomplish. Retest effects may also be remedied with random assignment to a measurement occasion. This creates intentional missing data, which may be addressed using one of the modern missing data approaches previously discussed. Regression to the mean is the tendency for extremes scores to move closer to the mean distribution at subsequent waves of the test. This may be remedied with latent-variable structural equation modeling (SEM). Because regression to the mean is a function of unreliability, SEM measures the variance in latent constructs rather than the variance in manifest variables, which contain measurement error (Little, 2013). The utility of SEM approaches, particularly the measurement model, will be reviewed throughout the sections in Chapter 4.

Central to this study are the threats to the validity of longitudinal studies associated with instrumentation effects arising from self-report measures. For example, treatment effects are confounded with instrumentation especially when the purpose of the treatment is to change the subject's understanding of the variable being measured (Howard et al., 1979). If the treatment has indeed changed how the participant conceptualizes the outcome variable, then the confounding effects will impact the post-intervention measures. Instrumentation effects may influence conclusions of longitudinal studies when the measurement properties of an instrument change over time. Understanding these confounds is central to observing stability of change over time (Golembiewski et al., 1976). Moreover, if a measured construct is not invariant then conclusions made would be erroneous (Little, 2013).

Instrumentation effects may be remedied by testing factorial invariance to determine if the construct under investigation has not changed across time-points. MI is the fundamental first step in the latent variable approach of SEM in which the measurement model is tested via confirmatory factor analysis (CFA). Evaluating the measurement model to determine change between data sets across time is referred to as a longitudinal CFA. When the measurement model is consistent across time or groups, MI is inferred.

Much evaluative research relies heavily on student report data and does not employ longitudinal designs (Howard et al., 1979). The next section will recap the one-group pretest-posttest design commonly used in these investigations into the impact of interventions. The Golembiewski et al. (1976) tripartite model of change will be discussed as a more valid and practical way to interpret outcomes measured via student self-report.

Research Designs and Measurement in Evaluation Studies

Historically, interventions have relied on one-group pretest-posttest designs in which outcomes are measured by self-report instruments evaluated using traditional analyses (e.g., repeated measures analysis of variance) (Howard et al., 1979). These types of design are often difficult to determine if a significant change has occurred because they do not include a comparison group. Furthermore, in certain instances, especially educational interventions, experimental designs, and random assignments may not be possible because of ethical violations **(**Shadish et al., 2002).

The most common approach to measure change in pretest-posttest designs is repeated measures ANOVA. The ANOVA family of statistics relies on using observed variables, which are assumed to be normally distributed, have equal population variances, and suffer from several limitations because they rely heavily on untested assumptions

(Little, 2013). Real data are seldom normally distributed and do not meet the homogeneity of variance assumption (Little, 2013). Violations of these two assumptions can result in Type-I and Type-II error. Type-I error is erroneously rejecting the null hypothesis when it is really accurate. Type-II error is failing to reject the null hypothesis when it is false. Statistical methods that use observed variables also assume that the constructs are invariant across time and groups (Tabachnick & Fidell, 2013). If these assumptions are not tested, the results can be under or overestimated, which may lead to erroneous conclusions. Golembiewski et al. (1976) implied that an observed true change cannot be observed from interventions using classical analysis procedures.

As previously discussed, intervention studies often rely heavily on self-report instruments. The reliability and validity of self-report data is widely perceived as flawed (Porter, 2011). Self-report data from students has been criticized because respondents may not be realistic or accurate in their judgments (i.e., skills, ability, learning) (Lublin, 1980). On the other hand, it is most useful for assessing emotional and cognitive engagement that are not easily observable (Fredricks & McColskey, 2012; McCroskey, 2009). Also, self-report measures are cost effective and easier to administer compared to objective and behavioral measures. Validity of student self-report data was supported when students' ratings matched exam performance (Benton et al., 2013).

A common assumption when using self-report instruments is that the respondent's standard measurement of the construct will not change from one testing to the next (Howard & Dailey, 1979). The Golembiewski et al. (1976) ABG change model shows that change may not be static, and stability of measurement should not be assumed. They identified three different types of change that may be found when using longitudinal designs and self-report instruments. Alpha change occurs when respondents use a stable consistent scale from T1 (T1) to T2 (T2). Beta change results from a recalibrated scale within a stable psychometric dimension from T1 to T2. Gamma change is a sharp departure from the original conceptual space because respondents have reconceptualized the construct, and the psychometric dimension is substantively altered. If gamma or beta change has occurred, interpretation of the data is very likely to lead to inaccurate conclusions. However, the authors are adamant that beta and gamma change should not be considered errors. They argue these forms of change could be hypothesized and anticipated outcomes and could represent meaningful contributions to understanding the impact of planned interventions.

Alpha, Beta, and Gamma Change and Assessment

History of Alpha, Beta, and Gamma Change

The best method to measure ABG change is a confirmatory factor analysis (CFA). In fact, the earlier studies assessing ABG change used CFA methods to detect gamma change (Golembiewski et al., 1976; Millsap & Hartog, 1988; Schmitt, 1982). There was consensus among the researchers that a change in factor structure was akin to gamma change but there was disagreement regarding how to determine beta change. There was also agreement that CFA analyses procedures offer the best combination, flexibility, and accuracy to compare within groups and across time. Furthermore, CFA analysis only require a minimum of two comparisons (Little, 2013; Schmitt, 1982; Vandenberg & Lance, 2000). Unlike traditional classical analysis methods, more assumptions can be tested using a CFA (Little, 2013). MI indicates the level of degree that the same underlying construct is being measured across time or groups (Little, 2013; Vandenberg

& Lance, 2000). There are several approaches to test MI of latent constructs lenient test such as exploratory factory analysis (EFA) and stricter test such as CFA. Both EFA and CFA can be used to determine factorial structures. The next section will explain differences between an EFA and a CFA and provide support that a CFA is the most efficient way test MI and examine ABG change.

Assessing Change using Factor Analysis

Utility of CFA verses an EFA. EFA is a data driven approach to discover factorial structures; however, CFA is theory driven to confirm hypothesized factorial structures. CFA techniques require a researcher to prespecify all aspects of the measurement model. This includes evidence based on theory to determine the number of factors that exist in the indicator-factor relationship.

Like EFA, CFA can produce standardized estimates but the strength of its analytic technique is the ability to produce unstandardized estimates, which are the unstandardized variance-covariance structures and means. Unstandardized solutions provide information on the true nature of relationship among the indicators and factors. A standardized solution would mask this relationship (Brown, 2015; Little, 2013). Thus, the typical analysis for CFA is a variance-covariance matrix. Variance is calculated as the indicator’s standard deviation squared, and the covariance is determined by multiplying two indicators’ correlation times their standard deviations. The unstandardized solutions provide estimates that are expressed in terms of the raw data metrics, which are the indicators. These unstandardized estimates give way to standard errors and significance testing of the model parameters (Brown, 2015).

The CFA is a measurement model within a larger analysis framework called structural equation modeling, (SEM). SEM uses latent variable modeling analysis to measure multiple indicators used to characterize an unobserved construct. A CFA is well equipped to address applied research questions aimed at psychometric soundness of a measure; item scale reliability, method effects; and measurement invariance, which is the comparison of factor models across groups and time and analysis of mean structures. One of the key differences between CFA and EFA is the ability of CFA to specify measurement errors also known as correlated errors in model estimation. Correlated errors examine the relationship among unique variance among indicators. In a CFA, measurement errors of indicators can be pre-specified; however, in EFA the assumption of measurement of error is considered random. Correlated error can be pre-specified based on method effects.

The CFA is more parsimonious than the EFA because simple structure is obtained with fewer parameter estimates. The goal of the CFA is to find the appropriate model parameter values that make the observed data most likely, also referred to as scale reliability (Brown, 2015). CFA analysis of mean structures is an estimation of latent means and indicator intercepts. Mean structure analysis allows researchers to compare groups on the latent mean, which is akin to analysis of variance (ANOVA). Also, these estimates give rise to evaluating measurement invariance, which is determining the equivalence of a measure across time and groups.

Comparative model evaluation is also a strength of CFA. To access comparative models, constraints are imposed on the factor solution such as “constraining all the factor loadings or all the unique variances to be equal” (Brown, 2015, p. 41). The adequacy of the model is evaluated using the goodness-of-fit index, chi-square (x^2), to test the adequacy of the model after the model is fit to the data. One of the most invaluable strengths of CFA compared to EFA and other traditional methods is the ability to test for MI (Brown, 2015). In SEM, the CFA is called the measurement model. The subsequent sections will discuss MI analyses and specifying, identifying, and determining the adequacy of the measurement model.

Using the Measurement Model to Test Invariance

Measurement Invariance

MI is evaluated using a longitudinal CFA, the measurement model, which lies within a latent-variable framework of structural equation modeling (SEM). It is also referred to as the measurement model, which is the first and most crucial step of a SEM analysis (Little, 2013). SEM is a statistical method that allows researchers to identify a parsimonious model that gives rise to the latent variable and measured indicators (Karakaya-Ozyer & Aksu-Dunya, 2018; Little, 2013). The measurement model is "basically a confirmatory factor analysis model which confirms if the data fit the proposed model" (Karakaya-Ozyer & Aksu-Dunya, 2018, p. 279). In comparison to traditional classical analysis procedures, such as ANOVA and multiple regression that are often used to evaluate educational interventions, SEM makes the fewest assumptions and allows testing of most assumptions. The key strength of this approach is the ability to analyze the observed variance-covariance matrix against model implied variance-covariance matrix to test the psychometric soundness of latent constructs across time. The technique has been further developed to include comparison of construct mean levels across time or groups (Meredith, 1993; Widaman & Reise, 1997). The two combined approaches are the central idea for testing MI.

MI indicates the level of degree that the same underlying construct is being measured across time (Little, 2013; Vandenberg & Lance, 2000). Conceptually, MI is the statistical evaluation to determine if a measure has the same psychometric properties across times or groups. Since the evaluation of most social and behavioral designs are complex and dynamic, MI is a prerequisite to evaluate if an observed true change over time has occurred after an intervention (Little, 2013; Vandenberg & Lance, 2000). Otherwise, reporting any observed change would be biased and deleterious (Chen, 2007; Little, 2013; Vandenberg & Lance, 2000). Even though factor invariance has been recommended as a necessary practice in longitudinal designs (Little, 2013), the field is still considered novel. "Only two simulation studies have been conducted on determining good criteria using alternative fit indexes" (Little, 2013, p. 155). For example, once an indicator is considered problematic when evaluating factor invariance, there are not established guidelines on how to address the issue. The PRCA-24 is theorized to be a second-order measurement model. Testing second-order model invariance has even less guidelines in the literature. MI analyses for both first- and second-order models will be reviewed next.

First and Second-Order Measurement Invariance

First-order factor models represent reflective relationships between indicators and latent variables. Second-order factor models represent reflective relationships between first-order factor and second-order factors. The measurement invariance of second-order factor structures generally works in the same manner as the first-order factor models.

There are three levels of invariance that follow a nested sequence. Level 0, which is configural invariance, Level 1, which is weak invariance, and Level 2, which is strong invariance.

Configural Invariance. This is level 0 for invariance testing. When the data are from a homogeneous group, MI will be evaluated across time. The subscript *o* represents the across occasion, and Σ is the model implied covariance. Configural invariance means that that same construct is measured across time but does not guarantee that constructs are measured on the same scale with the same zero point. Configural invariance is a qualitative evaluation to determine if the relations among constructs and indicators have the same pattern. This pattern is based on fixed and freed loadings at each time point. The model is considered the baseline model, and if it is deemed acceptable, then the other levels of invariance can be evaluated against it (Little, 2013). As it pertains to the ABG model of change, this level assesses if gamma and beta change exists, but neither can be ruled out even if the data support invariance since all parameters are freely estimated. However, if invariance is not supported, it might indicate the patterns are not the same across time and constructs at pretest and posttest do not hold the same psychometric properties (gamma). If invariance is not supported, it could also indicate instability of the construct dimensions (beta).

Weak Factorial Invariance. The next level, Level 1, of factor invariance involves constraining the factor loadings, Λ, to be equal across time. Weak invariance evaluates whether the measured constructs have the same unit of measurement over time. Like the previous level, it does not guarantee that constructs are measured on the same scale with the same zero point. Weak invariance “implies that any difference in one unit of latent variable results in the same differences of the observed indication of variables across time" (Rudnev et al., 2018, p. 51). This level of invariance is met when factor loadings are the same across time and model fit is compared to the previous model. If invariance is supported, then gamma change can be ruled out as constructs are being measured by the same measurement units. However, beta change cannot be ruled out because constructs may not be "measured on the same scale with the same zero point.

Strong Factorial Invariance. Level 2 is the strong invariant model. This is a test that item intercepts are invariant across time in addition to the constrained loadings. At this level "the latent variables scales are measured with the same units and have the same zero point for all constructs for both time points" (Rudnev et al., 2018, p. 51). If strong invariance is upheld, manifest intercepts are the same for each construct overtime in addition to factor loadings. Also, latent constructs relate to the same levels in the observed variables across time, and latent means can be compared across time. If all three levels are upheld, gamma and beta change can be ruled out, and constructs are considered invariant and can be compared across time (alpha change).

Strict Invariance. Little (2013) discourages using strict invariance. Strict invariance tests equality of indicator uniqueness over time or groups. He argues that it is reasonable to assume that indicator uniqueness is the same across time, but it is unreasonable "to assume that the amount of random error present in each indicator at each time point (or across groups) would be the same" (Little, 2013, p. 143). As a result, only configural, weak and strong invariant models should be tested.

Second-Order Invariance

According to Rudnev et al. (2018) and Dimitrov (2010), the second-order model should follow a bottoms-up strategy. In this regard, the analysis will begin with the least restrictive model, the configural first-order model. Additionally, the second-order model will follow the same logic as the ordinary model, the first-order model with few differences. The prerequisite for the second-order model is configural and weak invariance of the first-order model. Rudnev et al. (2018) recommends the next level of invariance is the weak invariance for the second-order model. He states, “If metric invariance of the first-order model is supported…it implies that covariances between the first-order factors are comparable. Therefore, loadings of the second-order factors are meaningfully compared over time” (Rudnev et al., 2018, p. 52). So, weak invariance of the second-order model follows same strategy as the first-order model and factor loadings of the first-order model factor are equated over time. Next, the second-order strong invariance model should be tested after first-order strong model invariance is supported. However, the process for second-order model strong invariance is slightly different than the first-order invariance model. Instead of equated intercepts like first-order models, the first-order latent constructs mean should be equated across time. Rudnev et al.’s (2018) rationale is that equated first-order latent construct means rather than first-order latent intercepts allow the constructs to be compared meaningfully. Additionally, by constraining the means “is preferrable and more convenient to implement, because its indicators [the first-order factors] are latent variables themselves who means may be of interest…” (Rudnev et al., 2018). The below sections will critique the complexities of specifying, identifying, and testing measurement models.

Testing Measurement Invariance Models for Data Fit

Longitudinal Null Model

The default longitudinal null model in SEM is usually wrong (Little, 2013; Widaman & Thompson, 2003). Null model expectations of single-group and single time-points specify that all indicators have only a variance and all covariances are fixed to 0. The model gauges the fitting of the highly constrained model to the data to approximate how badly the model fits. “If covariances are generally small, the amount of information will be small and if covariances are large, the amount of information will be large” (Little, 2013, p. 112). However, in longitudinal models, multiple constructs and indicators are represented using multiple time points where the means, variances, and covariances are modeled.

The longitudinal model expectation for indicators is that indicators are repeated at more than one time point. So, the observed matrix will have the same things measured repeatedly. “A reasonable null expectation is that the variances of the like indicators have not change…so the model estimates potential changes in constructs that are repeated measured” (Little, 2013, p. 112). The null expectation for the means of like indicators should not vary across time points. This is central to the concept of factorial invariance, so this null expectation should be considered in evaluating the longitudinal null model (Little, 2013).

“The null model should be nested within the tested hypothesized model. A model, C for child, is nested when it can be derived from another model, P for parent” (Little, 2013, p. 113). “Two models are nested [when] the difference in chi squared (χ^2) between the two models is also distributed as a x^2 with degrees of freedom equal to difference in the degrees of freedom of the two nested models" (Little, 2013, p. 128). Also, models are considered nested when one model was derived by placing one or more constraints on the original model (Little, 2013). To obtain the traditional independence null model, constrain all the parameters, except the residual variances, to 0. That is, the covariances among the constructs would be 0, the variances of the constructs would be 0, and the loadings would be 0. For the longitudinal panel null model, it is nested within the strong factorial invariant model.

Specifically, regarding cross-time measurement, the longitudinal null model constrains the means and variances to be equal for all indicators across occasions. The difference between the traditional independence null model and the longitudinal null model is that the means and variances do not change between subsequent measurements in the longitudinal model. The evaluation of mean and variance stability is the central focus of measurement invariance testing.

Statistical Rationale and Modeling Rationale

Little (2013) classifies model fit based on two approaches: statistical rationale and modeling rationale. The first approach uses χ^2 to determine absolute fit to measure the difference between the implied model estimates and the observable data. The measurement issue that encompasses this test of absolute fit is highly sensitive to sample size and degrees of freedom. Conceptually, the x^2 test will produce a significant indication that the covariance matrix and mean vectors are not statistically equal to the observed matrix. More precisely, it implies that there is no difference in the observed matrix (S) and model implied matrix (Σ). Specifically, the x^2 test is testing exact fit in the population (S = Σ), which is contrary and flawed hypothesis testing (Little, 2013). Little

(2013) explains that since model are "gross approximations of the actual processes," the appropriate logic is a test of not-close fit (p. 108). The statistical rationale 2 has a preferred outcome of accepting the null outcome of exact as opposed to reject or fail to reject it. Even if p-values are adjusted for Type I and Type II error, this erroneous logic is still not feasible (Little, 2013).

A modeling rationale is more akin to recognizing that models are gross approximations. Relative model fit is examined instead of absolute or exact fit, akin to x^2 testing. Relative fit is essentially testing the amount of misfit per degrees of freedom in the model. The most popular relative fit measure is root mean square error of approximation (RMSEA), which was introduced by Steiger (1980). Although the RMSEA is classified in the absolute fit category, it can be used to determine a null test of acceptable fit and alternative test of not acceptable fit. The RMSEA uses the saturated model as the comparison model, and an index is determined based on the amount of misfit per degrees of freedom. The value, the non-centrality parameter, is "divided by sample size minus 1 [to] remove the effect of sample size on the χ^2…an additional correction factor for the number of groups misfit is then divided by the degrees of freedom of the model" (Little, 2013, pp. 108-109). Strong guidelines suggest null RSEMA should be specified no larger than 0.08, which is a test of acceptable fit and not acceptable fit, use a 90% confidence interval with the upper bound of its confidence not higher than 0.10 (Browne & Cudeck, 1992; Little, 2013). The main reason for the 90% confidence interval is "tests of model fit are one-tailed…therefore we want to lump all of the alpha Type I error rate on one side or the other, which is what a 90% confidence interval provides" (Little, 2013, p. 111).

Scale Setting

One of the first issues that gives rise to measuring MI is determining the appropriate scale setting method. The parameters in the model are estimated based on a pre-specified scale setting. These specifications are also referred to as constraints in the measurement model. There are three approaches used to identify the scale of measurement models: marker variable, fixed factor, and effects coding (Little, 2013; Rudnev et al., 2018). The marker variable approach is the most common method used. The first indicator of each construct is chosen as the marker variable with a set to a value equal to 1. The fixed factor approach relies on constraining the factor variance of each factor to 1. Effects coding sets the scale by setting the sum of all factor loadings equal.

It is common among SEM software programs to use the first indicator as a marker variable. Little (2013) has regarded this method as problematic because other parameter estimates are in relation to the marker variable. Because the parameter estimates rely on the selected marker variable, any estimates obtained are considered arbitrary (Little, 2013).

The fixed factor method uses the metrics of the latent variable. The variance of the latent variable is constrained with a positive non-zero value. It is common to use 1.0 because of the easy mathematical properties. The fixed factor approach assumes that the latent variables are equal across groups even when testing for measurement invariance. Additionally, the scale provided by the fixed factor method makes the scale of the indicators meaningless as it loses information about the scale on which indicators are measured (Little, 2013).

The effect coding method is one of the newer methods. Little (2013) recommends this method over the other two methods. Unlike the other two methods, the estimates are non-arbitrary and provide a real scale. For this reason, its utility of use has an advantage over the other two methods.

The three scaling methods are mathematically equivalent for determining model fit; however, choice of scaling method is not equal when determining partial invariance. In fact, effects coding method is least likely of the three to find non-invariant indicators. Little (2013) recommends switching to one of the other methods to find offending indicators. According to Little (2013), the marker variable method should be used only if indicators are invariant over time. He suggests using the fixed factor method, which equates factor variances across time and groups, to determine which indicators meet partial invariance. Fixed factor method is the recommended scaling method when determining invariant indicators (Little, 2013).

Identification

Identification refers to the “balance of known information available with the unknown parameters that are estimated from the data” (Little, 2013, p. 85). In SEM, the known information is the number of variances, covariances, and means. Identification involves both construct identification and model identification. They are both different.

In construct identification, after a scaling constraint is placed on one of the parameters, the construct is defined from parameter estimates. A construct is under-identified, over-identified, or just-identified. An over-identified construct has more known parameters than unknown. Just-identified constructs have equal known and unknown information. Under-identified construct does not have enough known information to estimate parameters. When a construct has three indicators, the solution is just-identified. After the scaling constraint, just-identified constructs have just as many freely estimated parameters as unique bits of information. Just-identified solutions do not add degrees of freedom to the overall model fit. However, when a construct is over-identified, there will be fewer freely estimated parameters than unique bits of information after the scaling constraint is placed, thereby adding degrees of freedom to the overall model fit degrees of freedom. Little (2013) asserts that within construct degrees of freedom can arbitrarily influence improvement in the overall model fit. To minimize this arbitrary improvement, he recommends using just identified constructs, if possible. However, “very reliable indicators with good scale qualities will provide precise estimates of sufficient statistics…power to detect a correlation of a given magnitude” (Little, 2013, p. 213). Additionally, Fabrigar et al. (1999) noted that over-identified factors with higher communality and larger samples sizes also lead to accurate factor loadings.

Power

Power is the ability to detect a parameter that is different from zero or an estimated parameter. Power also refers to the ability to determine whether a model is reasonable or ridiculous (Little, 2013). Power estimation depends on reliability of indicators, sample size, and number of model parameters (Little, 2013). Power is also largely based on degrees of freedom, a just identified construct model is recommended. In the context of testing MI, power is the ability to compare two models that are nested. As reviewed earlier, models in a sequence are nested when a set of parameter constraints are equated across time or groups in restricted model but not in the less restricted model.

Sample Size

SEM estimators are based on asymptotic theory, which means very large sample sizes are required. A key assumption in SEM estimators is multivariate normality and larger samples increase the likelihood that the data will be multivariate normal (Little, 2013). Little (2013) discourages using heuristics such as a 5:1 or 10:1 ratio of observations to parameter estimates when determining sample size for SEM. He considers a sample size of 100 adequate, as the determination of sample size should be based on a detectible effect size, Cohen's *d*. Other key factors include:

> heterogeneity and representativeness of the sample, the precision of the measures in terms of reliability and scaling, the convergent and discriminant validity of the indicators and the constructs, and model complexity (complex models typically have highly correlated parameter estimates, which makes estimating them harder with smaller sample sizes. (Little, 2013, p. 121)

Relative Fit Measures

A statistical rationale is well suited to test the differences between nested models but is not suited to evaluate model fit or approximations. Chi-square difference test compares the null expectation that the difference in two nested models is non-significant. Chi-square difference test is also sensitive to sample size (n > 300) and often to rejects nested model even when violations are minor (Cheung & Rensvold, 2002; Little, 2013). Invariance tests are approximate similarity test and are evaluating only trivial differences (Little, 2013).

It is recommended that model fit for invariance testing should use alternative criteria in conjunction with x^2 difference testing (Cheung & Rensvold, 2002; Little, 2013). The comparative fit index (CFI) is considered a well-performing measure. The CFI evaluates the ratio of model misfit. The Tucker-Lewis (TLI), also known as non-normed fit index (NNFI), was developed by Tucker and Lewis (1973), and it is also a ratio of the chi-square degree of freedom. Models with 0.90+ values for the CFI and the TLI/NNFI can be quite acceptable models and comparative fit index difference less than 0.01 implies the assumption of invariance is met. These fit indices are considered guidelines and not hard rules and will be used to determine MI testing in the study. The ABG change process and methods may offer more practical ways to explain communication intervention studies that have relied heavily on the student self-report PRCA-24 measurement of communication apprehension (CA) and prior inconsistent results. The below section will review CA, PRCA-24 (CA measurement), and CA interventions and the utility of using MI and ABG change processes in communication studies.

Research on Communication Apprehension

Communication Apprehension Construct

CA is most widely studied construct within the field of communication avoidance and “has been central to the study of communication avoidance since 1970” (McCroskey, 1984, p. 13). Communication scholars have given it substantial attention because it plays an important role in explaining why people avoid communication. Hancock et al. (2008), for example, note that:

> If an individual has a high level of CA, application of [communication] techniques will not result in improved communication performance.

> Consequently, for the effective development of communication skills it is necessary to diminish CA. (p. 93)

In the communication literature, CA is considered one of the most useful and practical constructs to measure students' communication competence and program evaluation (Hsu, 2009; Kelly & Keaten, 1992; Levine & McCroskey, 1990; Pribyl et al., 1998; Rubin et al., 1990).

The construct has been associated with many names: stage fright, audience sensitivity, social anxiety, unwillingness to communicate, reticence, and speech anxiety (McCroskey, 1982). Earlier development of the construct focused exclusively on oral communication as a broadly based anxiety in the areas of stage fright, shyness, and reticence (McCroskey, 1970). CA was later broadened to writing and singing apprehension (Andersen et al., 1978; Daly & Miller, 1975). McCroskey (1982) reconceptualized CA as "the fear or anxiety associated with real or anticipated communication with others" (p. 137). Currently, this remains the central conceptualization of CA.

CA is viewed as an affective measure that is best measured using a self-report instrument (McCroskey, 2009). While there may be some behaviors that can be attributed to CA, these behaviors are more internal than external (McCroskey et al., 2009). "Self-report, then, are the most appropriate when they are directed toward affect and/or perception in circumstances where respondents have no reason to fear negative consequences from a given answer…Self report measures are amenable to either trait or state concerns…" (McCroskey, 2009, p. 176). While researchers have investigated the state and trait-like orientations of the construct, state versus trait distinction of CA still poses conceptualization issues (Keaten & Kelly, 2009; Sawyer & Behnke, 2009; Wadleigh, 2009). Although addressing conceptualization issues regarding CA is not a goal of this study, psychometric soundness is at play when determining factorial invariance.

Originally, McCroskey (1977) gave no indication if CA was to be viewed as a trait or state of an individual. However, it was clear that the construct was “directed toward a response generalized across situations and time, [and] the measures advanced clearly focused on a trait-like pattern” (McCroskey et al., 2009, p. 103). A valid measure should show that traits of an individual are enduring and do not fluctuate from one time to another. Considering this, some argue if CA is viewed as a trait or biological temperament, it would not be amendable to change (Beatty et al., 1998). However, McCroskey (1978, 1984) counters and reports that a valid measure of CA is amenable to change from one time to another if an intervening variable such as an intervention is introduced (McCroskey, 1978, 1984).

Measuring Communication Apprehension

CA is primarily measured using McCroskey’s (1982) self-report PRCA-24. The PRCA-24 is the most used self-report measure of CA. The measure is a set of 24-items using a five-point Likert response that attempts to measure a second-order latent trait of CA in four generalized contexts: group, meeting, dyadic, and public (Levine & McCroskey, 1990; McCroskey, 1984). The measure consists of four factors composed of six items, three positively and three negatively worded to reduce response bias. In general, the PRCA-24 measures respondents’ feelings toward communicating in each distinct-like context. Scoring is done by summing context scores to obtain a global trait-like CA measure or by summing each context individually to obtain each context measure. Scores on the scale range from 24 to 120 for the global measure and 6 to 15 for context measures. Additionally, scores in the higher range (max) represent high communication apprehension (HGA) and scores that fall in the lower range (min) represent low communication apprehension (LGA). Cronbach's alpha reliability estimates range are high and range from 0.93 to 0.95 (McCroskey et al., 1985). The mean for the total score on the PRCA-24 is 65.48 with a standard deviation of 16.46. The PRCA-24 has been administered to roughly one half million students in over 100 colleges and universities (Rudnev et al., 2018). The content validity of the measure was substantiated as scores were highly correlated with another pre-dispositional measure for CA (McCroskey et al., 1985). However, in the study, group and meeting factors (r = 0.69) correlated higher than other correlations (rs = 0.40-0.64) between any other two pair of factors of the PRCA-24. Although Levine and McCroskey (1990) posited a second-order factor model, the previous referenced study results suggested a two-factor model. Additionally, cultural studies using students from other countries have also found different factor structures (Hsu, 2009; Levine & McCroskey, 1990; Pribyl et al., 1998).

McCroskey's (1970) earlier version of the PRCA, which was referred to as the PRCA-College, also produced mixed results. When the instrument was administered to two different mid-western colleges during a basic speech course, 542 students at one university versus 2,479 students at another, factor analysis and varimax rotation produced different solutions. The sample of 542 produced a two-factor solution, while the larger sample produced one factor solution for positive worded items and a factor for negative worded items. The results were interpreted "to be indicative of two response patterns relating to item wording rather than item content" (McCroskey, 1970, p. 274) and not multidimensional. It was concluded that the PRCA-College was unidimensional. The PRCA-24 was created in response to criticism about the PRCA-College.

From the initial design of the PRCA-24, McCroskey (1984) implicitly hypothesized a second-order factor structure as CA was believed to be a global response pattern of apprehension across contexts. However, the second-order factor model was never substantiated. As a result, just like the previous version, the PRCA-College, the PRCA-24 led to mixed results. Levine and McCroskey (1990) noted three common patterns:

> These include a four-factor solution with each factor corresponding to each subscale, a unidimensional solution, and a two-dimensional solution with the dyadic and the group items loading on one factor and the meeting and public speaking items loading on the other factor. Such mixed results led some to believe that the scale has an unstable factor structure, which would challenge the validity of the PRCA-24 and would lead [researchers] to question the results of prior research in which the scale was used. (p. 62)

In response, Levine and McCroskey (1990) tested three rival models to evaluate the second-order factor model for PRCA-24 using a linear unidimensional model, Guttman simplex, and a second-order factor model. The researchers used three different data sets, longitudinal, cross-sectional, and cross-cultural to show support for the second-order model. However, they did not find support for the theoretical second-order model using either sample but still posited a second-order model for both U.S. samples based on theory as the other rival models did not fit the data. Specifically, results indicated that the first-order factor structure was internally consistent and parallel, and context items were tapping a high order construct. The reliability of the first-order factor model was 0.87 for group, 0.89 for meeting, 0.86 for dyadic, and 0.86 for meeting. The second-order model reliability was 0.81. Internal consistency and parallelism were met for the first-order model but not the second-order model. For the test of internal consistency 9 of 60 deviations were greater or equal to 0.10. The magnitude of 0.10 was an arbitrary critical value set a priori in the study. “It was reasoned that deviations amounting to less than 1% of the variance would be considered trivial by most communication researchers” (Levine & McCroskey, 1990, p. 67). Levine and McCroskey (1990) determined that the offending deviations for the second-order factor model were few and the magnitude was small. However, for the first-order, 5 of 216 deviations were greater than or equal to 0.10, and they did not report the number of deviations for the second-order factor model. Their report concluded that the second-order factor structure was the most plausible model compared to the other models tested. It was also advisable in future studies to reduce the PRCA-24 to PRCA-20, and refrain from using items 1, 10, 17, and 24. Research may include the PRCA-24 in cross-cultural studies (Levine & McCroskey, 1990). Despite the warnings, the PRCA-24 remains the most used version in U.S. and cross-cultural studies (Hsu, 2007; Keaten & Kelly, 2009; Pribyl et al., 1998). The below sub-sections will review communication interventions using the PRCA-24 and how traditional classical analyses have led to mixed outcomes. The ABG change model will be explained as an alternative application in communication interventions using the PRCA-24.

Assessing the Quality of Communication Apprehension Interventions

The interventions used to reduce CA are systematic desensitization (SD), skills training (ST) (Fremouw & Zitter, 1978), visualization (VIS) (Ayres & Hopf, 1992), performance visualization (Ayres & Hopf, 1992), communication-orientation motivation or cognitive orientation modification (COM) treatment (Motley, 1991), and multidimensional model therapy (Dwyer, 2000). In VIS, individuals positively imagine themselves giving a speech after listening to a script. ST gives students specifics skills training to improve speech performance competency. In SD, students are exposed to a stressful and anxious stimulus while using deep muscle relaxation techniques. COM is one on one counseling sessions where each student is given a pamphlet and asked to think positively about communication and imagine it as common everyday speech. Multimodal uses various dimensions of student's anxiety to select the appropriate treatment. Except for ST, these communication interventions are considered cognitive-oriented treatments. Communication researchers have questioned the value of these interventions to reduce CA (Allen, 1989; Ayres et al., 2000; Beatty et al., 1998).

Intervention outcomes using the PRCA-24 have been mixed. Some researchers believe the results are mixed because of conceptualizations and measurement validity issues (Allen, 1989; Beatty et al., 1998; Frantz et al., 2005; Hsu, 2009; Kelly et al., 1990; Kelly & Keaten, 1992). Since the construct was introduced by McCroskey (1977), it still lacks a unifying conceptualization (Conduit, 2000). A primary concern regarding conceptualization is that CA is viewed as a trait or biological temperament that would not be amendable to change. However, McCroskey (1982) reported CA is amenable to change through an intervention. Nevertheless, the effectiveness of communication interventions to reduce CA have raised concerns that results were due to experimental artifacts, or demands, and not treatment effects (Hsu, 2009). For example, Hsu (2009) meta-analysis study on employing communication interventions on trait CA reported small effect sizes for single intervention treatments. In the study, three treatments, usually cognitive-oriented, were found the most effective (r = 0.55) compared to two treatments combined (r = 0.24 to 0.29) and skills training alone had the least effect size (r = 0.09). In fact, the effect size for skills training was considered no effect. Additionally, Allen (1989) and Duff et al. (2007) found regression to the mean issues when they compared measurement techniques in communication studies. As a result, effect sizes were overestimated, and treatment effect were inconsistent. One possible explanation offered by Allen (1989) is that self-report data overestimates.

All intervention research on the treatment of CA has relied on the self-report measure, PRCA-24 (Hsu, 2009). Allen's (1989) meta-analysis study compared measurement techniques in communication studies and found that self-report effect sizes were overestimated, and treatment effects were inconsistent. Regression to the mean may have also contributed to the different effect sizes because most intervention studies used students who were categorized as high communication apprehension (HCA). Often in communication studies, far fewer students were categorized as low communication apprehension (LCA). Students with low and normal levels of CA have not been represented. Also, past studies have primarily used non-diverse populations and the outcomes do not generalize to diverse populations or cultural studies (Hsu, 2009; Levine & McCroskey, 1990; Pribyl et al., 1998). Longitudinal designs in cultural communications studies have also reported different factor structures using the PRCA-24 as a repeated measure (Hsu, 2007; Levine & McCroskey, 1990; Pribyl et al., 1998**).** However, most communication research is conducted in the classroom settings, so results generalize to classrooms settings.

Hsu's (2009) meta-analysis study reported that most studies used pretest-posttest control group designs, and all used the within-subject design to observe changes from pretest to posttest after treatment. These types of designs highlighted validity concerns because students in the placebo groups were not led to believe that the placebo study would reduce CA. Additionally, most interventions happened after regular class, so demand effect might have occurred in treatment groups, and pretesting may have sensitized participants to receive treatment (Hsu, 2009). Moreover, successful treatment in past communication studies has solely relied on outcomes measured using traditional classical analyses. As reviewed earlier, traditional classical analyses produce outcomes based on if the treatment group that had a larger change score from pretest to posttest (alpha change) (Allen, 1989; Golembiewski et al., 1976; Hsu, 2009; Kelly & Keaten, 1992). Previous communication studies did not use the MI nor ABG change processes.

Applying the Alpha Beta Gamma Change Model to Communication Interventions

The inclusion of the ABG change analysis for assessing patterns of change in CA may provide practicality, clarity, and utility to explain validity issues associated with communication interventions. Additionally, exploring ABG change provides integrity of measurement to ensure items are still relevant and terminology is not outdated. The PRCA (McCroskey, 1977) was created in the1970s, primarily using a non-diverse and Generation X population. Today, communication contexts may have a different meaning for the current college population, which is a more diverse student body and a different generation.

CA reflects a cognitive change in individuals (McCroskey, 1977). These cognitive levels may manifest in various ways (Golembiewski et al., 1976). The dimensions of CA may result in different types of change. Alpha change has been the predominant type of change accessed by classical analysis procedures (Vandenberg & Self, 1993). It is change measured in a relatively fixed and stable system. As it pertains to evaluating an intervention seeking to decrease CA scores, a positively worded item with a Likert score of 4 (agree) that changes to 5 (strongly agree) means an actual improvement in CA scores. Although beta change is measured in a relatively stable dimension, it represents a metric change of the measurement intervals, and the respondent interprets the rating intervals differently across administrations. Specifically, an individual’s own yardstick for assessing and valuing pretest scores after gaining new experiences in CA may have changed within the conceptual framework at posttest (Karltun Erlandsson, 2006). The respondent’s interpretation of the response scale may differ between each occasion. For example, a pretest Likert rating of 4 of a positively worded item might be perceived as a Likert rating of 3 during posttest rating. The “intervals of the measurement continuum that are associated with a conceptual domain have been recalibrated” (Karltun Erlandsson, 2006, p. 2), and the pretest score of 4 and Likert posttest rating of 3 may conceptually hold the same value after the intervention. In this case, a traditional analysis will report a lowered score and may indicate that the intervention was ineffective. If beta change has occurred, a recalibration of Likert intervals then a comparison of pre-intervention versus post-intervention using traditional analysis techniques may result in erroneous conclusions. However, if beta change is considered in the analysis, the recalibrated scales might indicate a clearer communication reality by the respondents.

Gamma change is a total departure from beta and alpha change as it “involves the basic redefinition of the relevant psychological space” (Golembiewski et al., 1976, p. 138) because of the communication intervention. “It refers to a change from one state to another, as contrasted with a change of degree or condition within a given state” (Golembiewski et al., 1976, p. 138). More clearly, students might shift the way they understand CA after the intervention. Gamma change would mean instability of the construct and interpretation of preintervention versus postintervention scores would be meaningless as the instrument is no longer appropriate to use.

The ABG change model shows that change may not be static. In fact, the communication interventions that are considered the most effective may also implicitly suggest that an outcome of decreased CA may not be static also. As reviewed earlier, interventions that had the largest effect size in treating CA were cognitive-oriented treatments. Cognitive-oriented treatments work by “getting people to change their cognitions about communication or anxiety-eliciting stimuli…by redefining stimuli so that they are no longer seen as threats of punishment or reward cessation” (Keaten & Kelly, 2009, p. 52). So, an explicit expectation or outcome from these treatments is for students to shift their cognitive orientations toward communication. For example, individuals with HCA have negative experiences and expect punishment rather than reward (McCroskey, 1982). Cognitive-oriented interventions are focused on changing these negative expectations toward positive expectations. Specifically, if HCA students have initial negative perceptions toward communication before cognitive-oriented interventions, the expected outcome is reduced or changed negative perceptions. This change process could be perceived as changing from “one state to another, as contrasted with a change of degree or condition within a given state” (Golembiewski et al., 1976, p. 138). Thusly, these expected outcomes from cognitive-oriented interventions may implicitly induce a change in students’ perceptual understandings of CA after the treatment. Moreover, this change process over the duration of interventions using communication-oriented treatment may possibly induce gamma or beta change rather than commonly assumed alpha change. In this case, if the intervention using cognitive-oriented treatments have been effective then measurement that indicates gamma or beta change would be a positive outcome. Therefore, assuming stability of students’ pretest-posttest outcomes in communication interventions seems like misnomer. The ABG model of change seems practical and useful for explaining communication intervention outcomes. Thusly, this study may help clarify the conflicting or disappointing results (e.g., prior work has failed to improve communication skills among graduates). The aim of the proposed study is to illuminate the way forward—new, productive ideas to improving the design of university-wide intervention.

Hypotheses

The hypothesized second-order factor for CA is depicted in Figure 1 and will be specified in the following ways: (a) each indicator will have a non-zero loading on the first-order factors (group, meeting, interpersonal, public speaking) (b) all covariances between each first-order factors will be explained by communication apprehension, the second-order factor. Measurement invariance of the second-order model will be hierarchically tested using the following hypotheses.

H1. The same first-order factor structures will hold in both T1(T1) and T2 (T2). There will be no difference in the observed matrix (S) and the model-implied matrix (Σ). A four-factor model will adequately describe the before and after data for the sample.

H2. The first-order factor loadings are equal across both T1 and T2.

H3. The first-order factor loadings and second-order factor loadings are equal both T1 and T2.

H4. The first- and second-order model factor loadings and first-order model intercepts of indicators are equal across T1 and T2.

H5. The first- and second-order factor loadings, intercepts of first-order model and indicators and second-order factor latent means are equal across T1 and T2.

As a result of testing these hypotheses, we will assess whether a change occurred and what type of change occurred after a communication intervention that aimed to reduced students' communication apprehension. As reviewed, CA is considered one of the most useful and practical constructs to evaluate the effectiveness of communication intervention programs **(**Hsu, 2009; Kelly & Keaten, 1992; Levine & McCroskey, 1990; Pribyl et al., 1998; Rubin et al., 1990).

CHAPTER 2

METHOD

The study conducted a secondary analysis using archival data from a university-wide intervention program. As discussed earlier, traditional analyses rely on limiting assumptions and may reveal inaccurate results. Therefore, factorial invariance was examined through the evaluation of three hierarchical levels of MI: configural, weak and strong invariance. Additionally, the study analyzed change and determined applicability of the Golembiewski et al. (1975) model of change to the observed data.

Participants and Measure

The sample data were accessed from a self-study conducted at a public, mid-sized university assessing whether a college communication program decreased communication apprehension. The study used the PRCA-24 developed by McCroskey (1982) to determine pretest and posttest communication apprehension scores in students. The aim of the program was to decrease communication apprehension in students through planned interventions.

The intervention program was part of a university-wide strategic plan to address communication skill gaps in students. As part of the intervention, an interdisciplinary first-year basic communication course was implemented. The program goals were to decrease student communication apprehension in a variety of contexts particularly group, meeting, and interpersonal. While the course provided some content about at public speaking skills, public speaking was not a targeted context taught in the course. In the course, students were taught to organize and present ideas through group interactions and teambuilding. They were encouraged to develop critical thinking skills to make thoughtful decisions regarding communication media. Additionally, a communication center was implemented as part of the program to further address students' communication needs and provide a collaborative space for students to work. Students were required to meet in the communication center during the term with groups and/or get help with upcoming projects and speeches. Also, instructors rotated daily to provide coverage in the center to meet with students individually or groups to address communication needs. At the end the course, students deliver a small group project utilizing the tools and skills acquired through the course as part of their final grades.

The program also implemented a policy of common planning hours for all instructors in the program. Each week, the instructors met as a group to facilitate discussions about content, introduce new ideas and provide feedback to peers. The content and learning outcomes were the same across teachers; however, each teacher had discretion how this was managed. Some instructors may have required longer presentations, some may have required shorter, some may have required more presentations, and some may have required less. Nevertheless, the content and objectives were the same across teachers.

The PRCA-24 was administered to all sections of the communication course during all academic terms. There were approximately 24 sections with 30 students each for all terms except summer sessions (i.e., two sections with approximately 15-20 students each). These 24 sections also included honors sections. The PRCA-24 is a 24-item scale and responses to each item were on a 5-point Likert scale allowing participants to rate how well the statements apply to themselves (1 = *strongly disagree* and 5 = *strongly agree*). The course instructors assigned the pretest-posttest administration of the PRCA-24 to students using the university's learning module system (LMS) during the first week (i.e., pre) and last week (i.e., post) of class. Students were encouraged to complete the survey; it was not required.

Approval from the university's Institutional Review Board (IRB) was granted to use the archival sample. The researcher received a de-identified data sample drawn from 3 years of fall term pretest-posttest scores. Fall term data were requested because it best represents students who are first-term freshmen. Winter and spring terms data often has transfer or repeaters. The course program had designated sections for general, honors, and transfer sections; most of the sections were general.

Honors students completed an additional project and may demonstrate homogenous ability. Transfer students are students who have completed 30 hours or more of course work and thus would not represent first-term students. Data from honors and transfer sections were not retrieved for the analyses. The retrieved sample consisted of 1,445 students.

Previous studies have shown that the PRCA-24 factor structure does not hold over time for cultural studies and when an international sample is used (Hsiao, 2010; Levine & McCroskey, 1990). Therefore, approximately nine cases were removed who identified as an international student.

The PRCA-24 was completed online through the LMS. The default settings through the LMS gives students the ability to complete the test more than once during the pre-post wave. As discussed in Chapter 1, most communication interventions have relied on HCA students as the primary sample, and some researchers believe demand characteristics are at play when students participate in communication interventions. Some students may take the test more than once with the hopes to impress their teachers with low apprehension scores, or they may take it more than once with the hope to opt out of the communication course entirely if they score low apprehension scores during the pretest. Some students may not complete the test or take it more than once if their internet connections are lost. Since retest effects contribute to validity concerns in longitudinal studies, only students who took the test once at pretest and once at posttest will be retained in the data set for analysis. Approximately 70 cases were removed that had more than one pre and posttest scores. Other cases that were removed were students who did not provide any demographic information, non-traditional students (>25 years old), mislabeled ID numbers, and students who only completed posttest surveys. The resulting sample size was 1,332: 80% White, 10% African American, 54% Male, 92% Non-Hispanic, and 99% traditional students approximately 18-19 years old.

Data Analytic Approach

Power, Sample Size, and Degrees of Freedom

SEM estimations rely on asymptotic theory which requires very large sample sizes to satisfy the assumption of the maximum likelihood estimator (ML). The data set was a one-group convenience sample. Non-random convenience samples may require a larger sample size (Little, 2013). Previous investigators have proposed heuristics using a sample size-to-parameters ratio of 20:1, 5:1, and 10:1, with the first being the currently accepted norm (Brown, 2015; Little, 2013). Given there are 54 parameters in the proposed model, using the 20:1 ratio would require a sample size of 1,080. The available archival sample size is 1,332 and thus met the requirement with a sample size of 1,332.

The hypothesized lower-order measurement model for both pretest and posttest data is shown in Figure 1. Indicators grp1 through grp6 are hypothesized to measure the factor “Group,” indicators mtg1 through mtg6 measure the factor “Meeting,” indicators ip1 through ip6 measure the factor “Interpersonal,” and indicators ps1 through ps2 measure the factor “Public Speaking.” It is also hypothesized that the factors do co-vary, but indicators do not co-vary. Each indicator has a unique measurement error term, represented as u1 for indicator grp1 and thus forth for all other measurement terms. With p = 24 indicators, there are 300 observations available to estimate a total of 54 free parameters resulting in 246 degrees of freedom in the model.

The issue of model identification is a crucial one in any SEM modeling, including longitudinal SEM for measurement invariance. As Little (2013) states, “identification refers to the balance of known information available with the unknown parameters that are estimated from the data” (p. 85). Constructs are considered over-identified when there are “more known variances and covariances than parameter estimates” and constructs that have more than three indicators (Little, 2013, p. 85). He also points out that nearly all SEM models will be over-identified. The hypothesized second-order CA factor model is considered an over-identified model with over-identified constructs. As reviewed in Chapter 1, over-identified constructs are considered problematic as they may influence arbitrary model improvement. It is recommended to reduce the number of items by creating parcels to create just-identified construct (Little, 2013).

Parcels

Over-identified constructs add additional degrees of freedom after estimating model parameter and these additional degrees of freedom may arbitrarily influence the over a model fit (Little, 2013). As a rule of thumb, Little (2013) recommends using parcels to create a just-identified model. “Parceling reduces both the sampling variability of the selected sample and the amount of incorrectness of [a] model in the population” (Little, 2013, p. 24). More specifically, parceling produces few parameter estimates and reduces the likelihood of dual loadings and correlated residuals which can be simple sample fluctuations or populations misfit.

Analysis Software

Mplus7 was used to run the analysis. Mplus7 is a statistical modeling program that offers the most options for handling missing data. Mplus also has a wide selection of models, algorithms, and graphical displays of data and analysis. The IBM Statistical Package for Social Science (SPSS) 28.0 for Windows was used to prepare the data. SPSS was also used to format the data file required for the input in Mplus7.

Missing Data

Listwise deletion has the practical advantage of including only matched cases in the analysis. However, listwise deletion may cause an overestimation of parameter estimates, range restriction, and loss of power. As a result, using listwise deletion is strongly discouraged by many researchers (Enders, 2010; Little, 2013). Fortunately, the Mplus statistical software package is equipped to handle missing data and offers several viable options to handle missing data. Mplus does not support data imputation techniques but handles missing data in a general way using ML under a MCAR and MAR process, which was explained in Chapter 1.

Model Fit Test Indices

As reviewed in Chapter 1, chi-square tests as metrics of model fit are known to reject models even when violations are minor when sample sizes are large. The sample size in the present study is considered large, *n* = 1,332, which resulted in chi-square tests for all hypotheses to be significant. So, in addition to x^2 and the chi-square difference test (Δx^2) (likelihood ratio test), I used other model fit criteria (including absolute and comparative fit) to evaluate the reproduced variance-covariance matrix, model fit, and nested model deterioration. Model fit was evaluated using the Tucker-Lewis index (TLI) and the comparative fit index (CFI). TLI and CFI values ≥ 0.95 indicate acceptable fit while values ≥ 0.90 reflect good fit. I used root mean square error of approximation (RMSEA) to assess model fit. RMSEA values up to 0.05 indicate close fit while values between 0.06 and 0.08 provide acceptable fit (Little, 2013). Model deterioration was evaluated using the rationale suggested by Cheung and Rensvold (2002) for samples greater than 300. Their recommendation was that difference in models (ΔCFI values) should not be larger than 0.01 across models.

In addition to chi-square tests, model improvement was evaluated based on a change in CFI (ΔCFI). It was necessary for the criterion (i.e., difference of no larger than 0.01 across models) to be met to continue hypothesis testing of MI (Little, 2013; Rudnev et al., 2018). The criteria for ΔCFI (i.e., difference of no larger than 0.01 across models) was evaluated for Hypothesis 2 through Hypothesis 5 to determine if each subsequent hypothesis could be tested. Each previous hypothesis needed to pass the criteria to proceed to the next.

Scale Setting and Identification

The fixed factor method was used to set the scale. As discussed in Chapter 1, there are pitfalls associated with using the default scale setting in Mplus, the marker variable method. When using this method, the amount of reliable variance captured will be estimated based on the marker variable chosen (Brown, 2015; Little, 2013). Additionally, the marker variable method has set an a priori constraint that fixed loadings are invariant. The fixed factor method relies on the recognition that metric of the latent variable is arbitrary. As such, the latent variances were set equal to the 1.0, and the latent variable relations were estimated in a standardized metric for the baseline model.

Hypothesis Testing

The hypotheses were tested using the bottom-up stepwise procedure, proposed by Rudnev et al. (2018) and invariance testing proposed by Little (2013). Little (2013) and Brown (2015) both recommend testing second-order models using a bottom-up strategy in which first-order invariance testing should be upheld before proceeding to higher-order invariance testing in a stepwise manner.

Hypothesis 1

The same first-order factor structures will hold in both T1 and T2. There will be no difference in the observed matrix (S) and the model-implied matrix (Σ). A four-factor model will adequately describe the before and after data for the sample. All items are tested in Table 1.

Table 1

Indicators of the PRCA-24

Question, Indicator	Item	First-Order Factor	Second-Order Factor
Q1, grp1	I dislike participating in group discussions.		
Q2, grp2	Generally, I am comfortable while participating in group discussions.		
Q3, grp3	I am tense and nervous while participating in group discussions.		
Q4, grp4	I like to get involved in group discussions.	Group	CA
Q5,grp5	Engaging in a group discussion with new people makes me tense and nervous.		
Q6, grp6	I am calm and relaxed while participating in group discussions.		
Q7, mtg1	Generally, I am nervous when I have to participate in a meeting.		
Q8, mtg2	Usually, I am comfortable when I have to participate in a meeting.		
Q9, mtg3	I am very calm and relaxed when I am called upon to express an opinion at a meeting.	Meeting	CA
Q10,mtg4	I am afraid to express myself at meetings.		
Q11, mtg5	Communicating at meetings usually makes me uncomfortable.		
Q12, mtg6	I am very relaxed when answering questions at a meeting.		
Q13, ip1	While participating in a conversation with a new acquaintance, I feel very nervous.		
Q14, ip2	I have no fear of speaking up in conversations.	Interpersonal	CA
Q15, ip3	Ordinarily I am very tense and nervous in conversations.		
Q16, ip4	Ordinarily I am very calm and relaxed in conversations.		
Q17, ip5	While conversing with a new acquaintance, I feel very relaxed.		
Q18, ip6	I'm afraid to speak up in conversations.		
Q19, ps1	I have no fear of giving a speech.		
Q20, ps2	Certain parts of my body feel very tense and rigid while giving a speech.	Public Speaking	CA
Q21, ps3	I feel relaxed while giving a speech.		
Q22, ps4	My thoughts become confused and jumbled when I am giving a speech.		
Q23, ps5	I face the prospect of giving a speech with confidence.		
Q24, ps6	While giving a speech, I get so nervous I forget facts I really know.		

Note: Bracket symbol (]) denotes which indicators were averaged together to create parcels.

Table 2 shows guidelines for how each hypothesis was tested.

Table 2

Hypothesis Testing for MI and Parameter Constraints for One-Group Repeated Second-Order Measure

Hypothesis	Nested	First-order factors			Second-order factor		Outcome	
MI Level, Model Tested		Factor loadings	Item intercepts	Latent means/ intercepts	Factor loadings	Latent means	Model better? CFI ≥ .90; ΔCFI <0.01; χ^2 < p < .05; RMSEA <0.08	ABG Change
H1, Configural, Model 1	--------	Free but construct variances are fixed @1 T1&T2	Free, fixed to 0	Free	n/a	n/a	No, Hypothesis testing ends	Gamma and/or Beta
							Yes, continue to H2	Gamma and/or Beta
H2, First-order Weak, Model 2	Model 1	Constrained (T1 = T2), construct variances only fixed @T1	Free, fixed to 0	Free	n/a	n/a	No, Hypothesis testing ends	Gamma and/or Beta first-order
							Yes, continue to H3	Beta first-order
H3, First- order Strong Model 3	Model 2	Constrained (T1 = T2), construct variances only fixed @T1	Constraine d (T1 = T2)	Set to 0/Free	n/a	n/a	No, continue to H4	Gamma and/or Beta for second-order
							Yes, continue to H4	Alpha
H4, First and Second-order Weak, Model 4	Model 3	Constrained (T1 = T2), construct variances only fixed @T1	Constraine d (T1 = T2)	Set to 0/Free, variances of second ordetT1 set to 0	Constrained (T1 = T2), construct variances only fixed at T1	Set to zero	No, Hypothesis testing ends	Gamma and/or Beta for 2nd order
							Yes, continue to H5	Beta or Gamma change for 2nd-order
H5, First- and second-order Strong, Model 5	Model 4	Constrained (T1 = T2), construct variances only fixed @T1	Constraine d (T1 = T2)	Set to 0/Free, variances of second ordetT1 set to 0	Constrained (T1 = T2), construct variances only fixed at T1	Free	No, Hypothesis testing ends	Gamma or Beta change for second-order
							Yes data supports MI for second-order model	Alpha change for both second and first-order

Note. The models are based on the fixed factor method recommended by Little (2013). Procedures were followed from Rudnev et al. (2018).

Hypothesis 1 Rationale

The configural model was Model 1 in the analysis as it is the first model in the nested model sequence. Configural invariance is also considered level 0. Factor variances for all factors at each time point were set equal to zero to establish the scale (fixed factor method). The intercepts for all factors were also set equal to 0. No parameters were constrained, and all parameters were freely estimated. Additionally, latent means for the second-order model constructs “Group,” “Meeting,” “Interpersonal,” and “Public Speaking” were fixed to 0 at T1 and T2. Chi-square (x^2) criteria was used to evaluate observed and reproduced covariance models and should be nonsignificant to proceed with analyses. If this criterion is met, one may infer that the pattern of relationship across time for the factors “Group,” “Meeting,” “Interpersonal,” and “Public Speaking” are the same before and after the intervention. Specifically, every place there is an estimate at one-time point, there should be an estimate at the other; everywhere there is a 0 estimate at one-time point, the estimate should be 0 at the other time point. For example, if the indicator grp2 for the factor “Group” has the largest factor loading at T1 then it should also have the largest factor loading at T2. The pattern of loadings should be the same, but the value of the parameter estimates could differ. Similarly, if the latent construct “Group” had the highest mean at T1, then it should be the highest at T2. This phase of the analysis was more qualitative rather than quantitative (Little, 2013; Widaman & Reise, 1997).

As it pertains to the ABG model of change, I was assessing whether gamma and beta change exist. If invariance is not supported, then it is possible that both gamma and beta change have occurred. However, even if the data support invariance, neither gamma nor beta change may be ruled out because all parameters are freely estimated. Configural invariance does not require any parameter constraints over time; "therefore configural invariance may also be assessed with lenient methods…" (Rudnev et al., 2018. p. 50). Therefore, invariance was primarily evaluated using the criteria for CFI ≥ 0.95. If these criteria were not met, the model would not be considered invariant, Hypothesis 1 will be rejected, and no other hypotheses could be tested. Hypothesis 1 was supported based on model fit estimates reflecting good fit (x^2 = 601.394, CFI = 0.985, TLI = 0.980, RMSEA = 0.040).

Hypothesis 2

The first-order factor loadings are equal across both T1 and T2.

Hypothesis 2 Rationale

A first-order weak invariance model is referenced as Model 2 and is nested within Model 1. Weak invariance is also considered the second step and represents Level 1 in MI testing for second-order models.

> The hypothesis being tested [for weak invariance] is that the elements of λ, the factor loadings, are equal across measurements. Weak invariance testing was done by constraining factor loadings across time, additionally, the construct variance scale was removed. Since factor loadings "are the maximum likelihood estimates of the regressions of observed scores on true scores, the constraint of equality across time tests the equality of the scaling units. (Schmitt, 1982, p. 350)

Factor loadings represent the strength of the linear relation between each factor and its associated items (Bollen, 1989; Little, 2013). This determines if the latent constructs could be compared over time. When the loadings of each item on the underlying latent construct are equal at both T1 and T2, the unit of the measurement of the underlying factor is identical.

Of importance, weak invariance testing does not require that the scales of the factors have a common origin; this was determined in strong factorial invariance testing (Chen et al., 2005; Little, 2013). Even if measurement invariance is upheld for the weak invariance model, beta change could not be ruled out. However, if the data support invariance for this model, relations between the factor and other external variables may be compared across time because one unit of change at one time would be equal to one unit of change in another (Chen et al., 2005; Little, 2013; Widaman & Reise, 1997). Weak invariance tests reveal the equality of scaling units in the factors, not a single item. Thus, the multi-item measurement of constructs is important (Schmitt, 1982). However, the latent means of the scale still should not be compared across groups, as the origin of the scale may differ, and beta change has not been ruled out. Lack of beta change also means that the variances of the constructs did not change because of the intervention.

The criteria used evaluate weak invariance were chi-square difference (Δx^2) and Δ CFI (i.e., difference of no larger than 0.01 across model) test. First-order weak invariance is a prerequisite for second-order weak invariance testing. Hypothesis 2 was supported, and gamma and beta were possibilities. The analyses proceeded with Hypothesis 3.

Hypothesis 3

The first-order factor loadings and second-order factor loadings are equal at both T1 and T2.

Hypothesis 3 Rationale

The second-order weak invariance model was Model 3 and nested within Model 2. This was considered the third step and represents Level 2 in MI testing for second-order models. The rationale for the second-order model's weak invariance testing is the same as first-order model as explained in Hypothesis 2. The first-order models factor loading remains constrained and second-order factor loading for T1 was constrained to be equal to second-order factor loadings at T2. The data support the model; therefore, the analysis proceeded to Hypothesis 4.

Hypothesis 4

The first- and second-order model factor loadings and first-order model intercepts of indicators are equal across T1 and T2.

Hypothesis 4 Rationale

First-order weak invariance is a prerequisite for first-order strong invariance (Model 4, level 3). Model 4 was nested within Model 3. An additional was placed on the first-order factors, and their intercepts were fixed to 0 and constrained to be equal across time. The data supported strong first-order invariance and the first-order model was invariant over time. As a result, gamma and beta change were not ruled out.

Hypothesis 5

The first-order and second-order factor loadings, intercepts of first-order model, and indicators and second-order factor latent intercepts are equal across T1 and T2.

Hypothesis 5 Rationale

Model 5 (level 4) is nested within Model 4, and second-order latent intercepts were set to 0 in T1 or T2 and equated over time. The means for the latent first-order and second-order model were freely estimated. Invariance was supported for the second-order model. This pattern of results may be interpreted as the means can be compared over time and gamma and beta change can be ruled out. The next section discusses the results in more detail.

CHAPTER 3

RESULTS

Pre-Analysis

The raw data were used as input in Mplus and was prepared using IBM SPSS. A Missing Values Analysis performed in SPSS indicated that Little's (1988) test of Missing Completely at Random (MCAR) was not significant, x^2 = 29.109, df = 24, p = 0.216. By failing to reject the null, it was concluded the data were MCAR. As discussed earlier in Chapter 2, FIML estimator used in Mplus, relies on the assumption that that data are MCAR and MAR to recover information lost from missing data. The sample size was 1,332 with approximately 24% of the missing data. Despite the number of missing data points, inferences were based on a large sample (i.e., 75% of 1,332, n = 990). Therefore, the information in the data may be considered reliable (Little, 2013). All missing values were assigned a code of (-99) to indicate a missing value for use in Mplus.

All negatively worded indicators were reversed coded so that a high score on each indicator had the same directional interpretation. Each first-order factor (Group, Meeting, Interpersonal, Public Speaking) had three negatively worded items that were reverse coded. The items recoded for each factor were: Group items 1, 3, and 5; Meeting items 1, 4, and 5; Interpersonal items 1, 3, and 6; Public Speaking items 2, 4, and 6.

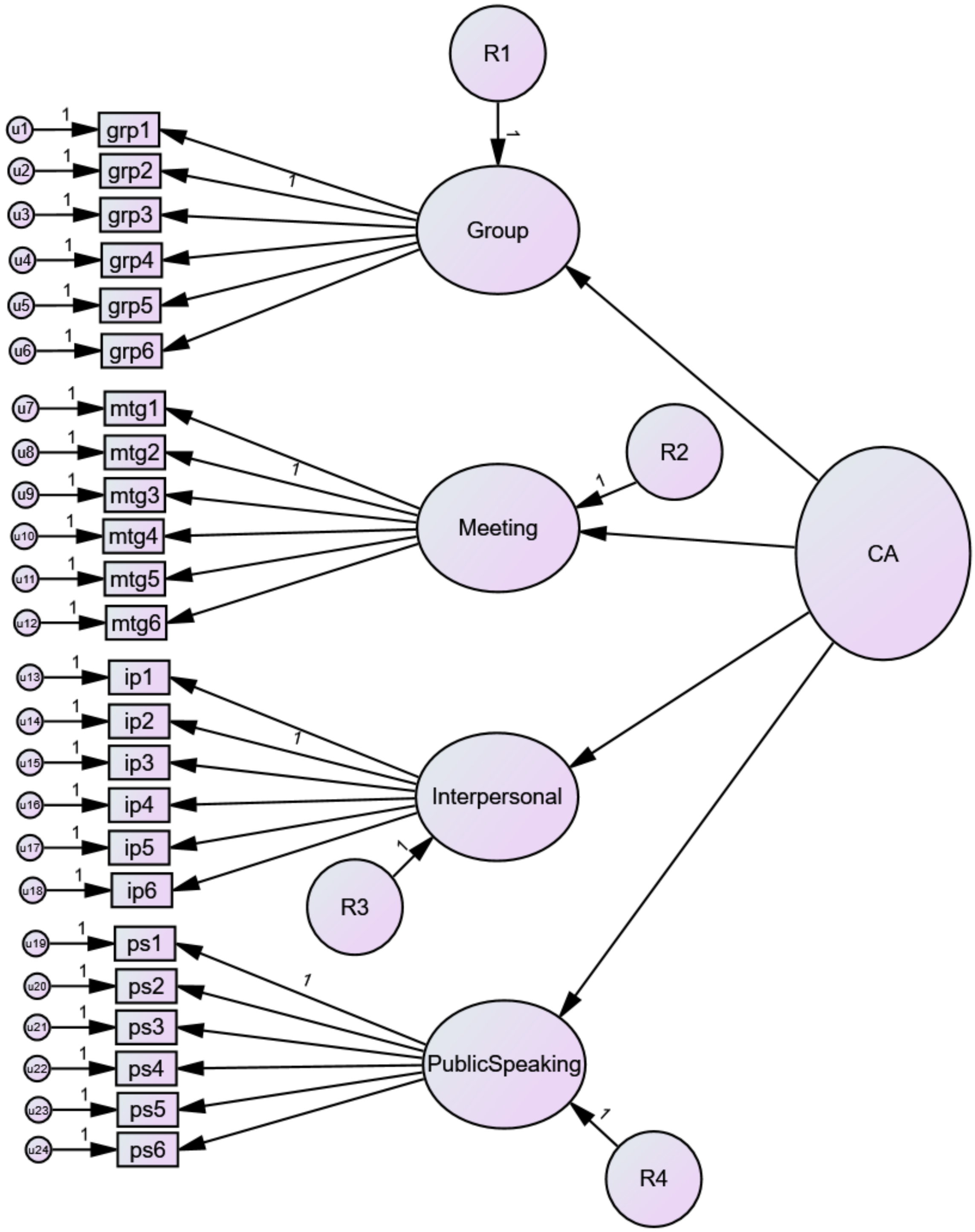
All scores were examined in Mplus for kurtosis and skewness. None of the individual variables had kurtosis and skewness values greater than 3.0. Each subscale factor was also tested for internal consistency using Cronbach's alpha. Group pretest ($\alpha = 0.87$), Meeting ($\alpha = 0.90$), Interpersonal ($\alpha = 0.89$), and Public Speaking ($\alpha = 0.89$). Cronbach's alphas for the posttest subscale (Group, Meeting, Interpersonal, Meeting) were 0.86, 0.91, 0.89, and 0.90, respectively. The Cronbach's alpha for both the pretest and posttest are considered high, and it was concluded that the scale was internally consistent (Lance et al., 2006; Nunnaly & Bernstein, 1994).

A Case for Parcels

As discussed earlier, the CA measurement model has over-identified factors (Figure 1). The model contains 24 items with each of the four lower-ordered factors consisting of six indicators (Figure 1). Over-identified factors are latent constructs with more than three indicators. A factor with more than three indicators may add additional degrees of freedom to the model's overall degrees of freedom. These additional within-construct degrees of freedom may randomly influence the overall model fit data (Little, 2013). To avoid these possible arbitrary influences, Little (2013) recommends reducing over-identified constructs in the model to just-identified constructs by creating parcels. Parcels are created by taking the average scores of two indicators to reduce the number of items per construct to three. By doing so, this creates a just-identified measurement model. The degrees of freedom in the just-identified models are "produced by between-construct relationships because the within-construct relationships are reproduced by the saturated parameter space associated with each construct's measurement element" (Little, 2013, p. 90).

Figure 1

Over-Identified Second-Order Measurement Model No Parcels



However, researchers who do not support using parcels insist that modeled data should be very close to the individual responses and parcels should not be used (Little et al., 2013). Considering this argument, the theorized over-identified CA measurement model with six indicators per factor was first tested for invariance without using parcels to determine if the baseline (configural) model could produce good model estimates to continue invariance testing. If the measurement model did not produce good fit estimates, then parcels would be used to improve model fit estimates as recommended by Little (2013).

The fixed factor method of scaling was used to set the scale for the over-identified model, and all factor variances for the first-order factors (Group, Meeting, Interpersonal, Public Speaking) were set equal to 1.0 for both T1 and T2. The model fit estimates for the baseline measurement model indicated that it was not a good fitting model (x^2 = 5869.41, CFI = 0.881, TLI = 0.869, RMSEA = 0.059). Moreover, there were also eight high modification values above the critical value of 59. The critical value of 59 was determined by using Little (2013) heuristics guidelines for determining high modification indices. The guidelines recommend researchers use 10% of the model’s degrees of freedom as a threshold to set critical values for modification indices such values are not available from prior published research. This heuristic guideline was used throughout the analysis to determine high modification indices.

Since the over-identified measurement model invariance testing produced poor model fit estimates, parcels were used to create a just-identified model to continue invariance testing.

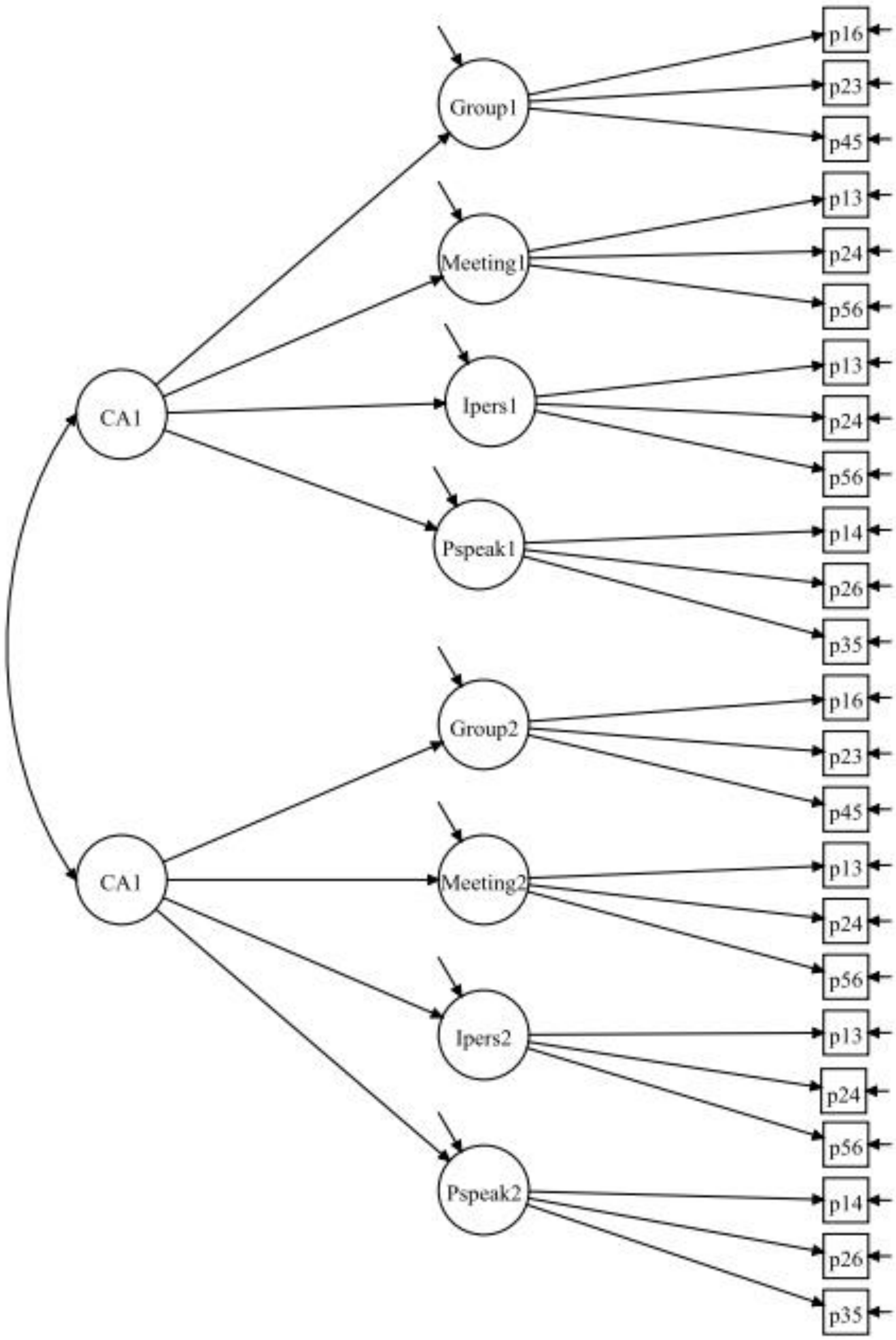
Parcels Improve Baseline Model

The configural model should be the best-fitting model specified in the longitudinal analysis as it is the baseline model that allows all possible latent correlations, loadings, intercept, and residual relationships to be estimated (Little, 2013). If the model does not show or suggest acceptable fit, adjustments should be made on this model before any additional invariance testing. Since the over-identified configural model did not produce good fit estimates, and all items in the measure showed high internal consistency with high Cronbach's alpha, parcels were used to improve model fit estimates.

Parcels were created by taking the average of two of the six items to create three parcels for each subscale factor. Indicators were paired by selecting the highest item-scale correlation to lowest item-scale correlation and then the next highest and next lowest item scale correlation. The process continued until all items had been parceled and each construct had three parcels. For example, values for item-total correlation for Group1 at T1 (grp1, grp2, grp3, grp4, grp5, and grp6) were 0.529, 0.686, 0.699, 0.672, 0.711, and 0.722, respectively. Items grp1 and grp6 were averaged to create a parcel called gp16 (g denotes Group, p denotes parcel, and 16 represent the two items averaged for parcels (Table 1). Then the next two highest (grp3) and lowest (grp2) were averaged to create the next parcel, gp23, for Group1 and continued in this order until all factors had parcels. The selection of item parcels was the same for both timepoints for all factors. As a result, the measurement model with parcels had a reduced number of items (three per factor) and 12 per time point (Figure 2) when compared to the model without parcels (six per factor) and 24 per time point (Figure 1). As discussed earlier, just-identified constructs provide less arbitrary influences from within-construct degrees of freedom.

Figure 2

Just-Identified Longitudinal Second-Order Measurement Model with Parcels



Lower-Order Factor Invariance Testing

The first step in the invariance analysis was the specification of the longitudinal null model. The null expectation is no changes in means or variances of the constructs over time. The variances of the indicators are equal across time, but no associations are estimated among them. Table 3 shows the descriptive statistics for the parceled items, which are the means, standard deviations, and correlations.

Hypothesis 1, level 0, invariance testing began with the configural model (Table 2). Fixed factor method was used to set the scale. All four of the first-order factors (Group, Meeting, Interpersonal, Public Speaking) variances were set equal to 1.0 and equated across time. This model allowed latent construct to be correlated, freely estimated loadings, and intercepts at both time points and did not restrict the pattern of residual relationships. The criterion for configural invariance is that all relations between each indicator and latent construct should demonstrate the same pattern across time.

Table 3

Correlations, Means, and Standard Deviations Among Parceled Indicators for First-order Factors

Parcel	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1 pGRP161	3.44	0.95																								
2 pGRP231	3.44	0.95	.71**																							
3 pGRP451	3.44	0.95	.74**	.75**																						
4 pMTG131	3.25	0.97	.59**	.61**	.66**																					
5 pMTG241	3.25	0.97	.63**	.61**	.68**	.81**																				
6 pMTG561	3.25	0.97	.63**	.63**	.69**	.77**	.78**																			
7 pIP131	3.41	0.95	.54**	.56**	.62**	.58**	.62**	.64**																		
8 pIP2412	3.41	0.95	.57**	.58**	.64**	.64**	.66**	.69**	.73**																	
9 pIP561	3.41	0.95	.57**	.57**	.65**	.64**	.67**	.68**	.76**	.76**																
10 pPS141	2.84	1.05	.46**	.48**	.53**	.59**	.56**	.57**	.45**	.51**	.49**															
11 pPS261	2.84	1.05	.42**	.46**	.49**	.57**	.56**	.57**	.47**	.47**	.50**	.74**														
12 pPS351	2.84	1.05	.50**	.50**	.56**	.60**	.59**	.60**	.46**	.53**	.53**	.76**	.70**													
13 PGRP162	2.42	1.04	-.46**	-.45**	-.46**	-.35**	-.39**	-.40**	-.39**	-.39**	-.40**	-.25**	-.29**	-.29**												
14 pGRP232	2.42	1.04	-.44**	-.48**	-.48**	-.39**	-.42**	-.42**	-.42**	-.40**	-.42**	-.29**	-.32**	-.30**	.71**											
15 pGRP452	2.42	1.04	-.49**	-.50**	-.55**	-.42**	-.46**	-.47**	-.44**	-.45**	-.47**	-.30**	-.34**	-.37**	.74**	.73**										
16 pMTG132	2.36	1.08	-.45**	-.49**	-.49**	-.47**	-.48**	-.48**	-.43**	-.45**	-.46**	-.37**	-.39**	-.39**	.66**	.67**	.73**									
17 pMTG242	2.36	1.08	-.43**	-.45**	-.46**	-.43**	-.46**	-.45**	-.42**	-.43**	-.46**	-.33**	-.37**	-.36**	.68**	.67**	.72**	.87**								
18 pMTG562	2.36	1.08	-.42**	-.46**	-.47**	-.44**	-.44**	-.45**	-.41**	-.42**	-.44**	-.34**	-.36**	-.36**	.68**	.69**	.72**	.82**	.82**							
19 pIP132	2.42	1.04	-.42**	-.45**	-.47**	-.39**	-.43**	-.43**	-.54**	-.47**	-.51**	-.30**	-.32**	-.31**	.56**	.61**	.67**	.64**	.66**	.65**						
20 pIP242	2.42	1.04	-.41**	-.41**	-.42**	-.40**	-.44**	-.43**	-.44**	-.45**	-.47**	-.35**	-.33**	-.36**	.61**	.61**	.67**	.67**	.69**	.69**	.67**					
21 pIP562	2.42	1.04	-.43**	-.46**	-.49**	-.42**	-.45**	-.45**	-.48**	-.47**	-.51**	-.36**	-.34**	-.37**	.62**	.64**	.72**	.70**	.72**	.72**	.78**	.75**				
22 pPS142	2.97	1.13	-.38**	-.39**	-.41**	-.43**	-.45**	-.45**	-.37**	-.41**	-.41**	-.50**	-.46**	-.46**	.45**	.47**	.51**	.55**	.53**	.50**	.50**	.55**	.56**			
23 pPS262	2.97	1.13	-.36**	-.39**	-.39**	-.41**	-.43**	-.43**	-.37**	-.37**	-.38**	-.44**	-.49**	-.42**	.45**	.49**	.53**	.56**	.54**	.53**	.54**	.48**	.55**	.79**		
24 pPS352	2.97	1.13	-.38**	-.37**	-.39**	-.42**	-.42**	-.43**	-.33**	-.39**	-.39**	-.47**	-.41**	-.50**	.46**	.45**	.52**	.54**	.50**	.52**	.43**	.56**	.55**	.77**	.66**	

Both factors had only one factor loading that was the same across time (for Meeting p13; for Ipers p53). Only one intercept out of the three for the Ipers, Meeting, and Pspeak factors showed a similar pattern across time. Table 4 also has model fit statistics for all tests of invariance; the results of the longitudinal null test are shown in Table 4.

Table 4

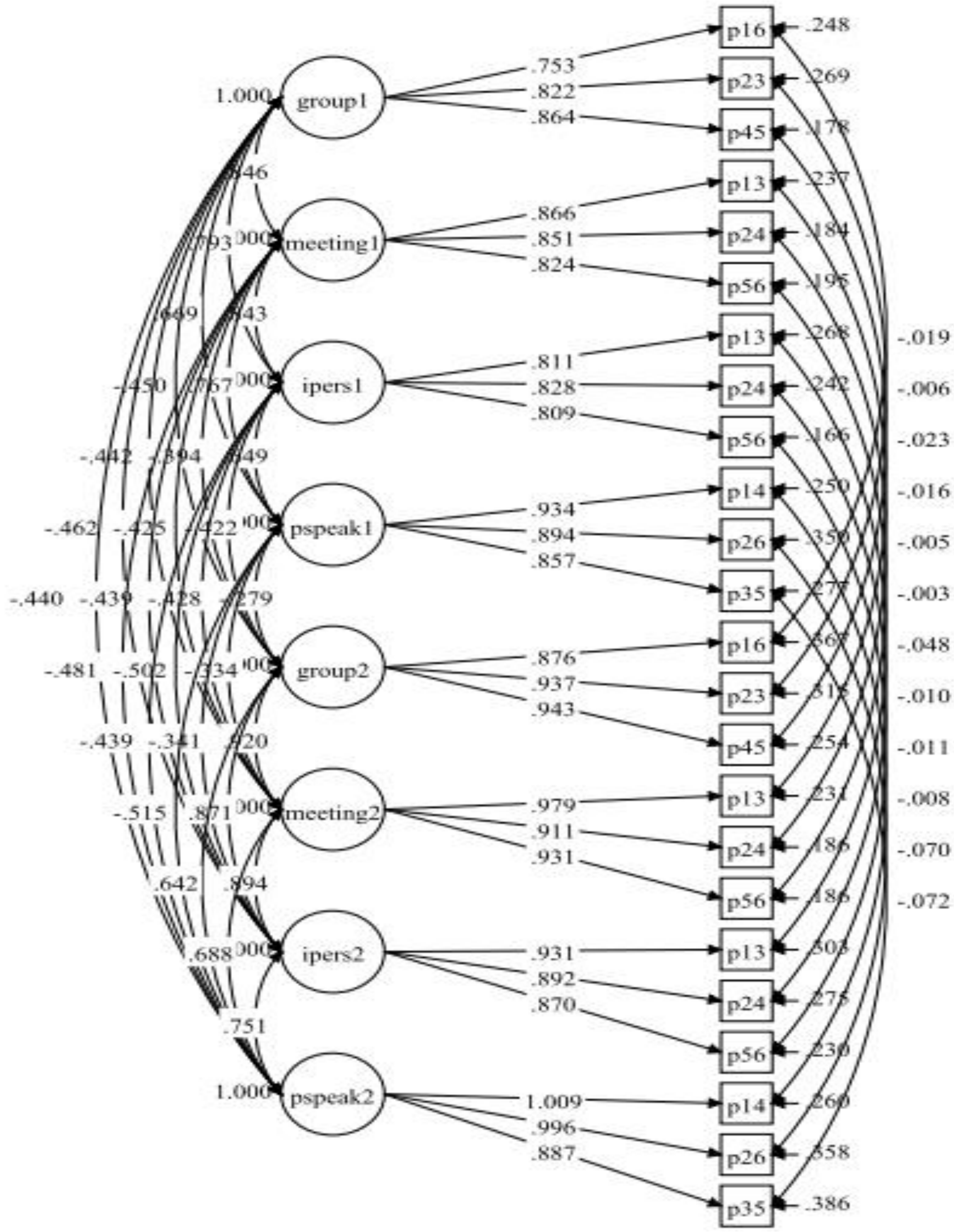
Model Fit Statistics for the Tests of Invariance for First-Order and Second-Order Factors for Communication Apprehension

Model testing	Chi-Square	*df*	*p*	RMSEA	RMSEA 90% CI	CFI	ΔCFI <0.01	Pass
Configural no parcels	5869.41	1028	<.001	0.059	1.158; 0.068	0.881	—	—
Null parcels	25,700.66	308	<.001	—	—	—	—	—
Configural parcels	601.39	212	<.001	0.040	0.036;0.044	0.985	—	—
Weak parcels first-order	616.96	220	<.001	0.039	0.036;0.043	0.984	0.001	yes
Strong parcels first-order	1163.24	228	<.001	0.059	0.056;0.063	0.963	0.021	no
Partial invariance strong	778.40	226	<.001	0.046	0.042;0.049	0.978	0.006	yes
Weak parcels second-order	932.62	242	<.001	0.050	0.046;0.053	0.973	0.005	yes
Strong parcels second-order	940.61	244	<.001	0.050	0.046;0.053	0.972	0.001	yes

All the parceled indicators for the Group and Public Speaking first-order factors showed the same pattern of factor loadings when each was compared with its complementary factor at both time points (Figure 3). For example, the highest factor loading for the Group factor was also the highest loading, and the highest factor loading for the Public Speaking (PSpeak) factor was also the highest at both time points. However, the pattern of factor loadings for Meeting and Interpersonal (Ipers) factors did not follow the same pattern across time for all indicators (Figure 3).

Figure 3

Configural (Baseline) First-Order Model



Note. All other models were nested within the baseline model. Parcels were used, and scale was set using fixed factor method. In the configural model, the association between constructs are correlational because the factor variances are set to 1, which produces standardized estimates. The label used for Mplus was longer and unique for each factor. For example, Group 1, first item was labeled pgrp161, and Group2 first was labeled pgrp162. See Table 4 for model fit estimates.

None of the indicator intercepts for the Group factor showed a similar pattern. Nevertheless, the model fit estimates did support good fit for the model (x^2 = 601.394, CFI = 0.985, TLI = 0.980, RMSEA = 0.040). The model fit estimates suggest a good model fit and modification indices do not approach 10% of the model overall x^2, it was determined that the configural model met the assumption of invariance. The inconsistent factor loadings and intercepts patterns are likely due to sampling variability and not model fit. The invariance testing continued to Hypothesis 2, weak invariance testing.

Hypothesis 2, level 1, tested the weak invariance assumption for the first-order factors. To set the scale, the construct variance constraint remained in place for all four factors at T1, but the factor variance constraints at T2 were relaxed. The factor loadings for all first-order factors were constrained equal with its corresponding construct across time and all other parameters were freely estimated at each time point (i.e., means, variances of constructs, residual variances). In contrast to the configural model, the weak invariant model constraint forced the loadings to be estimated as the optimal balance within and across time, thereby using information from both time points. Hypothesis 2 was supported, and the findings for first-order weak invariance testing were upheld (x^2 = 616.956, CFI = 0.984, TLI = 0.980, RMSEA = 0.039).

The next level of testing, Hypothesis 3, proceeded by testing the strong invariant model within the weak invariant model. The purpose of the model is to evaluate observed means and estimated intercepts of indicators. To set the scale for the fixed factor method, the means of the latent constructs at T1 were fixed to 0 to determine the latent intercepts and means at both time points. Also, the indicators intercepts were placed with a cross-time constraint. For this model, fit estimates did not support invariance. The model's CFI value of 0.021 was beyond the threshold of less than 0.01suggested by Little (2013). Table 4 has the model fit estimates for the strong invariant model. Since the model did not produce good model fit estimates and the findings did not support a full invariant model, modification indices were evaluated to determine if a model revision would support partial invariance.

To determine whether partial invariance is present, modification indices' critical values were evaluated based the guideline discussed earlier (i.e., a candidate for model revision is any parameter greater than 10% of models' chi-square estimates). Also, parcel invariance may only be considered if small number of indicators per construct deviate from others. Since there are only three parceled indicators per factor, only one indicator could have deviated from others to pass the test of partial invariance (Little, 2013).

The critical value was set at 116 for the strong invariant model and 4 out the 24 indicators were identified that exceeded the critical threshold value. One of three indicators from the group factor and one of three indicators from the meeting factor at both time points had high modification indices for their means and intercepts (Group, p45, T1 and T2; Meeting, p13, T1 and T2; values were 166.60, 181.99, 166.61, and 182.00, respectively). As a result, the mean and intercept constraints of these indicators were allowed to be freely estimated. The model fit estimates improved, and the partial strong invariant model fit estimates were (x^2 = 778.404, CFI = 0.978, TLI = 0.973, RMSEA = 0.046), with a change in CFI of 0.006, less than the threshold of 0.01. Table 4 has all model fit estimates.

Higher-Factor Invariance Testing

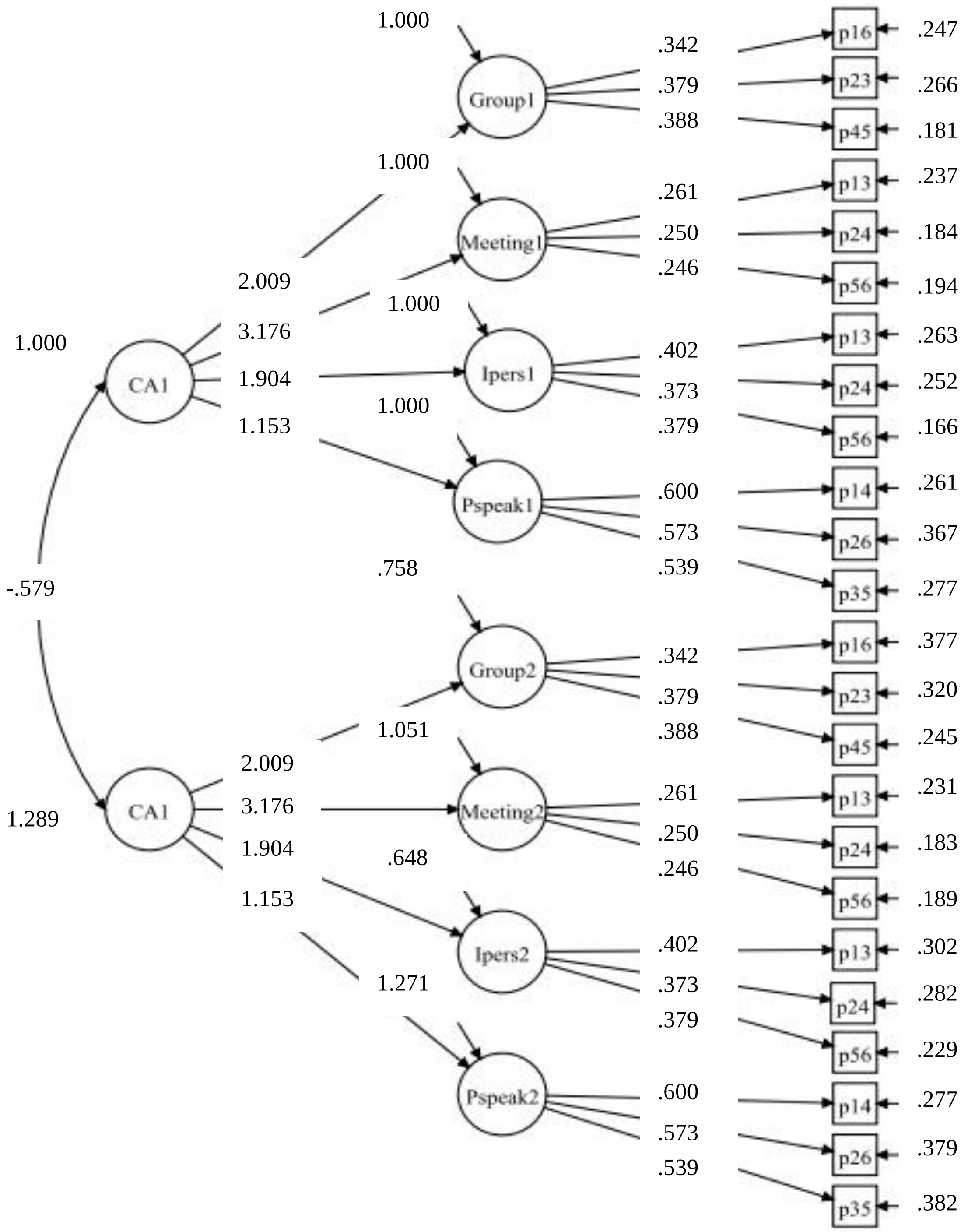
The higher-factor model should be built on the strong invariant version of the lower-order constructs and in the present study, the partial strong invariant model. Since all hypotheses were supported for the first-order model, second-ordered testing continued. The higher-factor model invariance testing, Hypothesis 4, began with testing the weak model within the first-order weak invariant model (see Table 2). In addition to the cross-time loading constraints already placed on the indicators in weak first-order model, factor loadings constraints were placed on the second-order construct across time. The weak invariance testing for the second-order model showed good model fit estimates (x^2 = 953.284, CFI = 0.972, TLI = 0.968, RMSEA = 0.050) (see Table 4).

The last level of invariance testing, Hypothesis 5, tested the second-factor model within the partial strong invariant lower-order model. The means of the first-order constructs were still being estimated, so the scale was set by setting CA at T1 to 0, constraining the lower-order means and higher-order regressions and estimating the mean of CA at T2. The final model shown in Figure 4 had good model fit estimates (x^2 = 940.612, CFI = 0.972, TLI = 0.969, RMSEA = 0.050).

Since partial strong invariance was supported by all five hypotheses, this would suggest that constructs are the same across time and the means can be compared. The means were compared using a global test of latent mean differences across time with the partial strong invariance testing, (x^2 = 940.612, CFI = 0.972, TLI = 0.969, RMSEA = .050). There was not a significant mean change overtime at the 0.05 level for CA.

Figure 4

Second-Order CA Partial Strong Invariant Model



Note. Parcels were used and scale was set using fixed factor method. The indicators' labels have been trimmed for demonstration purposes. The label used for Mplus was longer and unique for each factor. For example, Group 1, first item was labeled pgrp161, and Group2 first was labeled pgrp162. Table 4 shows model fit estimates.

CHAPTER 4

DISCUSSION

Principal Findings

Findings supported all five hypotheses that the CA measurement model is partially invariant. Accordingly, this suggests that participants' conceptual models of CA did not change from pretest to posttest, leaving open the possibility for alpha change as suggested by Golembiewski et al. (1976). When alpha change occurs, this indicates participants' fundamental understandings of CA from pretest to posttest remained stable and mean change across time is observed. However, partial invariance may also mean beta change may have occurred in the strong invariant first-order model. The strong invariant model fit estimates (x^2 = 1163.24, CFI = 0.962, TLI = 0.955, RMSEA = 0.059, ΔCFI = 0.021) indicated that some parceled items were not invariant. Specifically, evaluating modification indices for the model suggested that two parceled item-level intercepts for the Group and Meeting context factors (Group, p45, T1 and T2; Meeting, p13, T1 and T2) could not be constrained equal across time. Subsequently, relaxing the equality constraint across time and allowing the items parameters to be freely estimated improved the model fit estimates (x^2 = 778.404, CFI = 0.978, TLI = 0.973, RMSEA = 0.046, ΔCFI = 0.006). This improvement indicated that the parcels (Group, p45, T1 and T2; Meeting, p13, T1 and T2) were different from the other items for Group and Meeting factors. This is beta change.

A possible reason for beta change might be attributed to focus of the communication intervention program on developing skills that could have led to students to a new understanding of what represents high, medium, and low apprehension. Students participated in group meetings that may have changed their experience of communication apprehension and the scaling they used on the PRCA-24 instrument. Specifically, the program sought to improve and strengthen students' abilities to actively participate and engage in groups. Students were grouped with 4-5 other peers during the first week of class and assigned the task of creating a group presentation that required students to use standardized procedures when conducting their group meetings. The completed group projects were heavily weighted as part of the course grade. Students were required to meet often outside of class as a group throughout the quarter. During their meetings, students were required to record meeting notes using a standardized template created by the faculty. They were also required to create group charters with established boundaries. Considering the extent of required activities and importance of the group projects to their overall grades, students may have likely gained more knowledge and insight about their apprehension levels within the context group work and meetings. This increased knowledge may have had an impact on the responses for the non-invariant parcels. For example, the parcels (Group, p45; Meeting, p13) were composed of four item-level questions, *I like to get involved in group discussions*, *Engaging in a group discussion with new people makes me tense and nervous*, *Generally*, *I am nervous when I have to* *participate in a meeting*, and *I am very calm and relaxed when I am called upon to express an opinion at a meeting*, with intercepts that could not be constrained equal across time. In order to actively participate and earn high scores each student had to *get involved in group discussion, engage in group discussions with new people, participate in a meeting*, and *express their opinions at a meeting*, so the students may have changed their conceptual frameworks, thus changing their interpretations of their response scales for the items at posttest that led to beta change.

The results of invariance testing did not suggest gamma change occurred. Support for gamma change would mean instability of the constructs, which would infer the constructs could not be compared over time. Since all model fit estimates supported good model fit for all invariance testing, this indicates that the factors for the measure could be compared over time. Therefore, it is not likely that gamma change occurred as a result of the communication intervention.

Limitations and Future Directions

The present study is not without limitations, some of which arose from revisions to improve the measurement model while other limitations are common to any longitudinal design. The first limitation is that it is not clear whether the criteria used to determine whether measurement invariance models were supported by the data, such as exploring differences in CFI across models (Chen, 2007; Cheung & Rensvold, 2002), apply also for models testing for measurement invariance of second-order factors. There have been only two simulation studies conducted for determining good guidance on criteria for evaluating MI (Little, 2013). Of these two studies, neither was conducted using a second-order factor model. There is still a lot of work that needs to be done to improve guidelines for determining MI, particularly for second-order measurement models.

Second, the theorized CA model was over-identified, and model fit had to be improved by creating a just-identified model using parcels. In fact, the measurement model was supported for invariance using parcels and not supported for invariance without them. There is debate in the SEM literature regarding using parcels. Researchers opposed to parcels argue that parcels induce arbitrary manufacturing and create false structure within the model. As a result of these arbitrary influences parceled data is considered akin to cheating because data is based on averaged scores as opposed to raw individual scores (Little et al., 2013). Little (2013) offers a pragmatic view and suggests that parcels have greater reliability than items used to create them. Particularly, parcels will have more true score variances, stronger factor loading, and smaller unique variances. In the current study, the CA measurement model without parcels was over-identified and did not produce good model fit estimates (x^2 = 5869.408, CFI = 0.881, TLI = 0.869, RMSEA = 0.059). Parcels were used to create just-identified constructs to create a just-identified model to improve model fit. Future studies should evaluate reducing the number of items for each factor within the communication apprehension model.

Thirdly, the present study design relied solely on first-year students and a short (3 month) measurement occasion between pre-test and post-test. Implementing true longitudinal designs requires sufficient time to allow growth and change to emerge. Some developmental changes and growth arising from educational interventions may take many years to emerge (Little, 2013). Moreover, previous studies in communication research have revealed that CA increases during the first years in college and decreases during junior and senior years (McCroskey et al., 1989; Rubin et al., 1990). As a result of decreased CA, junior and senior students' communication competence increases because of the inverse relationship between CA and communication competence (McCroskey et al., 1989; Rubin et al., 1990). CA may be a lag experience. Future studies should investigate and compare CA scores between first year students and graduating students.

Additionally, the current study relied heavily on skills training techniques rather than other cognitive-oriented modifications interventions that directly address communication apprehension (i.e., visualization (VIS), cognitive orientation modification (COM)). A primary concern hinges on whether CA is viewed as a trait (a more fixed and enduring predisposition) or a state (amenable to change). Communication literature suggests that CA is amenable to change, and interventions that had the largest effect size in treating CA were cognitive-oriented treatments (Keaten & Kelly, 2009). Cognitive-oriented interventions are focused on changing negative expectations toward positive expectations. By design, communication-oriented treatments re-orient individuals toward positive communication experiences and may possibly induce gamma or beta change rather than proficiency (alpha change). This change process could be perceived as changing from "one state to another, as contrasted with a change of degree or condition within a given state" (Golembiewski et al., 1976, p. 138). The expected outcomes from cognitive-oriented interventions may implicitly induce a change in students perceptual understanding of CA after the treatment. In this case, if interventions using cognitive-oriented treatments have been effective then gamma or beta change would reflect positive outcomes of the process. Future designs that facilitate growth and improvement may benefit from an explicit focus that pursues gamma or beta change as a measure of effectiveness. One alternative method that seeks to capture participants' perceptions of change is the retrospective pretest-posttest (RPP) design. The RPP design ask respondents during the posttest to retrospectively respond to questions thinking back to a specified pretest period (Howard et al., 1979). After completing interventions, participants may increase their awareness and understandings of the construct. Moreover, the RPP design allows one to evaluate themselves more accurately during posttest than pretest. RPP designs can be used to assess program impact for both cognitive and non-cognitive constructs (i.e., skills trainings). The RPP design may be better suited design to capture change in communication intervention program evaluation.

The study design only investigated invariance overtime using one group. Future study directions should also investigate group-level invariance (i.e., culture, gender). Group-level factorial invariance is different from determining whether the latent constructs (i.e., covariances and mean levels) are the same; rather, it seeks to also determine group differences on the construct (i.e., reliable, and true properties of the construct). In other words, constructs can demonstrate different latent relations across subgroups, yet still be defined equivalently at the measurement level (Little et al., 2007). Gamma, beta, and alpha change may occur differentially across groups and these changes will be missed if designs rely solely on measuring invariance across-time.

Recommendations for Educational Leaders

Higher education leaders rely on data to make data informed decisions to meet the needs of their stakeholders. The utility of the tripartite model of change demonstrated the potential of how to assess change in a meaningful way to make informed decisions. For example, when leaders create pilot programs, particularly for QEPs, pilot data should be analyzed to account for ABG change before full program implementation to determine validity of instruments. In doing so, feasible instruments may provide valid and reliable data to determine program goals and effective improvements. For leaders to make data-driven decisions about university-wide programs, multilevel MI should also be investigated. This study provides a framework for testing MI on student learning using a basic foundational model across-time; however, leaders will need to investigate multi-level invariance to determine if identified measures are invariant across groups. In practicality, student learning is nested in groups (e.g., classrooms, programs, departments, university) and learning outcomes may be statistically dependent on different group complexities. Though MI was upheld in the current study, it may not have been the case if multilevel invariance was investigated.

QEP data is often collected from many sources. Those leaders central to the data collection processes should ensure that data collection procedures are standardized. Making informed decisions relies on having standardized data collection procedures. In conclusion, when institutions fully and validly assess the impact of interventions, the feedback loop between intervention, assessment, and redesign becomes stronger and holds the potential for greater insights and, ultimately, greater effectiveness.

Conclusion

The study highlighted that MI is a necessary first condition to determine meaningful comparison across time. In the context of communication, the present study contributed to a better understanding of both CA and the concept of change as a result of an intervention. The ABG change model was used to consider the multiple types of changes that can occur after a communication intervention. A participant may experience change in proficiency (alpha), a change in scale interpretation (beta), or a change in understanding the construct (gamma). Communication interventions that rely on skills training to decrease CA may primarily seek to change a student's proficiency (alpha), but measuring proficiency levels may not lead to improved levels, which are required for achieving communication competence. Communication interventions that are cognitive-oriented may facilitate a gamma or beta change in which participants might change their underlying understandings of CA. It is this change process that might lead to decreased levels of CA for improving communication competence. Designs such as RPP may yield better results for measuring change of communication interventions.

REFERENCES

Allen, M. (1989). A comparison of self-report, observer, and physiological assessments of public speaking anxiety reduction techniques using meta-analysis. *Communications Studies*, *40*(2), 127-139.

Andersen, P. A., Andersen, J. F., & Garrison, J. P. (1978). Singing apprehension and talking apprehension: The development of two constructs. *Sign Language Studies*, *19*(1), 155-186.

Ayres, J., Hopf, T., & Will, A. (2000). Are reductions in ca an experimental artifact? A Solomon four-group answer. *Communication Quarterly*, *48*(1), 19-26.

Ayres, J., & Hopf, T. S. (1992). Visualization: Reducing speech anxiety and enhancing performance. *Communication Reports*, *5*(1), 1-10.

Beatty, M., McCroskey, J. C., & Heisel, A. D. (1998). Communication apprehension as temperamental expression: A communibiological paradigm. *Communication Monographs, 65*, 197-219.

Becker, S., & Eckdom, L. R. V. (1980). That forgotten basic skill: Oral communication. *Journal of Speech Communication*, *12*, 1-18.

Benton, S. L., Duchon, D., & Pallett, W. H. (2013). Validity of student self-reported ratings of learning. *Assessment & Evaluation in Higher Education*, *38*(4), 377-388.

Bollen, K. A. (1989). *Structural equations with latent variables.* Wiley.

Boyas, E., Bryan, L., & Lee, T. (2012). Conditions affecting the usefulness of pre- and post-tests for assessment purposes. *Assessment & Evaluation in Higher Educations, 37*(4), 427-437.

Brown, T. (2015). *Confirmatory factor analysis for applied research* (2^{nd} ed). The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230-258.

Cavanagh, R., Kay, K., Klein, D., & Meisinger, S. (2006). Communication apprehension among business and accounting education Ireland. *Journal of Accounting Education*, *21*(3), 197-213.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal, 12*, 471-492.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Conduit, C. M. (2000). Culture and biology in human communication: Toward a multi-casual model. *Communication Education, 49*(1). 7-24.

Daly, J. A., & Miller, M. D. (1975). The empirical development of an instrument to measure writing apprehension. *Research in the Teaching of English, 9*, 242-249.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121-149.

Du-Babcock, B. (2006). Teaching business communication: Past, present, and future. *Journal of Business, 43*(3), 253-264.

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

Dwyer, K. K. (2000). The multidimensional model: Teaching students to self-manage high communication apprehension by self-selecting treatments. *Communication Education, 49*(1), 72-81.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Ericson, P. M., & Gardner, J. W. (1992). Two longitudinal studies of communication apprehension and its effects on college students' success. *Communication Quarterly, 40*(2), 127-137.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi:10.1037/1082–989X.4.3.272

Frantz, J., Marlow, A., & Wathen, J. (2005). Communication apprehension and its relationship to gender and college year. *Journal of Undergraduate Research at Minnesota State University, Mankato, 5*(7), 1-11.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763-779). Springer Science.

Fremouw, W. J., & Zitter, R. E. (1978). A comparison of skills training and cognitive restructuring-relaxation for the treatment of speech anxiety. *Behavior Therapy, 9*, 248-259.

Gaff, J. (1981). Reconstructing general education: Lessons from project GEM. *Change: The Magazine of Higher Learning, 13*(6), 52-58.

Gagliardi, J. S., Parnell, A., & Carpenter-Hubin, J. (2018). *The analytics revolution in higher education: Big data, organizational learning, and student success* (1st ed.). Stylus Publishing.

Garside, C. (2010). Seeing the forest through the trees: A challenge facing communication acrossthe curriculum programs. *Communication Education, 51*(1), 51-64.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Wadsworth/Thomson Learning.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, *12*(2), 133-157.

Hancock, A., Stone, M., Brundage, S., & Zeigler, M. (2008). Public speaking attitudes: Does curriculum make a difference*? Journal of Voice: Official Journal the Voice Foundation*, *24*(3), 302-307.

Hodis, G. M., 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.

Howard, G., & Dailey, P. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, *64*(2), 144-150.

Howard, G., Schmeck, R., & Bray, J. (1979). Internal invalidity in studies employing self-report instruments: A suggested remedy. *Journal of Educational Measurement*, *16*(2), 129-135.

Hsiao, T. (2010). Testing rival measurement models of the personal report of communication apprehension in two Taiwanese samples. *Psychological Reports, 107*(1), 127-138.

Hsu, C. (2007). A cross-cultural comparison of communication orientations between Americans and Taiwanese. *Communication Quarterly, 55*(3), 359-374.

Hsu, S. (2009). Treatment assessment of communication apprehension: A meta-analytic review. In J.A. Daly, J.C. McCroskey, J. Ayres, T. Hopf, D. Sonandre, & T. Wongprasert (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (3rd ed., pp. 171-192). Hampton Press, Inc.

Karakaya-Ozyer, K., & Aksu-Dunya, B. (2018). A review of structural equation modeling applications in Turkish educational science literature, 2010-2015. *International Journal of Research in Education and Science*, *4*(1), 279-291.

Karltun Erlandsson, A. (2006). Evaluation of intervention measures–a methodological approach. In *International Ergonomic Association 16th World Congress on Ergonomics, Maastricht* (pp. 11-14).

Keaten, J., & Kelly, L. (2009). Disposition versus situation: Neurocommunicology and the influence of trait apprehension versus situational factors on state public speaking anxiety. *Communication Research Reports, 21*(3), 273-283.

Kelly, L., Duran, R., & Stewart, J. (1990). Rhetoritherapy revisited: A test of its effectiveness as a treatment for communication problems. *Communication Education, 39*, 207-226.

Kelly, L., & Keaten, J. (1992). A test of the effectiveness of the reticence program at the Pennsylvania State University. *Communication Education, 41*, 361-374.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202-220.

Levine, R., & McCroskey, J. C. (1990). Measuring trait communication apprehension: A test of rival measurement models of the PRCA-24. *Communications Monographs, 57*, 62-72.

Little, R. J. (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. (2013). *Longitudinal Structural Equation Modeling*. The Guilford Press.

Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 121-147). Lawrence Erlbaum Associates Publishers.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285-300.

Lublin, J. R. (1980). Student self-assessment-a case study. *Assessment in Higher Education, 5*(3), 265-272.

McCroskey, J. C. (1970). Measures of communication bound anxiety. *Speech Monographs, 37*, 269-277.

McCroskey, J. C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communication, 4*, 78-96.

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

McCroskey, J. C. (1982). Oral communication: A reconceptualization. *Communication Yearbook*, *6*, 137-170.

McCroskey, J. C. (1984). The communication apprehension perspective. In J.A. Daly & J.C. McCroskey (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 13-338). Sage.

McCroskey, J. C. (2009). Self-report measurement. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, D. Sonandre, & T. Wongprasert (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (3rd ed., pp. 171-192). Hampton Press, Inc.

McCroskey, J. C., Beatty, M. J., Kearney, P., & Plax, T. G. (1985). The content validity of the PRCA-24 as a measure of communication apprehension across communication contexts *Communication Quarterly, 33*, 165-173.

McCroskey, J. C., Booth-Butterfield, S., & Payne, S. K. (1989). The impact of communication apprehension on college retention and success. *Communication Quarterly, 37*(2), 100-107.

McCroskey, J. C., Richmond, V., & McCroskey, L. (2009). Willingness to communicate communication apprehension, and self-perceived communication competence. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, D. Sonandre, & T. Wongprasert (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (3rd ed., pp. 97-128). Hampton Press, Inc.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, *73*(3), 574.

Mitchell, G., Skinner, L., & White, B. (2010). Essential soft skills for success in the twenty-first century workforce as perceived by business educators. *The Delta Pi Epsilon Journal*, *12*(1), 43-53.

Morreale, S., Backlund, P., & Sparks, L. (2014). Communication education and instructional communication: Genesis and evolution as fields of inquiry. *Communication Education*, *63*(4), 344-354.

Morreale, S., & Pearson, J. (2008). Why communication education is important: The centrality of the discipline in the 21st century. *Communication Education, 57*(2), 224-240.

Motley, M. T. (1991). Public speaking anxiety qua performance anxiety: A revised model and an alternative therapy. In M. Booth-Butterfield (Ed.), *The interrelationship of communication, cognition, and anxiety* (pp. 85-104). Sage.

Nunnaly, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Porter, S. (2011). Do college student surveys have any validity? *The Review of Higher Education*, *35*(1), 45-76.

Pribyl, C., Keaten, J., Sakamoto, M., & Koshikawa, T. (1998). Assessing the cross-cultural content validity of the personal report of communication apprehension scale (PRCA-24). *Japanese Psychological Research*, *40*(1), 47-53.

Rios, J., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-centuryskills for workplace success: A content analysis of job advertisements. *Educational Researcher, 49*(2), 80-89.

Rubin, R., Graham, E., & Mignerey, J. (1990). A longitudinal study of college students' communication competence. *Communication Education*, *39*, 1-14.

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

Rudnev, M., Lytkina, E. I., Davidov, E., Schmidt, P., & Zick, A. (2018). Testing measurement invariance for a second-order factor. A cross-national test of the alienation scale. *Structural Equation Modeling*, *12*(3), 471–492.

Southern Association of Colleges and Schools Commission on Colleges. (2020). *Resource manual for the principles of accreditation: Foundations for quality enhancement*. https://sacscoc.org/app/uploads/2019/08/2018-POA-Resource-Manual.pdf

Sawyer, C., & Behnke, R. (2009). Communication State Anxiety. In J. A. Daly, J. C. McCroskey, J., Ayres, T., Hopf, D., Sonandre, & T. Wongprasert (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (3rd ed., pp. 171-192). Hampton Press, Inc.

Schale, K. (1998). Advances in longitudinal research methodology. *Australian Journal on Aging*, *17*(1), 78-91.

Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research, 17*(3), 343–358.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized casual inference.* Houghton-Mifflin.

Shanahan, D. (2013). High oral communication apprehensives: How can students be helped to reduce their fear of public speaking? *Irish Journal of Academic Practice, 2*(9), 1-27.

Steiger, J. H. (1980). Statistically based tests for the number of common factors. In *The Annual Meeting of the Psychometric Society. Iowa City, IA.*

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1-10.

Vandenberg, R. J., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitments to the organizations during the first 6 months of work. *Journal of Applied Psychology*, *78*, 557-568.

Wadleigh, P. (2009). Contextualizing Communication Avoidance Research: Research, scope, realm and paradigm. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, D. Sonandre, & T. Wongprasert (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (3rd ed., pp. 171-192). Hampton Press, Inc.

Webber, K., & Zheng, H. Y. (2020). *Big data on campus: Data analytics and decision making in higher education*. Johns Hopkins University Press.

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). American Psychological Association.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*(1), 16-37.

Winsor, J. L., Curtis, D. B., & Stephens, R. D. (1997). National preferences in business and communications educations: A survey update. *Journal of the Association of Communication Administration, 3,* 170-179.

APPENDIX A

HUMAN USE EXEMPTION LETTER

LOUISIANA TECH
UNIVERSITY

OFFICE OF SPONSORED PROJECTS

EXEMPT MEMORANDUM

TO:	Dr. Mitzi Desselles and Ms. Tonya Calloway
FROM:	Dr. Richard Kordal, Director of Intellectual Property & Commercialization (OIPC) rkordal@latech.edu
SUBJECT:	HUMAN USE COMMITTEE REVIEW
DATE:	April 2, 2019
TITLE:	"Communication Apprehension of First Year Students: A Longitudinal Analysis of Confirmatory Factor Structure and Measurement Invariance"
NUMBER:	HUC 19-088

According to the Code of Federal Regulations Title 45 Part 46.104(d)(1), your research protocol is determined to be exempt from full review under the following exemption category(s):

(1) Research, conducted in established or commonly accepted educational settings, that specifically involves normal educational practices that are not likely to adversely impact students' opportunity to learn required educational content or the assessment of educators who provide instruction. This includes most research on regular and special education instructional strategies, and research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

Thank you for submitting your Human Use Proposal to Louisiana Tech's Institutional Review Board.