GREATER THAN THE SUM OF ITS PARTS:
ISOMORPHISM IN MEASURES OF TEAM CONSTRUCTS

by

Elizabeth Diane Clayton

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Organizational Science

Charlotte

2022

Approved by:

____
Dr. David Woehr

____
Dr. Janaki Gooty

____
Dr. Scott Tonidandel

____
Dr. Eric Heggestad

ABSTRACT

ELIZABETH DIANE CLAYTON. Greater Than the Sum of Its Parts: Isomorphism in Measures of Team Constructs (Under the direction of DR. DAVID WOEHR)

Work teams are an ever-growing structure as organizations seek to become more agile and achieve better outcomes (Bersin, 2016; Deloitte, 2018). Therefore, organizational researchers seek to accurately recognize and understand various aspects of team dynamics, which are often measured by capturing team-member perceptions. When these perceptions are shared among team members, team consensus constructs (e.g., team cohesion, conflict, psychological safety, satisfaction, task interdependence, liking, and viability) shed light on team functioning and performance. Researchers typically assess the psychometric properties of these measures at the individual level (e.g. factor analysis, covariance/variance matrices) without examining if the strength of and relationship among measures' indicators vary at the between-team level where the constructs theoretically operate (Carless & De Paola, 2000; Edmondson, 1999; Jehn & Mannix, 2001; Van der Vegt et al., 2001).

This misalignment between theory and measurement brings into question the quality of measures of team consensus constructs and the theoretical development based on the research associated with them. I examined the extent to which this misalignment is problematic and potential reasons for cross-level measurement and structural variance in and among measures. I used archival data to examine over 3,000 project-based teams using *R* and *MPlus* assessing measures in a multilevel factor analytic framework and examined for cross-level measurement and structural variance. The results demonstrated measurement quality should be assessed at the theoretically relevant level of analysis, the degree of psychometric isomorphism is in part a feature of within-team agreement and the wording of the measure, and there are consequences of misalignment regarding convergent and discriminant validity. Future research needs to address the need for discriminant validity among some measures and the potential for construct proliferation.

ACKNOWLEDGMENTS

My dissertation work and development as an academic is the result of the support and sharing of knowledge by talented faculty and colleagues who shaped me as a scholar, challenged me, and inspired me to pursue a career in academia. I would like to recognize and show my appreciation to my dissertation committee members. First, thank you to Dr. David Woehr, my committee chair. His passion for research is infectious. His patience is balanced with his ability to challenge and push me to grow as an academic. I am especially grateful for his generosity with his time helping me work through complex ideas. Additionally, I would like to thank Dr. Janaki Gooty who consistently encouraged me to develop my analytical skills, challenged me both in the classroom and during this dissertation process, and generously shared resources and opportunities with me. I am also deeply grateful for the guidance from Dr. Eric Heggestad whose scholarly curiosity and passion for measurement encouraged me to pursue this dissertation work. Finally, I would also like to thank Dr. Scott Tonidandel who patiently helped me learn how to manage and analyze data in *R* (which did not come naturally to me) and for his feedback focused on my future success.

Beyond my dissertation committee members, I would like to thank Dr. Amy Canevello who spent countless hours facilitating my development as a scholar. Her guidance has been pivotal to my success over the course of my time in graduate school. Her rigorous approach to answering the pressing questions of our time always inspires me to improve as a researcher. Thank you as well to Dr. Jaime Bochantin who served as my academic advisor early in graduate school and challenged my assumptions and approach to research, generously shared her knowledge and resources to develop my qualitative skills, and always took the time to encourage me.

Furthermore, I would like to recognize the Organizational Science Alumni Dissertation Fund for supporting my dissertation work by purchasing the necessary statistical software to run the analysis for my dissertation. .

DEDICATION

I am dedicating this dissertation to the very special group of people who have supported my 20-year dream of becoming a professor. First, thank you to my husband, Odise Adams Jr., whose sacrifices are too numerous to mention, and whose encouragement and support carried me through undergraduate and graduate school. His dedication to my future prompted him to find the University of North Carolina Charlotte Organizational Science program where I pursued my graduate education. Thank you, Odise, for moving across the country to help me achieve my dreams and raising three lovely children with me. To my children – Isabella, Senna, and Carver – whose smiles and zest for life keeps my spirit light, while I hope this process will serve as a reminder to always strive to reach your goals, our family is my greatest measure of success in life.

To the women in my life who have offered emotional and financial support, spent hours encouraging me, and helping me raise my children to the wonderful people they are today. First, Loretta Raiford, thank you for helping me return to college, reading my papers, and always believing in me. Sarah M. Lugo, for never doubting my ability to succeed and pushing me to pursue my passion for research. Vasilia Hilios Cottrell, thank you for all the emotional and financial support. You always provided a way forward for me to succeed. To Lee-Ann Marie Costa Gibson, for always showing up when I need you no matter the personal sacrifice. I would not be obtaining this doctorate without you. Rachel M. Koch, thank you for always helping me through your generosity in time, resources, and professional development. Robin and Jeff Smithberger, thank you for opening your home to my family during our move to Charlotte and continued friendship. Christine Lopez, your optimism, compassion, and kindness are

unparalleled. Thank you for always giving me a place of refuge and the support I needed to be the mother, wife, and scholar I am today. You ladies are my rock and motivation.

To my Aunt Diane and Uncle Robert Richey, who took me in as their own daughter and whose wisdom have helped me succeed in life and in school. To my late grandfather and grandmother, William (Pop Pop) Pascoe and Lucille Pascoe, who gave their family a hunger for knowledge and passion for service to the community. Pop Pop, you set the example to never stop learning and were smartest person I ever met who married the wittiest and kindest woman I have ever known. To my late parents, Janith and Jack Clayton, who encouraged spirited debates in our household, taught me knowledge is freedom, and believed in my ability to overcome adversity. To Todd Clifton who shared his home with my family so I could quit my job and gain research experience to become a better candidate for graduate school.

Much gratitude also goes to the Organizational Science Writing Group who have kept me focused and given me a place to discuss my ideas and challenges during graduate school. Specifically, my colleagues Dr. Lea Williams, Professor Melissa Medaugh, and Amber Davidson who began this writing group with me over three years ago. Thank you for always supporting my progress in graduate school.

Last but not least, I dedicate this dissertation to my cohort, the Salty Six, who have studied, laughed with me over the six years, and together mourned the loss of our beloved member, Jared Borns, may he rest in peace.

TABLE OF CONTENTS

LIST OF TABLES xiii

LIST OF FIGURES xv

INTRODUCTION 1

Part I: Psychometric Isomorphism in Measures of Team Constructs 5

1. Review of the Current Literature: Are the current practices evaluating the measures of common team constructs' psychometric properties sufficient? 5

2. Psychometric Properties Across Levels of Analysis: Are there major differences in the psychometric properties for these measures from the individual to the between-team level of analysis? 8

3. Examining for Cross-level Variation: Do the relationships among indicators (e.g., dimensionality) within these measures stay consistent from the individual to the team level of analysis? 14

Part II: Potential Reasons and Consequences for Varying Degrees of Psychometric Isomorphism 19

4. Measures' Characteristics: How does a measure's referent (e.g., "I" versus "team") and target (e.g., member–member relationship versus team-member?) relate to a measure's psychometric properties at different levels of analysis? 19

5. Team-member Agreement and Psychometric Isomorphism: To what extent does team-member agreement influence a measure's psychometric isomorphism? 25

6. Psychometric Isomorphism and Relationships Among Team Consensus Constructs: How do relationships among variables differ at various levels of analysis? 26

METHOD 30

Participants 30

Materials 31

Measures 31

Hypothesis Testing and Research Questions 33

Hypotheses 1-1e 33

Hypotheses 2-2b 35

Hypotheses 3-3b 36

Hypothesis 4 39

Research Questions 40

RESULTS 41

Sample and Participants 41

Part 1 Results: Examination of Psychometric Properties 42

Hypothesis 1: Metric Isomorphism and Measures of Common Team Constructs. 42

Hypothesis 2: Factor Structure of Multidimensional Measures 49

Part II Results: Influences of Psychometric Isomorphism and Relationships Among Team Constructs 53

Hypothesis 3: Psychometric Isomorphism and a Measure's Characteristics 53

Hypothesis 4: Psychometric Isomorphism and Team-member Agreement 56

Research Questions: Relationships Among Variables 58

DISCUSSION 62

Part 1 Discussion: Psychometric Isomorphism in Measures of Team Constructs 62

Part 2 Discussion: Reasons and Consequences for Varying Degrees of Psychometric Isomorphism 63

Theoretical Implications 65

Practical Implications 67

Limitations and Future Research 68

Conclusion 69

REFERENCES 71

TABLES 85

FIGURES 129

APPENDIX A 137

CFA & MCFA Equations 137

CFA & MCFA Results 140

APPENDIX B 149

MCFA Cohesion Syntax in R 149

Step 1: CFA of sample total covariance matrix at the individual level 149

Step 2: Estimate between-group level variation 151

Step 3: Within-group factor structure 151

Step 4: Between-group factor structure 157

Step 5: MCFA 159

MCFA Step 5 in MPlus 162

LIST OF TABLES

TABLE 1: Analogy Between Measurement Equivalence and Cross-level Isomorphism 85

TABLE 2: Model Fit Indices in a Multilevel Context 86

TABLE 3: Summary of Articles from Literature Review Examining Psychometric Properties of Team Consensus Measures 88

TABLE 4: Summary of Hypotheses & Research Questions 90

TABLE 5: List of Measures 92

TABLE 6: Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Measure of Team Cohesion 94

TABLE 7: Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Measure of Team Conflict 96

TABLE 8: Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Uni-dimensional Measures of Team Constructs 96

TABLE 9: Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Measures of Team-member Liking and Viability 97

TABLE 10: Standardized Factor Loadings for the Measure of Team Cohesion Across Level of Analysis & ICC(1) 98

TABLE 11: Residual Variance for the Measure of Team Cohesion Across Level of Analysis 99

TABLE 12: Standardized Factor Loadings for the Measure of Team Conflict Across Level of Analysis & ICC(1) 100

TABLE 13: Residual Variance for the Measure of Team Conflict Across Level of Analysis 101

TABLE 14: Standardized Factor Loadings and Residual Variance for the Measure of Team Psychological Safety Across Level of Analysis & ICC(1) 102

TABLE 15: Standardized Factor Loadings and Residual Variance for the Measure of Team Task Interdependence Across Level of Analysis & ICC(1) 103

TABLE 16: Standardized Factor Loadings and Residual Variance for Measure of Team Satisfaction and Task Interdependence Across Level of Analysis & ICC(1) 104

TABLE 17: Standardized Factor Loadings for the Measures of Team-member Viability and Liking Across Level of Analysis & ICC(1) 105

TABLE 18: Residual Variances for the Measures of Team-member Viability and Liking Across Level of Analysis 106

TABLE 19: Correlations and Standard Errors (*SE*) Among Latent Factors at The Individual Level Of Analysis 107

TABLE 20: Correlations Among Latent Factors at the Between-team Level of Analysis via S_B and (Multilevel Factor Model)) 107

TABLE 21: Correlations Among Latent Factors at the Within-team Level of Analysis via SW and (Multilevel Factor Model) 108

TABLE 22: Correlations and Standard Errors (*SE*) Among Latent Factors at The Between-Tea0 Level of Analysis Via S_B 108

TABLE 23: Inter-item Correlations Among Observed Variables at The Individual Level of Analysis 109

TABLE 24: Inter-item Correlations Among Latent Variables at the Between-team Level of Analysis via S_B 115

TABLE 25: Multigroup Analysis for the Measure of Team Cohesion via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median *(ADmd)* 121

TABLE 26: Multigroup Analysis for the Measure of Team Conflict via S_B: High & Low r_{WG} *(j)* & Average Deviation Median *(ADmd)* 122

TABLE 27: Multigroup Analysis for the Measure of Team Psychological Safety via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median *(ADmd)* 123

TABLE 28: Multigroup Analysis for the Measure of Team Satisfaction via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median *(ADmd)* 124

TABLE 29: Multigroup Analysis for the Measure of Team Task Interdependence via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median *(ADmd)* 125

TABLE 30: Multigroup Analysis for the Measure of Team Viability via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median *(ADmd)* 126

TABLE 31: Multigroup Analysis for the Measure of Team Liking via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median *(ADmd)* 127

TABLE 32: Participant Characteristics in Data Set by Measure 128

LIST OF FIGURES

FIGURE 1: Description of Level of Analysis and Associated Constructs: Individual, Within-, and Between-team 129

FIGURE 2: Description of Level of Analysis and Associated Constructs: Individual, Within-person, Between-person, and Between-team 130

FIGURE 3: Measurement Model for One-factor and Two-factor Solutions: One-level 131

FIGURE 4: Shared Cluster Construct and Two-level Factor Models 132

FIGURE 5: Multilevel Factor Model with Simplified Factor Structure 133

FIGURE 6: Shared Cluster Construct Model: Three Higher Level Factors 134

FIGURE 7: Multilevel Factor Model with Simplified Factor Structure & Cross-loading 135

FIGURE 8: Three-level Multifactor Model 136

INTRODUCTION

As teams are vital to an organization's ability to become more agile, enhance problem-solving, and incorporate diverse perspectives (Horwitz & Horwitz, 2007; Kozlowski & Bell, 2013; Somech & Drach-Zahavy, 2013), researchers seek to understand how team characteristics and dynamics influence team performance (Guzzo & Dickson, 1996; Jehn et al., 2008; Stewart, 2006). Team characteristics, attitudes, behaviors, and cognitions operate in a multilevel context in which researchers often collect individual team-member perceptions and then transform them (e.g., aggregate) to reflect team-level phenomena (Chan, 1998). Despite advances in multilevel research (Bliese, 2000; Cole et al., 2011; Morgeson & Hofmann, 1999; Vandenberg & Lance, 2000), a critical concern is the extent to which measures of team constructs accurately capture team phenomena (G. Chen et al., 2004).

This concern stems from the implicit assumption that the nature and structure of measures capturing team phenomena (i.e., constructs) remain consistent from the level of data collection (e.g., individual) to the level at which the constructs operate (e.g., team). For example, when common measures of team constructs (e.g., team cohesion, conflict, psychological safety, satisfaction, and task interdependence) are assessed at the individual level, the nature and structure of these measures are assumed to remain consistent at the between-team (i.e., team) level of analysis (Carless & De Paola, 2000; Edmondson, 1999; Jehn & Mannix, 2001; Loughry & Tosi, 2008; Van der Vegt et al., 2001). This is problematic because both the strength of a measure's indicators (e.g., items)[1] to detect a construct (i.e., latent factor) and the relationships

1 Survey items in a measure are often indicators of a latent factor (i.e., construct); thus, the terms are typically interchangeable. However, there are cases in which an indicator is not an item in a measure (e.g., an index, latent mean of an item). For example, the latent mean of an item can be an indicator when the construct resides at a higher level (e.g., team) than that of the of the data collection (i.e., individual; see Figure 4).

among indicators may vary across levels of analysis (Dedrick & Greenbaum, 2011; Huang et al., 2015; Whitton & Fletcher, 2014). Additional work needs to be done to understand the reasons for and potential consequences of such variation. Although Chen and colleagues (2004) detailed a systematic validation process for measures that use aggregated scores (e.g., consensus constructs), assessing the structure and nature of these measures at the level of the grouping factor (e.g., team membership) is a critical step in the validation process that must be considered.

While theoretically there is no construct at the individual level of analysis for team phenomena, measures of team consensus constructs are administered to individuals. Examining whether different conclusions can be drawn about the structure and nature of these measures at the individual versus the team level of analysis requires researchers to assess for isomorphism. In a multilevel context, *isomorphism* refers to the similarity in meaning, properties, and functionality of a construct (i.e., *latent variable*) across levels of analysis (Bliese et al., 2007; Zyphur et al., 2008). Researchers evaluate measures' isomorphism in terms of the degree of measurement and structural invariance (also referred to as c*ross-level measurement invariance* or *psycometric isomorphism* and *cross-level structural invariance*, respectively; Byrne et al., 1989; Zyphur et al., 2008).

Although common measures of team constructs are implicitly assumed to be psychometrically isomorphic, researchers should not ignore the potential implications of whether the relationships among team-related variables differ at the individual and the team levels of analysis. Therefore, I assessed for psychometric isomorphism in measures of team consensus constructs from the individual to the team level of analysis and for structural invariance across these measures. I also tested the assumption of cross-level consistency and demonstrated how to examine the nature and structure of these measures at the team level of analysis by applying measurement validation techniques designed for multilevel data.

The need to examine for psychometric isomorphism in measures of team consensus constructs is still in dispute. According to Chen and colleagues (2005), measures should only be evaluated at the level of analysis at which a construct operates. However, the nature and structure of measures of team consensus constructs are not typically examined at the team level (see Figure 3). Tay and colleagues (2014) argue that psychometric isomorphism from the lower (e.g., individual) to the higher (e.g., team) level of analysis is a necessary prerequisite for collective constructs, such as team consensus constructs– as it establishes similarity within a group. From either point of view, not knowing the nature or structure of a consensus measure designed to capture team phenomena is problematic.

In the current study, psychometric isomorphism is required across both levels of analysis for common measures of team consensus constructs to substantiate the implicit assumption that the meaning, properties, and functionality of these measures are consistent whether examined at the individual or the team level. While the need for psychometric isomorphism in measures of team consensus constructs more broadly has not been resolved, this study explores potential reasons and consequences of cross-level variation psychometrically and structurally via seven measures of team consensus constructs that operate in a multilevel context (i.e., team cohesion, conflict, psychological safety, task interdependence, satisfaction, liking, and viability).

First, I examined for psychometric isomorphism in common measures of team constructs by considering the first three questions: (1) Are the current practices evaluating the measures of common team constructs' psychometric properties sufficient? (2) Are there major differences in the psychometric properties for these measures from the individual to the team level of analysis?

and (3) Do the relationships among items (e.g., dimensionality) within these measures stay consistent from the individual to the team level of analysis? Answering these questions contributes to the broader literature on measuring team phenomena by describing best practices when assessing the measurement of team consensus constructs, providing practical examples for researchers to assess the psychometric properties of measures operating in a multilevel context in *R* and *MPlus*, and establishing norms for reporting and clarifying the psychometric properties of measures designed to capture higher-level constructs using data where observations are driven by a theoretically relevant grouping factor (e.g. team membership).

Second, I explored potential reasons and consequences for varying degrees of psychometric isomorphism from the individual to the team level of analysis for the seven measures of team consensus constructs by considering the next three questions: (4) How does a measure's referent (e.g., "I" versus "team") and target (e.g., member–member relationship versus team?) relate to a measure's psychometric properties at different levels of analysis? (5) To what extent does team-member agreement influence a measure's psychometric isomorphism? and (6) How do relationships among variables differ at various levels of analysis? This section links the construction of measures' items to differences in their psychometric properties across levels of analysis, investigates whether team-members' conceptualization of a construct is influenced by their shared perceptions of the teams' standing on a construct, and examines the potential influence of structural invariance across levels of analysis on the relationships among team constructs.

As the theoretical development of a construct and its measurement are inherently intertwined, I linked measurement theory and the theory of team consensus constructs by applying the multilevel factor analytic framework in a manner consistent with the development of the construct and its measurement. I also answered Chen and colleagues' (2005) call for greater clarity on modeling within and between levels of analysis across statistical packages via conducting analysis in both *R* and *MPlus*. As the ability to understand phenomena is limited by the ability to measure them, the current study contributes to the larger body of research on teams by more thoroughly scrutinizing measures of team consensus constructs that operate in a multilevel context.

Part I: Psychometric Isomorphism in Measures of Team Constructs

The accuracy of common measures of team constructs to capture team phenomena is vital to the development of theory, the confidence in conclusions drawn from research, and the real-world application of findings. I address the first question through a review of the current literature that reveals the extent to which the psychometric properties of common measures of team constructs are reported at the individual and the team levels of analysis. Regarding the second question, I discuss how to classify the degree of psychometric isomorphism across levels of analysis within a measure and propose tests for cross-level differences. For the third question, I review conflicting results in previous literature regarding the proposed dimensionality of the measures of team cohesion and conflict and propose how to assess theoretically relevant alternative measurement models for these measures at the team level of analysis.

1. Review of the Current Literature: Are the current practices evaluating the measures of common team constructs' psychometric properties sufficient?

Making theoretical claims without addressing methodological concerns is not a new problem as "theory often precedes measurement" and research on teams is no exception (Kuhn, 1961; Murphy & Ackermann, 2014, p. 17). While Muthén 's (1994) multilevel factor analysis (MFA) provides necessary methodological advancement in examining the psychometric properties of multilevel constructs, the large sample size requirement (i.e., 100 teams/groups and 300 individual observations) exceeds that which researchers typically use to test their hypotheses (Asparouhov et al., 2015; Dyer et al., 2005; Hox & Maas, 2001; Mok, 1995). Therefore, it is important to know the current practices that researchers use to examine the psychometric properties of measures of common team constructs.

This review is limited to specific measures of team cohesion, conflict, psychological safety, satisfaction, and task interdependence; it includes the articles in which the measure was initially published and other articles since 2005, the year after Chen and colleagues' (2004) review (Carless & De Paola, 2000; Edmondson, 1999; Jehn & Mannix, 2001; Loughry & Tosi, 2008; Van der Vegt et al., 2001). Journals[2] and dissertations were searched via the EBSCO library search database based on the name of the construct (e.g., team cohesion and team cohesiveness) and subconstructs (e.g., interpersonal cohesiveness, task attraction, and task commitment) for the period of 2005 to 2010. The articles (comprising 135 publications) are empirical studies that used at least one of the common measures of team constructs or its subdimensions, did not make major edits to the construction of the items (i.e., significant changes in the wording or number of items used), and reported at least one psychometric property of the measure (e.g., model fit index, factor loadings, residual variances, estimate of reliability). A breakdown by measure is as follows: team cohesion (3 articles and 4 dissertations); team conflict (16 articles and 13 dissertations); team psychological safety (46 articles and 42 dissertations); and team task interdependence (11 articles). There were no publications for the measure of team satisfaction.

[2] [*Organizational Science, Small Group Research, Organizational Research Methods, Journal of Management, Journal of Applied Psychology, Group and Organization Management, Leadership Quarterly, The Academy of Management Journal, Journal of Organizational Behavior, Journal of Occupational Behavior, Human Relations, Journal of Business Ethics, and Group Dynamics: Theory, and Research, Practice*.]

Following Chen and colleagues' (2004) recommendations, each article was examined based on four key areas researchers must address to draw conclusions regarding consensus measures' psychometric properties: examine the factor structure, evaluate intermember agreement, assess the measure's internal consistency, and ensure that within-team agreement justifies aggregation (p. 287-292). The first three areas directly relate to the measures' psychometric properties while the last area focuses on how team dynamics influence team level phenomena.

First, I examined for whether the researcher tested the factor structure via a CFA (i.e., individual level factor analysis) versus an MCFA (multilevel CFA) and the range of factor loadings. I then examined for the reporting of intermember agreement/deviation indices (e.g., $r_{WG(j)}$ and average deviation [*AD*]) and whether the results met the recommended thresholds for those respective indices (Dunlap et al., 2003; James et al., 1993; Smith-Crowe & Burke, 2003). Intermember agreement/deviation indices reflect the amount of variation within a team for a specific measure. Third, I examined for the reporting of the internal consistency of measures at the individual level (e.g., Cronbach's alpha; Cronbach, 1951) and/or scale reliability at the aggregate level or a multilevel composite reliability (G. Chen et al., 2004; Geldhof et al., 2014). Fourth, I determined justification for aggregation by examining the reporting of intra-class correlation coefficients (ICC(1); Bliese, 2000). ICC(1) captures the extent of influence team membership has on member scores and is an important piece of evaluating team consensus.

Of the scholarly publications examined, 16% reported conducting a CFA; only one article examined a measure using the MCFA framework, 12% reported either a range or a list of factor loadings associated with the measures' items, 4% reported at least one model fit index, and 41% reported the interrater agreement index, $r_{WG(j)}$, and none reported *AD*. For estimates of reliability, 97% of retained articles reported Cronbach's alpha results from a CFA, 1% reported an aggregated alpha, and 0% reported a composite reliability. 22% of the publications reported ICC(1) as evidence of variation in team member scores due to team membership, which was used as a prerequisite for examining a higher-level phenomenon. (For a more detailed review, see Table 3.)

The results of the literature review reveal that researchers typically evaluated measures at the individual rather than the team level of analysis. I was unable to find examples where any measure was fully evaluated at the team level based on the criteria as listed. Since the measures were designed in a consensus model, it was surprising that less than 52% reported any agreement indices. Based on these results, it is unclear if these common measures of team constructs truly capture team phenomena. Therefore, the current reporting practices on common measures of common team constructs' psychometric properties are not sufficient.

2. Psychometric Properties Across Levels of Analysis: Are there major differences in the psychometric properties for these measures from the individual to the between-team level of analysis?

Differences in the psychometric properties from the individual to the team level of analysis are problematic in common measures of team constructs if they are not psychometrically isomorphic, which refers to the measurement invariance across levels of analysis. In the current study, the degree of psychometric isomorphism (e.g., partial configural, strong configural, weak metric, and strong metric) reveals the extent to which conclusions drawn from individual-level data are consistent with that found at the theoretically relevant level of analysis (i.e., team).

Differences between the individual and the team levels of analysis highlight the potential consequences of misalignment in theory and measurement and should be assessed and interpreted in an MCFA via model fit, factor variances and loadings, and residual variance (Dyer et al., 2005; Geldhof et al., 2014; Tay et al., 2014). This misalignment is problematic if key assessments regarding a measure's quality do not remain consistent across levels of analysis: model fit (i.e., overall ability to capture the latent construct), indicator's strength (i.e., estimated factor loadings), factor variances (i.e., the communality of variance among items due to a common factor), and residual variance (i.e., unique variance due to a specific feature of the item and errors in measurement)[3].

In a multilevel context, these key assessments inform researchers as to a measure's degree of psychometric isomorphism, which indicates the extent to which conclusions drawn from lower-level data (e.g., team members) are consistent at a higher level of analysis (e.g., team; Meredith, 1993; Ryu, 2014; Tay et al., 2014). This differs from examining the homologous nature of a broader construct in which the lower-level construct is similar to its higher-level counterpart (G. Chen et al., 2004).

In the current study, the primary focus is the ability of these measures to capture team phenomena rather than lower-level individual differences. While theoretical meaning across levels of analysis is discussed (see Figures 1 and 2), comparisons across levels primarily focused on measures' psychometric properties, not on theoretical meaning. Cross-level comparisons were examined in terms of the consequences of misalignment in measurement and theory. Interpretation of construct meaning at the within-team and the within-person measurement

3 See Appendix for a review of the equations and an in-depth discussion detailing the different parts of an MCFA.

models was investigated to a limited degree and only when deviation within a grouping factor (e.g., person, team) was of substantive theoretical interest.

Degrees of Psychometric Isomorphism. Cross-level comparisons of measures' psychometric properties are best understood via Tay and colleagues' (2014) framework assessing the degree of psychometric isomorphism in measures where they make an important distinction between measurement invariance across groups and levels. Their framework is crucial for understanding whether measures designed to capture a lower-level construct (e.g., self-efficacy) are consistent in conceptual meaning and properties to measures capturing their higher-level counterpart (e.g., collective efficacy). Consistent with previous research on evaluating the degree of psychometric isomorphism, these increasingly stringent standards are broken into two broad categories (i.e., configural and metric isomorphism), which inform researchers as to the strength of their claims that a measure captures higher-level constructs when derived from lower-level data (Byrne et al., 1989; Ryu, 2014; Tay et al., 2014; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Simply put, the more stringent the standard, the stronger the claim.

While Tay and colleagues' (2014) framework compares distinct measures of a construct (e.g., self-efficacy and collective efficacy) in a multilevel context, this framework can be applied to categorize the degree of psychometric isomorphism in a single measure.[4] In the current study, the degree of psychometric isomorphism reveals the extent to which conclusions drawn from individual-level data are consistent with that found at the theoretically relevant level of analysis (i.e., team) in common measures of team constructs.

4 In Table 1, the comparison between measurement invariance/equivalence across groups and levels (i.e., cross-level isomorphism) is included as a reference point to help researchers who are more familiar with that concept.

Configural Isomorphism. The least stringent form of psychometric isomorphism, configural isomorphism, refers to the conceptual similarities across levels of analysis in a construct. In other words, does the meaning of the construct vary as a function of the level of analysis? Configural isomorphism examines the extent to which a measure's indicators relate to a factor. Within this broad category, there are three distinct types: partial, weak, and strong.

Partial configural isomorphism occurs when some, but not all, theoretical dimensions are found at different levels of analysis in a measure via factor analysis (see Figures 4 and 7). Some researchers argue that a simplified factor structure occurring at a higher level of analysis with similar meaning to its lower-level counterpart is a form of configural isomorphism (D'Haenens et al., 2012; Ryu, 2014; Stapleton et al., 2016). In other words, at least one factor remains consistent across levels, and/or a more broadly defined construct at a higher level of analysis encompasses some or all the nuanced subdimensions (i.e., multiple factors) found at lower levels, thus maintaining a degree of similarity in construct meaning at both levels.

A factor structure for a measure that remains consistent across levels implies *weak configural isomorphism*; that is, the construct's measure has the same number of dimensions across levels of analysis when using similar, but not necessarily the same, indicators and is also assessed via model fit indices and factor loadings. I examined well-established measures in a CFA framework and assumed the indicators loaded onto the same factor at the individual and the between-team levels. Because I constrained factor loadings with a marker variable (as typical in the CFA framework) and expected items to load onto the same factors at both levels, I did not examine for weak configural isomorphism.

The most stringent standard, *strong configural isomorphism*, refers to consistency in a measure's dimensionality and indicator quality across levels of analysis. Strong configural isomorphism is assessed via an MCFA by which researchers determine whether there is evidence for a higher-level construct (see B. O. Muthén, 1994; Tay et al., 2014). I examined psychometric isomorphism within a single measure. The indicators were the same and the factor structure was assumed to hold at both levels. That is, all indicators/items were expected to load on the same latent factor at both the individual and the team levels of analysis (see Tables 3 and 6). Therefore, at a minimum, I expected all measures of common team constructs to reveal strong configural isomorphism.

Metric Isomorphism. A stricter category of psychometric isomorphism than configural, metric isomorphism refers to consistency in the pattern and/or magnitude of factor loadings and to residual variances at different levels of analysis (Tay et al., 2014). Metric isomorphism has two distinct types: weak and strong. *Weak metric isomorphism* describes measures in which the rank order (i.e., pattern) of factor loadings remains consistent across levels of analysis. I expected that the measures' items would remain consistent in their ability to capture aspects of a construct at the individual and the team levels of analysis. In other words, the general ability of a measure's indicator to capture a latent factor is not influenced by the level of analysis.

The most stringent test of psychometric isomorphism, *strong metric isomorphism*, refers to both the pattern and magnitude of factor loadings being consistent across levels of analysis. Researchers have found that magnitude of factor loadings is greater at the level of analysis in which the construct is hypothesized to operate; therefore, there are likely to be differences in the magnitude of factor loadings at the individual and the team levels (Byrne et al., 1989; Dyer et al., 2005; B. O. Muthén, 1994). Whether these common measures of team-related constructs reveal strong metric invariance depends on whether the measure is designed to capture a team-related construct at a single level (e.g., team) or at multiple levels of analysis (e.g., individual, team).

There will likely be differences in the magnitude of factor loadings at the individual and the team levels for some of the measures of common team constructs as many are designed to solely capture team phenomena.

Therefore, in line with Tay and colleagues' (2014) framework, these measures are assumed, at a minimum, to have weak metric isomorphism from the individual to the team level. In other words, the pattern (i.e., rank order) of the factor loadings is expected to remain consistent at both levels but the magnitude of factor loadings may vary between levels. Regardless, weak metric isomorphism provides sufficient evidence of good quality measures of team phenomena, since the assumptions made about capturing a team construct based on the results of a CFA (as opposed to an MCFA) are, in large part, accurate.

Hypothesis 1: Common measures of team constructs (i.e., team cohesion, conflict, psychological safety, task interdependence, satisfaction) reveal metric isomorphism from the individual to the between-team level of analysis.

Hypothesis 1a: Team cohesion reveals metric isomorphism from the individual to the between-team level of analysis.

Hypothesis 1b: Team conflict reveals metric isomorphism from the individual to the between-team level of analysis.

Hypothesis 1c: Team psychological safety reveals metric isomorphism from the individual to the between-team level of analysis.

Hypothesis 1d: Team satisfaction reveals metric isomorphism from the individual to the between-team level of analysis.

Hypothesis 1e: Team task interdependence reveals metric isomorphism from the individual to the between-team level of analysis.

3. Examining for Cross-level Variation: Do the relationships among indicators (e.g., dimensionality) within these measures stay consistent from the individual to the team level of analysis?

Partial configural isomorphism occurs when fewer dimensions are found at a higher level of analysis. I expected partial configural isomorphism to occur if the multidimensional measures of team cohesion or conflict had fewer factors at the team versus the individual level. While a simplified factor structure implied a degree of similarity in the meaning of the construct across levels of analysis (e.g., overall sense of team cohesion or conflict), the indicators are loaded onto one or two factors, as opposed to three, at the team level. Psychometrically, partial configural isomorphism occurs when different factor structures best fit the data, via model fit indices and factor loadings, at different levels of analysis. This is consequential because if fewer dimensions/factors are found at the team level, researchers may need to refine the definition at this higher level of analysis to reflect the simplified factor structure (Tay et al., 2014). Therefore, it is important to test if multidimensional measures of team constructs operating in a multilevel context (e.g., people grouped in teams) reveal the same factor structure when examined at the theoretically relevant level of analysis (i.e., team). I expected that the three-factor models of team cohesion and conflict would hold when examined at the team level of analysis, as consistent with their theoretical development.

Hypothesis 2: Measures of team cohesion and conflict have the same factor structure at the individual and the between-team levels of analysis.

Factor Structure for Team Cohesion. Team cohesion reflects how well a team works together by assessing overall commitment to reaching team-related goals (i.e., task commitment) and how much the team enjoys working together (i.e., interpersonal cohesiveness) on team-related tasks (i.e., task attraction; Jehn et al., 2008; Loughry & Tosi, 2008; Marks et al., 2001; Mullen & Copper, 1994). Team cohesion encompasses three theoretically distinct but related constructs. In a previous CFA, the current study's measure of team cohesion revealed three latent factors at the individual level of analysis (Loughry & Tosi, 2008). To examine if this measure adequately captured these theoretically distinct dimensions, I examined alternative and theoretically relevant models (i.e., alternative factor structures) with varying latent factors at the team level of analysis that were not previously supported at the individual level via a CFA.

The current study examined model fit across levels of analysis in one-, two-, and three-dimensional models of team cohesion[5]. The one-dimensional model at the team level indicated that only a general latent factor operated at this level, suggesting that the interpersonal, task attraction, and task commitment dimensions were all heavily driven by an overall sense of team cohesiveness. Only the superordinate construct (e.g., team cohesion) was an identifiable factor at the team level (Johnson et al., 2011). Although the subdimensions may have theoretical meaning as they capture various aspects of team cohesion, researchers would not be able to psychometrically distinguish dimensions of team cohesion as drivers or outcomes of team phenomena.

In the two-dimensional model, the dimensions of task attraction and commitment were collapsed into a task cohesion factor while interpersonal cohesiveness represented the second factor of team cohesion. Theoretically, team cohesion is thought to span both vertically (i.e.,

5 The three dimensions of team cohesion included in the current study (i.e., interpersonal cohesiveness, task commitment, and task attraction) were chosen as they are more relevant to the temporary project team context. The dimension of interpersonal cohesiveness was included while the social cohesion dimension was not included. While Beal et al. (2003) found that social cohesion was a distinct component of group cohesion (as hypothesized by Festinger's theoretical development of the construct), social cohesion focuses more on group pride, which is less relevant in this context. Interpersonal cohesiveness is more appropriate as it focuses on whether team members enjoy working together (Festinger et al., 1950).

individual and team levels) and horizontally (i.e., social- and task-related dimensions; Beal, Cohen, Burke, & McLendon, 2003; Braun et al., 2020). The current measure captures both social (i.e., interpersonal cohesiveness) and task-related (i.e., task attraction and commitment) dimensions. In this model, the task-focused items measured the extent to which team members feel connected due to the nature of the team's work (i.e., team tasks; Schaffer & Manegold, 2018). The interpersonal factor taps into the relationships/social dynamics among team members by measuring how much they like each other and how well they get along and work together (Braun et al., 2020; Mullen & Copper, 1994). If the two-dimensional model has a better fit at the team level versus the three-dimensional model at the individual level, this measure of team cohesion has a simpler factor structure at the team level (often found in an MCFA), making it difficult to psychometrically distinguish between task attraction and commitment at the team level. If a simplified factor structure exists at the team level of analysis, then this measure of team cohesion would reveal partial configural isomorphism (see Figure 5 for an example of a measurement model with a simplified factor structure at the higher level of analysis).

Consistent with previous research, I tested for the three distinct factors found at the individual level of analysis (i.e., interpersonal cohesiveness, task attraction, and task commitment) and expected that these dimensions would be found at both the individual and the team levels of analysis across samples (Loughry & Tosi, 2008; Mullen & Copper, 1994; Schaffer & Manegold, 2018). If true, researchers would be able to psychometrically distinguish between the dimensions of team cohesion at the -team level of analysis, allowing them to investigate nuanced differences between how well a team works together due to their interpersonal interactions (i.e., interpersonal cohesiveness), interest in the team-related tasks (i.e., task attraction), and how united they are to finish the tasks (i.e., task commitment). Therefore, consistent with previous research via CFA, a three-factor model of team cohesion should show superior model fit at the between-team level of analysis versus a one- or a two-factor model (Beal et al., 2003; Loughry & Tosi, 2008).

Hypothesis 2a: The measure of team cohesion reveals a three-factor model (i.e., interpersonal cohesiveness, task attraction, and task commitment) versus a two-factor model (i.e., interpersonal and task-oriented cohesion) or a one-factor model (i.e., general team cohesion) at the between-team level of analysis.

Factor Structure of Team Conflict. Similar to the measure of team cohesion, Jehn and Mannix's (2001) multidimensional measure of team conflict contains subconstructs representing three distinct types of team conflict in which the factor model is typically assessed at the individual level via CFA (Jehn et al., 2008; Jehn & Mannix, 2001). Based on this three-factor model, Jehn (1997) found that people describe differences in team conflict by how team members interact with one another (i.e., relationship conflict); the extent to which people disagree on how to accomplish team goals, work delegation, and resource allocation (i.e., process conflict); and differences in opinions and disagreements regarding team tasks (i.e., task conflict). However, theoretical support and statistical evidence suggest that there is substantial overlap among the relationship, process, and task team conflict subconstructs, making it difficult to distinguish them psychometrically.

Task and process conflict are theoretically related and strongly intercorrelated, as both capture disagreements on team-related matters. Both process and relationship conflicts elicit negative feelings, with relationship conflict being more emotion-laden (Jehn et al., 2008; Wit et al., 2012). Latent profile analysis suggests that the relationship and process dimensions of team conflict tend to follow similar patterns while high/low levels of task conflict capture distinct profiles of team conflict (O'Neill et al., 2018). Due to this overlap, researchers have hesitated to examine the influence of process conflict in team dynamics (Shaw et al., 2011). To evaluate the degree of overlap psychometrically, this measure needs to be examined for its ability to capture distinct dimensions of team conflict at the team level of analysis.

Examining the factor structure of this measure at the team level can provide clarity on its ability to capture the theoretical dimensions of relationship, process, and task team conflict. Measures often have simpler factor structures at higher levels of analysis (Dedrick & Greenbaum, 2011; Dyer et al., 2005; Huang et al., 2015; Kim et al., 2016). If this were the case regarding team conflict, the influence of team membership would drive the similarity in team-member perceptions of team task and process conflict or relationship and process conflict, thereby making them psychometrically indistinguishable at the team level. Therefore, distinctions between the factors at the individual level would not be evident at the team level of analysis. As team conflict is theorized to primarily reside at the team level, distinguishing between these dimensions of team conflict via a CFA may provide misleading results regarding the measure's quality.

While a two-factor model of team conflict would explain some of the difficulties in isolating the influence of process conflict among teams, it is more likely that this carefully developed measure, through qualitative and quantitative methods, will reveal three distinct factors at the team level (Jehn, 1997; Jehn et al., 2008; Jehn & Mannix, 2001). Researchers have found that different emergent states influence these distinct dimensions differently; the influence of process conflict on team performance varies based on the team's developmental stage and degree of relationship conflict within the team (Jehn et al., 2008; Wit et al., 2012).

Hypothesis 2b: The measure of team conflict reveals a three-factor model (i.e., relationship, process, and task conflict) versus a two-factor model (combining the process and task or process and relationship dimensions) at the between-team level of analysis.

Part II: Potential Reasons and Consequences for Varying Degrees of Psychometric Isomorphism

Part II explores potential reasons and consequences for varying degrees of psychometric isomorphism from the individual to the team level of analysis for measures of team consensus constructs (i.e., team cohesion, conflict, psychological safety, satisfaction, task interdependence, liking, and viability). I discuss how a measure's characteristics (i.e., referent, target) and degree of agreement in a sample could be linked to its degree of psychometric isomorphism and how relationships among variables may differ when examined at different levels of analysis.

4. Measures' Characteristics: How does a measure's referent (e.g., "I" versus "team") and target (e.g., member–member relationship versus team-member?) relate to a measure's psychometric properties at different levels of analysis?

A measure's characteristics likely influence its degree of psychometric isomorphism because its referent and target determine the focus of respondents to a one level of analysis minimizing the variance associated with the other level (van Mierlo et al., 2009). The referent focuses the respondents on themselves or toward something else (e.g., team) and the target provides the overarching context (e.g., member-member relationship and team). As these measures best reflect constructs that reside at the level of analysis of the referent and/or target, the degree of psychometric isomorphism from the individual to the team level is influenced by both the referent and the target.

Hypothesis 3: The characteristics of measures of common team constructs relate to their degree of psychometric isomorphism, such that stricter measurement invariance from the individual to the between-team levels of analysis occurs based on the measure's referent and/or target.

Referent's Influence on Psychometric Isomorphism. Theoretically, measures designed in a referent-shift model primarily operate at the level of the referent regardless of the level of data collection. In the current study, measures of team constructs in a referent-shift consensus model (e.g., team cohesion, conflict, and psychological safety) reflect the latent factor primarily operating at the team level of analysis since the items encourage team members to report on team-level phenomena; measures in a direct consensus model (e.g., team satisfaction and task interdependence) encourage members to report their nuanced perceptions as a team member (van Mierlo et al., 2009). Direct consensus models introduce additional variation in scores since team members are asked to assess their own attitudes, beliefs, and cognitions, whereas referent-shift models ask members to describe aspects of the team (Arthur et al., 2007). Measures of team constructs in a referent-shift model primarily reside/operate at the team level; those in a direct consensus model are designed to reflect both individual- and team-level phenomena (Chan, 1998). While there is theoretical justification for measures in a direct consensus model to capture meaningful variance at multiple levels of analysis, statistical evidence is also required.

By linking theory and measurement, the MFA provides psychometric evidence for how measures in a referent-shift model differ psychometrically from those in a direct consensus model via model fit indices, factor loadings, and residual variance across levels of analysis. Superior model fit at one level of analysis over another is a psychometric indication of the level at which the construct primarily operates (Dyer et al., 2005; Tay et al., 2014). Measures of team cohesion, conflict, and psychological safety will reveal the best fitting model at the team level of analysis via evaluating the team level covariance matrix (S_B). Running a factor analysis on the S_B is akin to modeling variance based solely on the influence of team membership and is best visualized by the model of the shared cluster construct (see Figure 4).

Measures in a direct consensus model (e.g., team satisfaction, team task interdependence) are theoretically capable of capturing distinct but related constructs at the individual and team levels of analysis. Both levels capture team-related constructs but they are theoretically distinct. For example, the current study's measure of team satisfaction captures members' general satisfaction with their team regardless of team membership at the individual level, and overall team satisfaction at the team level of analysis. The two-level factor model best represents a direct consensus model as both the individual and team level of analysis are theoretically relevant (see Figure 4). Since these measures are designed to capture phenomena at both levels, model fit indices will reveal good fit at the individual and team level of analysis.

Differences in the magnitude of factor loadings across levels of analysis indicate the influence a measure's referent has on its psychometric properties and may occur based on the item wording in the measure, which determines the type of consensus measurement model (Kim et al., 2016; Stapleton et al., 2016). Measures designed to capture a construct that primarily resides at the team level (e.g., a referent-shift consensus model in the current study) will likely exhibit greater factor loadings at the team level, since the items are designed to reflect variation in team-level phenomena versus measures that focus on the team member working in a team context, as found in direct consensus measurement models (e.g., team satisfaction, team task interdependence).

Researchers have suggested that factor loadings increase or decrease from the lower to the higher level of analysis; three studies showed an average increase of .23 in factor loadings (Dyer et al., 2005; Reise et al., 2005; Whitton & Fletcher, 2014) in an MCFA when investigating for higher-level constructs. Since common measures of team constructs are designed to primarily capture team phenomena, the difference in the magnitude of factor loadings between measures in a referent-shift and a direct consensus model will likely be smaller than .23.

There is a lack of research on variation in the magnitude of factor loadings across levels of analysis and the nature of these measures. The current study takes a more tempered approach to the expectation that the factor loadings will increase by a minimum of .10 from the individual to the team level of analysis for measures in a referent-shift model and that a modest increase (ranging from .1 to .09) will occur in a direct consensus measurement model.

Residual variance, which identifies differences between referent-shift and direct consensus models, is influenced by the level of analysis. Residuals are likely greater in measures at the individual level in a referent-shift model because of variation due to individual differences not associated with higher-level team phenomena (Kim et al., 2016). The lower-level focus is on individual differences within the sample with the context being a team environment. These measures in a direct consensus model encourage substantive variation among team-members by asking them their individualized experience within a team minimizing the amount of residual variance at the individual level compared to measures in a referent shift model.

Taking these indications of psychometric isomorphism into consideration, measures in a referent-shift model will reveal more measurement invariance from the individual to the team level of analysis, which is a greater indication that the construct being measured primarily operates at a higher level of analysis. Measures in a direct consensus model that theoretically capture phenomena at multiple levels will indicate less measurement variance across levels (i.e., greater psychometric isomorphism) at the level of the referent (i.e., individual) and the context (i.e., between-team).

Hypothesis 3a: The referent of a measure's items influences the degree of psychometric isomorphism among measures of common team constructs, such that measures that refer to the self (direct consensus measurement models) reveal stricter measurement invariance from the individual to the between-team level of analysis than measures that refer to the team (referent-shift consensus measurement models).

Target's Influence on Psychometric Isomorphism. Researchers have frequently employed measures designed to capture individual, dyadic, or team/group phenomena and related but distinct constructs at different levels of analysis (Gooty & Yammarino, 2011; Stapleton et al., 2016). For example, a measure's target is the context of the items such as a person, relationship, team, or workplace. In teams research, phenomena are investigated both where the primary target/context is the team (e.g., team task interdependence and satisfaction) and the more specific context/target of team member-member relationships (e.g., team member liking and viability; O'Neill et al., 2018; Robert et al., 2019; Tekleab et al., 2009).

While a measure may capture theoretically related constructs at multiple levels of analysis, the way in which constructs are aggregated varies. For instance, common measures of team constructs typically require transformation (e.g., aggregation) of member perceptions to reflect team-level phenomena; measures such as member liking and viability are aggregated within individuals and across the team (i.e., overall team member liking and team viability; Thomas et al., 2019; Woehr et al., 2015). Researchers use these measures (aggregated differently) − to investigate related constructs across levels, enabling them to distinguish the individual, the within-team, and the team level influences on other team-related phenomena. However, these measures' ability (i.e., measure quality) to capture constructs at various levels of analysis will likely differ.

Researchers should be aware of how a lack of isomorphism reduces the ability to capture a construct when the measure's target differs from the level of analysis. Measures in a direct consensus model will differ in their degree of isomorphism based on the specificity of the measure's target because the amount of variation associated with the level of analysis will coincide with the level of the referent and the target. However, this assumption about the relationship between the measure's target and the measure's ability to capture a phenomenon at different levels of analysis has yet to be examined psychometrically regarding measures of team consensus constructs.

To test this assumption in the current study, the influence of the target on isomorphism was examined by comparing measures in a direct consensus model (e.g., team satisfaction and task interdependence, and team member liking and viability) that are theoretically capable of capturing constructs at multiple levels of analysis (e.g., within-person, individual, between-team). Measures with a member-to-member target should reveal superior model fit, factor loadings, residual variance, and estimate of reliability at the within-person and the between-team levels versus the between-person or within-team level; those with a team target (e.g., team satisfaction and task interdependence) should reveal superior metrics at the individual and the team levels of analysis. Therefore, indications of isomorphism across levels provide psychometric evidence for the link between the wording in a measure (e.g., a measure's target) and the ability of the measure to capture phenomena at various levels of analysis.

Hypothesis 3b: In direct consensus models, a measure's target influences the level of psychometric isomorphism such that measures with a target of member-member relationships (e.g., team member liking and viability) have poorer model fit, reduced factor loadings, and greater residual variance at the between-team level versus the within-team and the individual levels of analysis compared to measures with a team target, which will reveal superior indicators of psychometric isomorphism at the individual and the between-team levels of analysis.

5. Team-member Agreement and Psychometric Isomorphism: To what extent does team-member agreement influence a measure's psychometric isomorphism?

For researchers to claim the presence of a team consensus construct, team members must experience a degree of consensus regarding the phenomenon; for example, agreeing on how satisfied they are with their team (team satisfaction) to be examined as a team-level phenomenon. Common team constructs exist in a sample only when sufficient agreement among team members is present via ICC(1) cutoff scores relative to the specific measure (Woehr et al., 2015). The sample-specific index of ICC(1) indicates that the amount of variation in observed scores is due to group membership (Bliese, 2000). An essential step in MCFA, ICC(1) provides evidence for the presence of a higher-level construct and has implications regarding biased estimates in factor loadings and a measure's reliability (Can et al., 2015; Geldhof et al., 2014; Hox & Maas, 2001). However, the psychometric assessment of whether a common understanding/perception has emerged in a team is not sample-specific, but team-specific. Therefore, whether the presence of a consensus influences the degree of psychometric isomorphism must be assessed via a team-specific index (e.g., $r_{WG(j)}$ and *AD*) as opposed to a sample specific index (e.g., ICC(1) (Bliese, 2000; Burke, Finkelstein, & Dusig, 1999; James, Demaree, & Wolf, 1993).

While ICC(1) evaluates the viability of the sample in examining a higher-level construct, the $r_{WG(j)}$ and *AD* indices of interrater agreement allow researchers to uncover how teams shape members' views on team-level phenomena (B. O. Muthén, 1994). For example, low ICC(1) values likely bias factor loadings because teams that do not agree on how to describe their team do not share a similar conceptualization of team phenomena (Can et al., 2015; Hox & Maas, 2001). The degree of psychometric isomorphism (i.e., amount of variance across levels of analysis) will be linked to the degree that team members share a common understanding regarding a team construct. In the current study, this common understanding is operationalized via team-specific agreement indices Thus, a higher level of agreement within a team on a team construct influences a measure's degree of psychometric isomorphism.

Hypothesis 4: Team member agreement influences a measure's level of psychometric isomorphism such that measures reveal a stricter standard of measurement invariance across levels of analysis when there is greater agreement among members.

6. Psychometric Isomorphism and Relationships Among Team Consensus Constructs: How do relationships among variables differ at various levels of analysis?

While psychometric isomorphism refers to the internal properties of a measure (i.e., regression intercepts, factor loadings/regression slopes, and residual variance), structural invariance refers to consistency in factor/latent means and variance/covariance structures (Byrne et al., 1989). Researchers have found that a measure's degree of psychometric isomorphism influences its structural invariance (Byrne et al., 1989; Zyphur et al., 2008); that is, when variance in factor structures or loadings occurs across levels of analysis, the correlations among latent variables will also differ across levels. These patterns and magnitudes of correlations provide important evidence of discriminant validity among latent variables (Gerbing & Anderson, 1988; Malhotra et al., 2014). In the current study two questions emerge regarding measures of team constructs and structural invariance: to what degree does psychometric isomorphism influence structural invariance in team measures, and do the patterns and magnitudes of correlations among team variables still offer evidence for discriminant validity when accounting for the nested nature of the data?

When measures capture multidimensional constructs (e.g., team conflict and cohesion), structural invariance is more narrowly investigated in terms of configural isomorphism (Tay et al., 2014; Zyphur et al., 2008). The current study investigates structural invariance among measures by examining how the relationships among latent factors may vary across levels of analysis. Here, configural isomorphism refers to assessing the internal structure of a measure while structural invariance refers to assessing the relationship of latent factors of theoretically distinct constructs.

Psychometric Isomorphism as a Constraining Force. Based on the different degrees of psychometric isomorphism (e.g., partial configural, weak configural, strong configural, weak metric, and strong metric; Tay et al., 2014), common measures of team constructs may vary in the pattern and magnitude of the factor loadings across levels of analysis and still be considered psychometrically isomorphic, to a degree. If the relationship between the indicators and the latent factors is different at the team versus the individual level of analysis for any of the measures of team constructs, the relationships among these constructs may also vary across levels of analysis (Byrne et al., 1989). By examining how/if the degree of psychometric isomorphism within a measure influences structural invariance across levels of analysis among team measures, I hope to establish the extent to which relationships among team-related variables identified at the individual level remain consistent at the team level of analysis.

The measures are expected to have some degree of metric isomorphism. Weak metric isomorphism refers to consistency in the relative ordering of factor loadings across levels of analysis, and strong metric refers to the consistency in both the relative ordering and magnitude of factor loadings at the individual and the between-team levels of analysis. Therefore, it is important to understand if the consistency in the relative ordering and/or magnitude of factor loadings across levels of analysis (i.e., weak and/or strong metric isomorphism) constrains the degree of structural invariance among measures of team constructs.

Regarding the current study, structural invariance among common team measures means that the covariance/variance matrix between variables does not differ at the individual versus the between-team level of analysis. At the individual level, variance does not account for the nested nature of the data; at the between-team level the variance reflects the association between two variables based on the constructs' latent factor means (Byrne et al., 1989).

According to previous meta-analytic research, the type of consensus model (i.e., referent-shift and direct consensus) influences relationships among constructs (Wallace et al., 2016). The current study dives deeper by examining if the degree of psychometric isomorphism may be the root cause of this finding (Wallace et al., 2016); that is, if measures of team constructs that reveal metric isomorphism also hold in how they relate to each other across levels of analysis, then structural invariance may be inherently linked to the degree of measures' psychometric isomorphism. I attempt to provide more insight into the link between the psychometric properties within a measure and structural invariance among measures across levels of analysis by posing the following question:

Research Question 1: Does the degree of psychometric isomorphism constrain the degree of structural invariance among measures of team cohesion, conflict, interdependence, psychological safety, satisfaction, general member liking, and viability?

Psychometric Isomorphism and Discriminant Validity. As team phenomena operate in a dynamic environment in which constructs such as team conflict, cohesion, psychological safety, and satisfaction all influence each other across the team's life cycle, establishing they are distinct constructs (i.e., discriminant validity) is essential (Ilgen et al., 2005; Marks et al., 2001). For example, to provide evidence for discriminant validity, the extent to which latent variables relate to one another should be assessed via the covariance/variance matrix (i.e., Phi matrix, Φ; Gerbing & Anderson, 1988; Malhotra et al., 2014). However, because researchers often examine the relationships among variables at the individual level (Φ_I) rather than at the team level of analysis (Φ_B), the team's influence on perceptions of team phenomena is not incorporated in the variance/covariance matrix among latent variables (Φ_I, see Figure 2).

Regarding the current study, all the measures of team constructs are reflective constructs in that the observed variable is influenced by the latent variable (Edwards & Bagozzi, 2000). When establishing discriminant validity in reflective constructs, the variances/covariances among latent variables should be less than 1.00 and "greater than twice their respective standard error" (Bagozzi et al., 1992, p. 668). I investigated whether there was still evidence for discriminant validity at the team level of analysis by examining if the relationship among these reflective constructs remained consistent (i.e., structurally invariant) at the individual and the team levels of analysis and/or maintained commonly held standards for discriminant validity.

Research Question 2: Does the level of analysis relate to the degree of differentiation among measures of team cohesion, conflict, interdependence, psychological safety, satisfaction, general member liking, and viability?

METHOD

Answering these hypotheses and research questions is accomplished by conducting analysis of an archival data set. The data analysis focuses on an empirical examination of measures to test this study's hypotheses with subsequent analysis to address the research questions.

Participants

Participants were U.S. college/university students who were part of course-related project teams whose work contributed to their final course grade. These students were enrolled in a 15-week, semester-long course in either the fall or spring terms (i.e., regular academic terms) between 2006 and 2020 using the Comprehensive Assessment of Team Member Effectiveness (CATME) system. Teams were composed of three to ten team members working together for a minimum of 90 days over the course of the semester. Participants completed measures assessing perceptions on their team (e.g., team cohesion, conflict, psychological safety, satisfaction, and/or task interdependence) and/or team members (e.g., team-member liking and viability) at one point in time from midway through the end of the semester, keeping the last time point in the final dataset.

There are three important features about these project teams. First, the performance of these teams had real-life consequences to participants (e.g., contributed to the final course grade). Second, these teams all reported on team emergent states and processes midway through the end of the semester giving team members enough time to report on the dynamics and characteristics of the team. Third, these teams worked together over similar lengths of time ranging from 12 to 15 weeks over the course of the semester.

Materials

Estimates were calculated using full-information likelihood and maximum likelihood with robust standard errors that accounted for clustered data via *R* and *MPlus,* as recommended by Yuan and Bentler (2000) when evaluating multilevel and/or non-normally distributed data. *MPlus* does this by default when specifying multilevel models; it must be specified using the lavaan package in *R*. The exception is estimates based on covariance matrices which inherently do not allow for specifying robust estimation and are not structured in a multilevel format.

Measures

All measures of team dynamics were assessed on a Likert-type scale. These measures and their subscales were aggregated across the team for each measure's item and included team cohesion, conflict, psychological safety, satisfaction, and task interdependence. Item scores were aggregated across individuals' assessments of their team members, then across the team to reflect general liking and relationship viability among team members. Referent-shift consensus measures included team cohesion, conflict, and psychological safety, while direct consensus measures included team satisfaction, task interdependence, general team member liking, and overall team member relationship viability. (See Table 5.)

Team Cohesion. Overall team cohesion was a nine-item measure with three subscales developed from combining two existing measures capturing interpersonal cohesiveness among team members and attraction and commitment to team tasks (Carless & De Paola, 2000; Loughry & Tosi, 2008). The abbreviated task commitment subscale (Loughry & Tosi, 2008) containing the three items with the highest factor loading was used instead of the original four-item Carless

& De Paola's (2000) measure. Example items include: "Team members get along well," "Team members like the work that the group does," and "I'm unhappy with my team's level of commitment to the task" [reverse coded]. Responses ranged from *strongly disagree* (1) to *strongly agree* (5).

Team Conflict. This nine-item measure assessed team conflict along three dimensions: task, relationship, and process conflict (Jehn & Mannix, 2001). Examples include: "How much conflict of ideas is there in your work group?" "How much relationship tension is there in your work group?" and "How much conflict is there in your group about task responsibilities?" reflecting task, relationship, and process conflict, respectively. Responses ranged from *none or not at all* (1) to *very often* (5).

Team Psychological Safety. Edmondson's (1999) seven-item measure was used to assess psychological safety within a team. Examples include: "Members of this team are able to bring up problems and tough issues" and "It is safe to take a risk on this team." Responses ranged from *very inaccurate* (1) to *very accurate* (7).

Team Satisfaction. Team satisfaction was assessed on a three-item measure; for example: "I am satisfied with my present teammates" (Van der Vegt et al., 2001). Responses ranged from *strongly disagree* (1) *to strongly agree* (5).

Team Task Interdependence. Five items captured the degree to which team members were required to work together to complete tasks. Examples include: "I have to work closely with my teammates to do my work properly" and "I depend on my teammates for the completion of my work." Responses ranged from *strongly disagree* (1) to *strongly agree* (5).

Team Member Liking and Viability. Team member liking and viability were measured with Thomas and colleagues (2019) six-item measure capturing team members' perceptions of each other. Three items assessed liking and the other three assessed the desire to work with each other again (i.e., team-member viability). Examples include: "I like this person as an individual" and "I would gladly work with this individual in the future." Responses ranged from *strongly disagree* (1) to *strongly agree* (5).

Hypothesis Testing and Research Questions

Hypotheses 1-1e

The first hypothesis proposes that common measures of team constructs (e.g., team cohesion, conflict, psychological safety, satisfaction, and task interdependence) are psychometrically isomorphic and represent quality measures at the between-team level of analysis. The psychometric properties of a measure are typically inspected in a factor analytic framework via exploratory factor analysis (EFA) and CFA (F. B. Bryant & Yarnold, 1995; Crocker & Algina, 1986). In a multilevel context, the FA framework inspects a measure's properties across levels (e.g., individual, within-team, and between-team) via an MFA. The current study examined well-established measures using MCFA to inspect for psychometric isomorphism, as opposed to an EFA which is used for the development of measures (Kim et al., 2016; L. K. Muthén & Muthén, 2017). I chose MCFA techniques instead of an aggregated CFA to inspect team phenomena due to estimation problems (see Pornprasertmanit et al., 2014, for a review).

The psychometric properties of each measure were examined via measure reliability estimates and an MCFA. A measure's reliability was estimated with Cronbach's alpha (α) at the individual level for purposes of comparison with previous research; reliability at higher levels of analysis was estimated via a composite (ω) (Geldhof et al., 2014). The MCFA was conducted in a five-step process in *R* using the lavaan package and *MPlus* software; *R* syntax is available in the Appendix, with the exception of a three-level analysis conducted only in *MPlus* due to limitations in lavaan (Dyer et al., 2005; Huang, 2017; B. O. Muthén, 1994). The degree of psychometric isomorphism was established by following Muthén's (1994) five-step procedure:

- Step 1: Conduct a CFA to examine the factor structure at the individual level of analysis.
- Step 2: Examine whether it is appropriate to justify aggregation for a team-level construct via examining ICC(1).
- Step 3: Calculate the within-group covariance matrix (S_{PW}) and associated within-level CFA (appropriate for a deviation construct).
- Step 4: Calculate the between-group covariance matrix (S_B) and the associated between-level CFA (appropriate for a shared [e.g., team] construct).
- Step 5: Conduct the MCFA by combining within- and between-group covariance matrices to model the higher levels of analysis (appropriate for a two-level construct with a theoretically relevant deviation and shared construct).

After conducting these analyses, key indications regarding a measure's quality that must remain consistent for the measure to be considered psychometrically isomorphic were reviewed, such as model fit (i.e., measure's overall ability to capture the latent construct), indicator's strength (i.e., estimated factor loadings), factor variances (i.e., the communality of variance among items due to a common factor), and unique variance (i.e., item variance due to a specific feature of the item and errors in measurement; G. Chen et al., 2004). Regarding model fit, different indices were appropriate based on the level of analysis in a multilevel context. Model fit was estimated by various indices that collectively assess an indicator's ability to capture a latent variable.

In CFAs, a variety of fit indices are appropriate (e.g., X^2, *CFI, TLI, RMSEA*, and *SRMR*). However, Hsu and colleagues (2015) found that common fit indices are not sensitive to model misspecification at the within-group and between-group levels for a two-level factor model. Therefore, only appropriate model fit indices for these measures were examined in the MCFA. Although a variety of model fit indices are capable of detecting within-group model misspecification (e.g., *CFI, TLI, RMSEA*, and *SRMR-W*), only *SRMR-B* is appropriate for between-group assessments (Hsu et al., 2015). Table 2 reflects the level of the factor analysis in two-level models with the appropriate model fit indices and index description as relevant in an MCFA. While a CFA on the S_B is theoretically appropriate for shared constructs (e.g., common team constructs), the within-level CFA via S_W and two-level factor models via a covariance matrix combining the S_B and S_W were estimated for comparison purposes and to investigate the ability of these measures to capture deviation constructs (e.g., disagreement on the level of cohesiveness within a team).

All comparisons across levels of analysis for Hypotheses 1 and 2 used the results from Step one and Step four in an MCFA representative of Figure 3 and a shared construct in Figure 4, respectively. Step one consisted of a CFA modeling the individual level not accounting for team membership (S_T) and Step 4 consisted of a CFA modeling the team level (S_B), which is theoretically appropriate for shared consensus constructs such as team cohesion, conflict, psychological safety, satisfaction, and task interdependence.

Hypotheses 2-2b

Regarding the multidimensional measures of team cohesion and conflict, a three-factor model should reveal good fit at both the individual and the between-team levels of analysis consistent with the development of the constructs. However, if a simpler factor structure is found at the between-team level, the residual covariances should be examined. Differences in residual covariances across levels of analysis occur because the commonality among indicators is not always a result of the influence of the hypothesized latent factor. In a structural equation model (SEM), residual variance accounts for shared variance among the indicators that is a function of the item's wording or context and is not an aspect of the latent factor (Asparouhov et al., 2015). Therefore, residual variance was examined among hypothesized factor structures for measures of team cohesion and conflict. I expected that the residual variance would be minimized, resulting in greater reliability in the three-factor models and at the team level versus the more simplified factor structures, which is consistent with the theoretical development of the constructs.

For the multidimensional constructs (i.e., team cohesion, team conflict), after performing an MCFA, it is important to examine if the correlation among latent factors is small enough to support distinct constructs across the subdimensions (i.e., $|r| < .75$; Schmitt et al., 2018). By examining the correlations among latent factors represented by each respective subdimension, researchers can determine whether any moderate improvements in model fit, factor loadings, and residual variance were due to increasing the complexity of the model or if a distinct factor (e.g., team process conflict) explained substantive variance not explained by another factor (e.g., team relationship conflict).

Hypotheses 3-3b

Hypotheses 3a and 3b seek to understand how a measure's characteristics influence its degree of isomorphism. To test hypothesis 3a, I used the results of the MCFAs to compare the degree of psychometric isomorphism in referent-shift consensus measures (i.e., team cohesion, conflict, and psychological safety) and direct consensus measures (i.e., team satisfaction and task interdependence). I expected that the equality constraints (i.e., identical factor loadings and residual variance across teams) placed on the individual level of analysis in a CFA for referent-shift measures would result in poorer fitting models and reliability estimates compared to the team level, which accounts for the nested structure of the data, and that direct consensus measures would show a negligible difference. I also expected that referent-shift consensus measures would reveal higher factor loadings at the team versus the individual level of analysis compared to direct consensus measures, which would show similar factor loadings at the two levels. That is, referent-shift consensus measures should reveal strong metric isomorphism, while direct consensus measures should reveal only weak metric isomorphism.

Hypothesis 3b seeks to test whether a measure's target also influences its degree of isomorphism. I tested this hypothesis by conducting an MCFA and calculating reliability estimates on the measures of team-member liking and viability and then compared those results with those of common measures of team constructs designed as direct consensus measures (e.g., team satisfaction and task interdependence). I expected that measures with a team target would show greater psychometric isomorphism from the individual to the team level of analysis than those with a member-member relationship target from the within-person to the team level via a CFA at each respective level. Measures with a member-member relationship target should show meaningful variance at the lowest level of analysis (i.e., within-person), the level the measure was designed to capture, and the highest level (i.e., team), the level of the larger context (i.e., the team).

The measures with a team-member target are multilevel but have three levels of analysis (within-person, between-person, and between-team) with specific factors that need to be modeled (see Figure 2). Observations with one grouping factor (e.g., people grouped in teams) have two distinct levels of analysis (e.g., within-team and team); observations with two relevant grouping factors (e.g., multiple observations from a person and people grouped in teams) have three levels of analysis. These three levels have an associated variance/covariance matrix (e.g., S_{WP} = within-person, S_{BP} = between-person, S_{BT} = between-team). (See Figure 9 for the three-level factor model.) As with the two-level model, each level in the three-level model was examined via its relative variance/covariance matrix before the final multilevel model structure was assessed. In the MCFA examining three levels for the measures of team-member liking and viability, the ICC(1) scores were determined for each item at both the between-person and the team levels. An additional step with a CFA on the additional covariance matrix was conducted before the final step of examining a multilevel factor structure.

For measures of liking and viability, the within-person and the team levels and multilevel factor model were assessed and compared. Comparing these levels aligns with the theoretical examination of deviation and shared constructs (associated with the within-person and the team levels, respectively). There is no relevant theoretical construct using a traditional CFA on the lowest level of data (i.e., CFA on S_T) without accounting for a grouping factor for the measures of liking and viability. (See Figure 2 for a more detailed description.) These measures were originally designed to capture different perceptions people have of other team members; therefore, the measures' ability to model within-person variance is of substantive interest (Thomas et al., 2019).

Examining these measures' degree of psychometric isomorphism was achieved by comparing the within-person level (via CFA on the S_{WP}) and the team level (via CFA on the S_{BP}). Examining the between-team level via a CFA evaluated the ability of this measure to capture a shared construct typical of team phenomena. Comparing the within-person level to a multilevel factor model (i.e., the last step in a MCFA) differs in that a multilevel factor simultaneously models the variance due to differences in ratings within a person and to team membership. This multilevel modeling implies that this measure captures theoretically distinct but related constructs in which both a deviation and shared construct should be estimated (i.e., shared configural model; see Stapleton et al., 2016 for a review).

Differences in psychometric properties at the individual and the team levels were compared in the measures of team satisfaction and task interdependence to differences between the within-person and the team levels of analysis for the measures of liking and viability. I expected that the differences would be greater in measures with a “team-member” target. Based on the first steps in an MCFA, a multilevel factor model for measures in a direct consensus model was considered as a point of comparison, following recommendations by Muthén (1994), as measures in a direct consensus model inherently invite participants to introduce variation based on their own distinct perceptions as opposed to describing the team as a whole.

Hypothesis 4

Hypothesis 4 refers to whether a shared understanding among team members influences the conceptualization of a team construct. It tests whether agreement on the presence/magnitude of a common team construct influences the measure’s degree of psychometric isomorphism. First, the data was split into two separate datasets representing those teams that met the threshold for strong agreement and those teams that fell below. Determining the level of agreement was based on an interrater agreement/disagreement index calculated for each team; the process was conducted for both $r_{WG(j)}$ and *AD*. Following Woehr and colleagues’ (2015) recommendations, cutoff scores for $r_{WG(j)}$ and *AD* representing strong agreement for each measure were determined based on previous estimates (found during the literature review).

Second, the two groups were examined within a CFA framework for measurement invariance across groups at the individual level of analysis along the six increasingly stringent standards described by Vandenburg and Lance (2000). If measurement invariance across groups was not present, then Hypothesis 4 would be supported because the presence of a shared understanding (i.e., strong interrater agreement) influences the conceptualization of the measure. However, if at minimum the CFA model held across groups, I then examined the factor structure at the team level of analysis.

In a third step, I determined the degree of psychometric isomorphism for each group following MCFA procedures. If the level of agreement influenced the degree of isomorphism (e.g., partial, weak, and strong configural; weak and strong metric isomorphism), then Hypothesis 4 would be supported.

Research Questions

To understand how the relationships among measures of team constructs relate to isomorphism, I created variance/covariance matrices (i.e., Phi matrix, Φ) at every level of analysis (e.g., individual, within-team, and team). For the first research question, these matrices were examined to see if measures that are less/more isomorphic relate to each other differently at various levels of analysis; that is, if measures revealing strong metric isomorphism maintained similar relationships while measures with weaker forms of isomorphism had varying relationships.

Following Bagozzi et al.'s (1992) recommendations to examine for discriminant validity, I addressed the second research question by comparing the variance/covariance at the team level between two latent variables and their standard errors. If the variance/covariance was less than 1 but was two times the standard error for each construct, then there was evidence of discriminant validity at the level of analysis at which the construct is hypothesized to operate. I constructed correlation matrices across levels of analysis for the covariance/variance matrices from the respective levels to ease interpretation of cross-level variance/invariance; correlations ranged from 0 to 1 (Jak, 2019).

RESULTS

The results are broken down into two parts. The first focuses on the internal properties of the common measures of team constructs. It addresses Hypotheses 1 and 2 via an empirical study on archival data that examined these measures in an MCFA framework. Using the same data, the second part addresses Hypotheses 3 and 4 and both research questions. It consists of an empirical examination of what influences the degree of psychometric isomorphism within measures and if variance across levels of analysis within a measure relates to relationships among other measures of team constructs across levels of analysis.

Sample and Participants

After applying the inclusion criteria on the archival data, the sample retained for each respective measure is described in Table 32. Samples among the measures ranged from having 3,275 – 19,105 teams, 13,341 – 74,852 team-members, and 3.92 – 4.24 mean team size. Every sample reported to be predominantly male over female and White/Caucasian followed by Asian, Hispanic/Latino, Black/African American, Other, and Native American.

While only a portion of participants reported demographic information, there is no theoretical reason as to why there should be differences in the conceptualization of the measures in the current study based on such information. Therefore, participants who did not report demographic characteristics were retained. This robust sample is more than adequate to examine the conceptualization of team phenomena among project-based teams.

Part 1 Results: Examination of Psychometric Properties

Hypotheses 1 and 2 seek to confirm the factor structure at the team level of analysis (consistent with the theory of the constructs), provide psychometric evidence that the measure primarily captures team phenomenon, and highlight any measurement variance between the individual and the between-team levels of analysis.

Hypothesis 1: Metric Isomorphism and Measures of Common Team Constructs.

Hypotheses 1a – 1e test for metric isomorphism in measures of common team constructs via Steps 1 – 4 in an MCFA[6]. For simplicity, cross-level comparisons at the individual and the team levels of analysis are discussed in depth under their respective sub-hypotheses concerning model fit, factor loadings, residual variance, and reliability. Steps 2 and 3 in the MCFA are discussed across the measures of common team constructs before proceeding to the specific sub-hypotheses.

Step 2 calculated the variation in scores due to team membership in the sample via ICC(1); results are reported in Tables 10, 12, 14, 15, and 16. Before examining for a higher-level construct, there needs to be enough variation between teams before a factor structure can be estimated at the between-team level of analysis (B. O. Muthén, 1994). All measures' items, on average, had enough variation due to team membership in their respective samples ($\geq$ .10), except for team task interdependence (range of .05 - .08) and one item on the team psychological safety measure (= .08). While this ICC(1) is typically considered too low to investigate for a higher-level construct, an MCFA was conducted for the purpose of comparison in later analyses. Testing of Hypothesis 4, which separates the sample into high and low levels of agreement, will

6 As discussed in the Methods section, Step 5 in an MCFA models a two-level factor model. As these measures are designed to capture team phenomena, a one-level factor model at the team level of analysis modeling a between-team factor, results of Step 4 in an MCFA, were compared with those of Step 1, which models an individual-level factor.

examine the consequences of within-team agreement on the psychometric properties of the measure of team task interdependence.

Step 3 is not discussed in the respective sub-hypotheses as it is not directly related to the hypotheses. As expected, the results from Step 3 consistently reveal poor model fit, reduced factor loadings, increased residual variance, and lower reliability. Modeling a factor based on within-team variance provided evidence that a measurement model based on variation due to team membership was more suitable (see Tables 6 – 8 and 10 – 16).

Hypothesis 1a: Metric Isomorphism and Team Cohesion. Hypothesis 1a tests for cross-level variation at the individual and the team levels of analysis for the three-factor model of team cohesion (i.e., team task, relationship, and process conflict).

Model fit. The results indicate that, overall, there is consistency in estimating model fit for the individual-level factor structure (X^2 =5,972.72, p < .05; *CLI* = .94; *TLI* = .93; *RMSEA* = .10; and *SRMR* - .04) and the between-team factor structure in Step 4 of MCFA (X^2 = 2,962.17, p < .05; *CLI* = .96; *TLI* = .94; RMSEA = .12; and *SRMR* = .04). When comparing each proposed factor structure at the individual and the team levels of analysis, *CFI, TLI, RMSEA*, and *SRMR* did not vary more than .02. While the three-factor solution revealed the overall best fit at the between-team level of analysis, it missed the recommended threshold for *TLI,* which rewards more parsimonious structures, and *RMSEA*, which is more sensitive to badness of fit (see Table 6).

Factor loadings. The pattern (i.e., structure) of factor loadings at the individual and the team levels of analysis remained consistent. The strongest indicators (i.e., latent mean structure of the measure's items) at the team level of analysis also had greater factor loadings at the individual level (i.e., observed scores); across both levels the measure's indicators followed the same rank order (see Table 10). The magnitude of the factor loadings was greater at the team versus the individual level of analysis with a mean increase of .08. Therefore, model fit indices and the pattern and magnitude of factor loadings provided evidence for weak metric isomorphism.

Residual variance and reliability. Residual variance across all items was greater at the individual versus the team level of analysis, with a mean difference of .12 (see Table 11). As an estimate of composite reliability, (ω) is a function of residual variance. Cross-levels differences followed the same pattern as residual variance: ω ranged from .66 – .88 at the individual level and from .79 – .94 at the team level across all subscales (see Table 6). These cross-level differences are consistent with weak but not strong metric isomorphism.

Taken together, model fit indices, factor loadings, residual variance, and estimates of reliability for the measure of team cohesion provide sufficient evidence for weak metric isomorphism but not strong metric isomorphism. Thus, Hypothesis 1a was supported as team cohesion revealed a form of metric isomorphism.

Hypothesis 1b: Metric Isomorphism and Team Conflict. Hypothesis 1b proposes minimal cross-level variation at the individual and the team levels of analysis when examining the three-factor model of team conflict (i.e., interpersonal cohesiveness, task attraction, and commitment).

Model fit. The results are relatively consistent across model fit indices at the individual level as calculated in Step 1 of the MCFA (X^2 = 1842.00, p < .05; *CLI* = .98; *TLI* = .97; *RMSEA* = .04; and *SRMR* - .02) and Step 4, which examined the team factor structure (X^2 = 2092.22, p < .05; *CLI* = 98; *TLI* = .96; *RMSEA* = .09; and *SRMR* - .03). There was a cross-level difference in *RMSEA* (assessing badness of fit), which did not meet the recommended threshold of ≤ .06 when modeling the between-team factor structure. However, the general consistency in model fit indices at the individual and the team levels of analysis provides support for weak metric isomorphism.

Factor loadings. The pattern (i.e., structure) of factor loadings at the individual and the team levels of analysis remained consistent (see Table 12). While some items had equivalent factor loadings at the team but not at the individual level of analysis (e.g., items 4 and 5), this does not change the pattern (i.e., rank order) of loadings. Factor loadings were greater at the team level of analysis (an average of .09 increase across all items) versus the individual level. The consistency in the pattern of factor loadings combined with cross-level differences in the magnitude of factor loadings add additional support for weak metric isomorphism.

Residual variance and reliability. The average decrease in residual variance from the individual to the team level of analysis across all items was .14. This decrease accompanied an increase in ω from the individual to the between-team level of analysis across all subscales (task ω = .83, process ω = .82, and relationship ω = .85 at the individual level; task ω = .91, process ω = .89, and relationship ω = .92 at the team level).

These decreases in residual variance and increases in estimates of reliability for team conflict are evidence of weak metric isomorphism. Therefore, Hypothesis 1b (the presence of metric isomorphism) was supported.

Hypothesis 1c: Metric Isomorphism and Team Psychological Safety. Hypothesis 1c tests the unidimensional measure of psychological safety for metric isomorphism at the individual and the between-team levels of analysis.

Model fit. The model fit indices *CFI, TLI,* and *RMSEA* did not meet the recommended thresholds at the individual or the between-team levels of analysis (*CFI* = .85, *TLI* = .77, *RMSEA*

= .09 and *CFI* = .91, *TLI* = .86, *RMSEA* = .13, respectively), bringing into question the quality of this measure. There were moderate increases from the individual to the between-team level (.06 – .09) in the goodness of model fit indices (*CFI* and *TLI*) and minimal increases of .04 in the badness of model fit indices (*RMSEA* and *SRMR*). These contrary results reflect a small improvement in goodness of fit and worsening fit in the badness of fit indices. As many of the model fit indices did not meet the recommended threshold across levels of analysis, there were no meaningful cross-level differences in model fit, providing support for Hypothesis 1c of metric isomorphism.

Factor loadings. The factor loadings followed the same pattern at the individual and the team levels of analysis across the indicators; however, the individual-level loadings only moderately tapped the latent factor (range of .51 to .64) across the items (see Table 14)[7]. At the team level, factor loadings calculated from estimating the latent mean of each item revealed two items with a strong association y while the rest revealed a moderate association (λ = .61 – .73; see Table 14). The consistency in the pattern of factor loadings across levels and the mean difference across factor loadings from the individual to the team level of .10 provides support for Hypothesis 1c revealing weak metric isomorphism.

Residual variance and reliability. Residual variance was greater at the individual versus the team level of analysis (δ = .53 – .74 and ε = .46 – .63, respectively). Consistent with residual variance, ω was greater at the team level (.84) versus the individual level (.84), providing further support for Hypothesis 1c.

7 Research on drawing conclusions about the ability of items in a scale to tap a latent factor by examining the magnitude of factor loadings in a factor analysis is discussed in more depth in the Appendix B.

Taken together, psychological safety reveals weak metric isomorphism based on model fit indices, factor loadings, residual variance, and reliability. Therefore, Hypothesis 1c was supported.

Hypothesis 1d: Metric Isomorphism and Team Satisfaction. Hypothesis 1d tests for metric isomorphism at the individual and the team levels of analysis. The results of running an MCFA revealed that team satisfaction was underidentified and, therefore, unable to accurately estimate residual variance. I conducted an MCFA on the team satisfaction and task interdependence measures simultaneously to overidentify the parameters. The factor structure was specified as a two-factor model, with the team satisfaction items loading onto the first factor and the team task interdependence items loading onto the second.

Model fit. As detailed in Table 8, team satisfaction revealed good model fit across indices. While there were no differences in *CFI* and *TLI* at the individual and the team levels, there was a slightly worse fit at the individual versus the team level of analysis (*RMSEA* = .06 versus .08, *SRMR* = .05 versus .03). The inconsequential differences in model fit across levels provide support for metric isomorphism.

Factor loadings. The patterns of factor loadings remained consistent at the individual (λ = .92 – .95) and the team (λ = .96 – .97) levels of analysis. There was a modest mean increase from the individual to the team level of analysis ($\bar{x}$ = .03).

Residual variance and reliability. Concerning residual variances, the mean difference across items from the individual to the team level of analysis was .06. Composite reliability also improved slightly at the team versus the individual level (ω = .97 versus ω .95).

Therefore, the minimal cross-level differences in residual variance, factor loadings, and model fit for team satisfaction means that Hypothesis 1d (the presence of metric isomorphism) was supported.

Hypothesis 1e: Metric Isomorphism and Team Task Interdependence. A one-factor model was used to examine for metric isomorphism (via an MCFA) at the individual and the team levels of analysis. The model fit indices for *TLI* and *RMSEA* did not meet the recommended threshold at the individual or the team level (*TLI* = .89, *RMSEA* = .11 and TLI = .90, *RMSEA* = .12, respectively). Unlike the other measures where X^2 was consistently smaller at the team level of analysis, X^2 was greater at the individual versus the team level (X^2 = 1281.41 versus X^2 = 1531.22). This was likely a result of a low ICC(1) ranging from .05 – .08, which is also lower compared to the other measures. Examining the ICC(1) is part of Step 2 in an MCFA. If the ICC(1) is too low, a researcher should not continue with the rest of the steps as there is not enough meaningful variation to distinguish between group differences. Therefore, factor loadings, residual variance, and reliability were not further examined. There was insufficient evidence to support Hypothesis 1e.

Regarding the overarching Hypothesis 1, common measures of team constructs revealed weak metric isomorphism in that model fit, rank order of the factor loadings, residual variance, and reliability estimates were consistent from the individual to the team levels of analysis, except for team satisfaction, which had low ICC(1) values. This consistency across most measures means that Hypothesis 1 was partially supported (that is, common measures of team constructs reveal metric isomorphism).

Hypothesis 2: Factor Structure of Multidimensional Measures

Hypothesis 2 examines if the multidimensional structure of the measures of team cohesion and conflict hold when examined at the team level of analysis. The results compared theoretically relevant alternative measurement models at the team level of analysis that may not have been evident when examining the psychometric properties (e.g., model fit, factor loadings, residual variance, reliability, and correlations among subscales) of these measures at the individual level.

Hypothesis 2a: Between-team Level Factor Structure for Team Cohesion. While Hypothesis 1a examines for metric isomorphism, Hypothesis 2a tests if the three-factor structure holds at the team level of analysis when testing theoretically plausible alternative factor structures. That is, does psychometric evidence support a three-factor structure over theoretically alternative one- and two-factor models.

Model fit. At the team level, the three-factor model distinguished between task commitment, task attraction, and interpersonal cohesiveness; the two-factor model examined task-oriented and interpersonal cohesiveness as distinct factors; and the one-factor model represented an overall sense of team cohesion factor. In the three-factor model, model fit was superior across all indices compared to the one- or two-factor models (see Table 6). Only the three-factor model was close to or exceeded the recommended thresholds for the model fit indices (X^2 = 2,96.17, $p < .05$; *CFI* = .96; *TLI* = .94; *RMSEA* = .12; *SRMR* = .04). While all were statistically significant ($p < .05$), X^2 substantially improved with great model complexity (i.e., greater number of factors), which is typical for that index (1-factor = 7,390.26; 2-factor = 5,041.21; and 3-factor = 2,962.17). *SRMR*, which examined model misfit–(and not a function of

X^2) met the recommended threshold across all factor models (1-factor = .06, 2-factor = .06, 3-factor = .04). Overall, model fit indices supported a three-factor model of team cohesion.

Factor loadings. The magnitude of the factor loadings increased with the complexity of the factor structure: one-factor (λ = .52 – .89), two-factor (λ = .53 – .94), and three-factor (λ = .62 – .94). (See Table 11.) The increased factor loadings add further evidence for a three-factor model supporting Hypothesis 2a.

Residual variance and reliability. Following the same pattern, residual variance decreased with greater model complexity: one-factor (ε = .23 – .73), two-factor (ε = .12 – .72), and three-factor (ε = .12 – .62). (See Table 11.) Across all models, the residuals for the two reverse-coded items were substantially larger ($\varepsilon > .44$ versus $\varepsilon < .35$). These patterns provide additional support for a three-factor model. The estimates of composite reliability revealed that the one-factor model of team cohesion and the interpersonal cohesiveness and task attraction subscales had high reliability estimates (ω = .94, ω = .94, and ω = .91, respectively). The task commitment subscale, which contains the reverse-coded items, was lower (ω = .79). The ω for the combined subscales of task commitment and attraction representing the task-oriented factor was .89. The improved reliability in the two-factor model may address some of the problems associated with variance due to the reverse-coded items. The support for a three-factor model based on model fit, factor loadings, and residual variance is likely due to the task commitment scale containing reverse-coded items. To get a clearer picture of the measure of team cohesion's dimensionality, correlations among subscales were examined.

Correlations among factors. In the two-factor model of team cohesion at the individual level of analysis, the team interpersonal and task-oriented cohesion factors were heavily correlated (r = .89); in the three-factor model, latent factors were also heavily correlated (task attraction and task commitment: $r = .78$, task attraction and interpersonal cohesion: $r = .86$, and task commitment and interpersonal cohesion: $r = .78$). At the team level, the correlations among latent factors were even greater (two-factor model: team task-oriented and interpersonal cohesion: $r = .92$; three-factor model: task attraction and interpersonal cohesion: $r = .89$; task attraction and task commitment: $r = .85$; and interpersonal cohesion and task commitment: $r = .85$). The intercorrelations among latent factors exceeded the recommendations of Schmitt and colleagues (2018) of interfactor correlations being <.75. When examining the theoretically relevant level of analysis (i.e., team) for the measure of team cohesion, the high intercorrelations did not support a two- or a three-factor model.

There was not enough variation at the team level of analysis to investigate two or three distinct dimensions of team cohesion using the measure in the current study. Therefore, contrary to what was expected, Hypothesis 2a was not supported.

Hypothesis 2b: Between-team Level Factor Structure for Team Conflict. Hypothesis 2b examines if the psychometric properties at the between-team level of analysis for the measure of team conflict support three distinct factors (i.e., team task, relationship, and process conflict) over two-factor models (i.e., a team emotionally laden factor and team task conflict factor, or a team disagreements factor and team relationship conflict factor).

Model fit. The three-factor model revealed better model fit across all indices over both of the two-factor models in Step 4 of an MCFA that examined the team level of analysis (three-factor: $X^2 = 2429.50$, $p < .05$; $CFI = .98$; $TLI = .96$; $RMSEA = .09$; and $SRMR = .03$ versus the two-factor task and emotionally laden: $X^2 = 5178.78$, $p < .05$; $CLI = .94$; $TLI = .92$; $RMSEA = .14$; and $SRMR = .05$ or the two-factor disagreements and relationship: $X^2 = 7641.72$, $p < .05$; $CLI = .91$; $TLI = .88$; $RMSEA = .17$; and $SRMR = .05$). The two-factor model (task and emotionally laden) revealed better fit than the disagreements and relationship factor model. The former exceeded or came relatively close to the recommended cut-off values, with the exception of *RMSEA*, as opposed to the latter. Overall, the superior model fit in the three-factor model supports Hypothesis 2b.

Factor loadings. At the team level of analysis, the average magnitude of the factor loadings across the three- and the two-factor models of team conflict did not vary greatly (three-factor = .85, two-factor task and emotionally laden = .84, and two-factor disagreements and relationship = .87; see Table 12). As the magnitude of the factor loadings was similar, there was no evidence to support or reject the three-factor over the two-factor models of team conflict.

Residual variance and reliability. For the measure of team conflict, residual variance, on average, was smaller in the three-factor model (task, relationship, and process conflict: $\bar{x}$ = .23) versus the two-factor task and emotionally laden conflict model ($\bar{x}$ = .27) or the two-factor disagreements and relationship conflict model ($\bar{x}$ = .30). (See Table 11.) The reliability estimates were minimally better when combining subscales into a more general factor (task ω = .91, relationship ω = .92, and process conflict: ω = .89 versus emotionally laden conflict ω = .94 and disagreements ω = .92). The minimal differences across the different factor models did not clearly support one factor model over another for the measure of team conflict.

Correlations among factors. Across the multifactor models of team conflict, the intercorrelations among latent factors were high. At the individual level, the two-factor models were strongly correlated (emotionally laden and task conflict: *r* = .80, team-focused disagreements and relationship conflict: *r* = .83); the three factors were highly correlated (relationship and process conflict: *r* = .84, relationship and task conflict: *r* = .72, and process and task conflict: *r* = .80). At the team level, the two-factor models were strongly correlated (emotionally laden and task conflict: $r = .81$ and team focused disagreements and relationship conflict: $r = .87$). The three-factor model revealed greater distinction among factors (relationship and process conflict: $r = .82$, relationship and task conflict: $r = .36$, and process and task conflict: $r = .67$).

While the intercorrelations in the two-factor model of emotionally laden and task conflict exceeded the recommended threshold of $r < .75$, the low intercorrelation for relationship and task conflict in the three-factor model provided additional support that task conflict is psychometrically distinguishable from relationship conflict (Schmitt et al., 2018). Based on these psychometric properties at the team level of analysis, there was greater evidence for a distinct team task conflict factor but not for a distinct process and relationship conflict factor at the team level of analysis. As there was more evidence to support a two-factor model of team conflict, Hypothesis 2b was not supported.

Part II Results: Influences of Psychometric Isomorphism and Relationships Among Team Constructs

Hypothesis 3: Psychometric Isomorphism and a Measure's Characteristics

Hypothesis 3a: Influence of the Referent. The influence of a measure's referent on its degree of psychometric isomorphism was examined by comparing the model fit, factor loadings, and residual variances across levels of analysis for measures intended to capture team phenomenon in a referent-shift versus direct consensus model. While Hypothesis 1 examined for evidence of psychometric isomorphism via model fit indices, Hypothesis 3a examined for systematic differences in model fit indices at the individual and the team levels of analysis and if these differences are the result of a measure's referent. There were no consistent differences in model fit based on the level of analysis in the measures of common team constructs.

Concerning the factor loadings, measures of team cohesion and conflict revealed that the team level of analysis had consistently greater factor loadings; increases ranged from .05 to .16, with a mean increase of .09. This mean increase is slightly lower than expected (i.e., ≥ .10). However, the measures of team satisfaction and task interdependence showed much smaller increases from the individual to the team level as expected, ranging from .03 to .05 among their items, with an average increase across items of .04.

As there is less literature on differences in residual variance across levels of analysis in a measure, the current study did not propose a specific range or cutoff scores to compare the measures in a referent-shift versus a direct consensus model. Therefore, these results were examined for more general patterns. The measures (i.e., team cohesion, conflict, and psychological safety) in a referent-shift model's residual variance across items decreased at the team level from the individual level, with a range from .08 to .17 across all proposed factor structures. The measures (i.e., team satisfaction and task interdependence) in a direct consensus model's residual variance also decreased at the team level from the individual level, with a range from .04 to .07. There was a mean decrease across items of .06 for team satisfaction and -.05 for team task interdependence.

Directly related to a measure's residual variance, composite reliability increased consistently across all measures at the between-team versus the individual level of analysis. However, there were no meaningful differences between measures in a referent-shift versus a direct consensus model. (See Tables 6 – 8.) Therefore, based on a lack of meaningful cross-level differences from the results of the MCFA, Hypothesis 3a was not supported.

Hypothesis 3b: Influence of the Target. The influence that a measure's target has on its degree of psychometric isomorphism was examined in a direct consensus model (i.e., team satisfaction, task interdependence, general liking among team members, team viability). Team satisfaction and task interdependence had a "team" target; that is, items referred to members' personal perceptions about working in their team. General liking and team viability had a "team-member" target; that is, items referred to members' personal perceptions about working with specific members of their team. Unfortunately, the sample in the current study for the measure of team task interdependence had a low ICC(1); so, examining the degree of psychometric isomorphism was not warranted. The measure of team satisfaction, which had a "team" target, revealed metric isomorphism, where model fit indices, factor loadings, and residual variance were consistent at the individual and the team levels (see Tables 8, 15, and 16).

Measures of liking and viability were examined by comparing the within-person (modeling the deviation construct), the between-team (modeling the shared construct), and the three-level multilevel (modeling the shared configural construct) factor models (see Figure 8).[8] ICC(1) scores showed that at least 10% of the variance was due to between-team variance for all items which supported considering team-level phenomena. Model fit indices at the within-person, between-team, and multilevel factor models revealed good fit, with the exception of *RMSEA* (an index of badness of fit) at the team level via CFA on the S_{BT} (*RMSEA* = .15). The difference in factor loadings ranged from .03 – .20, with a mean difference of .07 across the items comparing the within-person and the between-person factor models. However, the reverse-coded item had more variance across levels (.20) than the other items (.03 – .08). The difference in residual variance ranged from .01 – .22, with a mean of .07 across the items. The reverse-coded item revealed greater residual variance than the other items (.22 versus a range of .01 –

8 While a one-factor model for a general feeling about team members was examined, consistent with previous research a two-factor model was supported, with liking and viability items loading onto distinct factors. It is discussed in the results from this point forward. See Tables 9, 17, and 18.

.06). The residual variance remained when comparing the CFA on the S_{WP} and that estimated at the within-person level in a multilevel factor model (δ = .10 – .21 and ε = .09 – .21, respectively). However, the residual variance from the CFA on the S_{BT} and that estimated at the between-team level in a multilevel factor model varied greatly (.10 – .28).

These results support a within-person construct, which is consistent with the theoretical development of the measure, but do not support a shared construct model of general team liking and viability. (See Figure 4.) However, modeling a multilevel factor structure with meaningful variance at the within-person, the between-person, and the team levels for the measures of liking and viability was supported. (See Figure 9.) As hypothesized, measures with a “team” target revealed a stricter degree of psychometric isomorphism across theoretical relevant levels of analysis (i.e., individual and team); measures with a “team-member” target revealed measurement variance in the pattern and factor loadings, residual variance across levels, and a poorer fit at the team level when not modeling them as multilevel factors. However, the measure of team task interdependence was unable to be included in the analysis due to a low level of agreement. Therefore, Hypothesis 3b was partially supported.

Hypothesis 4: Psychometric Isomorphism and Team-member Agreement

Splitting the data between high and low team-member agreement by $r_{WG}(j)$ resulted in a smaller sample for low agreement versus the high agreement group across all measures. Splitting the groups by average deviation from median score across all items in the measure (*ADmd*) resulted in relatively even groups according to sample size. (See Tables 25 – 31.) Across all measures split by high and low $r_{WG}(j)$, none revealed metric measurement invariance; the measures of team cohesion and conflict failed to converge when testing for configural invariance. Upon further inspection, the one-factor model for team cohesion and the two-factor model for team conflict was not identified in the low $r_{WG}(j)$ groups. To inspect the severity regarding the lack of measurement invariance across low and high $r_{WG}(j)$ groups, the results restricting the residuals to be equal across groups (i.e., strict measurement invariance) caused the largest decreases in *CFI* (.03 – .31) across all measures. As team-member scores should be theoretically interchangeable when measuring team phenomena, greater residual variance in the low agreement groups indicated greater residual error as team-members failed to coalesce on a shared understanding of their team.

Splitting the samples by high and low *ADmd* had slightly different results. The high and low *ADmd* groups for the measure of team cohesion were scalar-measurement invariant. The measures of team satisfaction, task interdependence, and psychological safety did not reveal metric invariance but did reveal scalar invariance. While scalar invariance is a stricter form of measurement invariance than is metric, the high and low groups were not statistically different. This is likely due to the increases in degrees of freedom from the metric to the scalar invariance and the fact that team intercepts did not differ greatly in either group. Therefore, minor differences in factor loadings became inconsequential with the increase in degrees of freedom. As with $r_{WG}(j)$, for high and low *ADmd* groups, the range of differences in *CFI* from the configural model when restricting factor loadings, intercepts, and residuals (i.e., strict invariance) ranged from .00 – .12 versus the other forms (i.e., metric and scalar invariance), which differed by < .02. As the factor models across measures were not comparable across groups at the team level of analysis, further investigation into the degree of psychometric isomorphism between high and low agreement groups at the individual level of analysis was unwarranted.

The differences in the *ADmd* groups were less meaningful than the $r_{WG}(j)$ groups. The differences in the $r_{WG}(j)$ groups were consequential, as the factor models were either not identified among teams with low agreement or the other psychometric properties (i.e., factor loadings, residual variance) of the measures differed significantly. Therefore, Hypothesis 4 was partially supported based on teams with low agreement via r_{WG} *(j)* values.

Research Questions: Relationships Among Variables

Two research questions examined relationships among variables at the individual and the team levels of analysis. To estimate these relationships, a subset of the current study's sample was used in which each team was administered all measures (*n* = 2,332 individuals, 541 teams). The only substantive difference between the larger sample and its subset in the initial hypotheses and the research questions was that the ICC(1) values were substantially higher across all items for every measure, ranging from .23 – 44. While the results are not reported, an MCFA was conducted specifying all latent factors found in earlier MCFA results using the subset. There were no meaningful differences in the psychometric properties across levels of analysis between the larger sample and its subset.

Research Question 1: Psychometric Isomorphism as a Constraining Force. All measures in the current study were categorized by the degree of psychometric isomorphism followed by a comparison of the covariance/variance matrices across levels of analysis. Concerning the degree of psychometric isomorphism, all measures of common team constructs had the same number of factors and were consistent in the pattern of zero and nonzero factor loadings at the individual and the team levels of analysis, giving support for strong configural isomorphism (see Tables 6 – 8 and 10 - 16). For the measures of team cohesion and conflict, the relative ordering of factor loadings remained consistent across levels of analysis. The magnitude was greater at the team versus the individual level of analysis, providing support for weak metric isomorphism.

For the measures of team satisfaction and task interdependence, the relative ordering remained consistent across the individual and the team levels. The minimal increase (< .03) suggests that these measures are best characterized by strong metric isomorphism. Regarding the measure of team psychological safety, many of the model fit indices did not meet the recommended cut-off criteria (e.g., *CFI*, *TLI*, *RMSEA*). The relative ordering of the factor loadings at the individual and the team levels was not consistent; however, the magnitude of factor loadings was greater at the team versus the individual level of analysis. Therefore, this measure did not adequately meet the quality standards in a CFA framework to categorize its degree of psychometric isomorphism with this sample.

For the measures of team-member liking and viability, general model fit indices were good for a two-factor model at the within-person (*CFI* = 1.00, *TLI* = 1.00, *RMSEA* = .04, SRMR = .01) and the team levels of analysis (*CFI* = .96, *TLI* = .95, *RMSEA* = .15, *SRMR* = .04), which supports strong configural isomorphism. A three-level multilevel factor model was supported as well (*SRMR-wp* = .02, *SRMR*-bp = .06, *SRMR-bt* = .06; see Table 9). The pattern and the relative ordering of factor loadings remained consistent at the within-person and the team levels of analysis; however, the magnitude of factor loadings was greater at the team level. This pattern followed suit in the multilevel factor model as well. Therefore, there was sufficient evidence for weak metric isomorphism in measures of team-member liking and viability.

The covariance/variance matrices at the team level of analysis via S_B and the MCFA differed slightly but followed the same pattern regarding model fit and factor loadings due to how residual variation was modeled in the respective factor structures (see Figure 4). As the current study primarily examined for shared constructs, the results focus on the differences in relationships among constructs at the individual (via S_T) versus the team (via S_B) levels of analysis. Relationships among variables differed greatly at the individual and the team levels of analysis, with consistently stronger correlations at the team level. However, there was no pattern of differences based on whether the measures revealed weak or strong metric isomorphism. For example, cohesion correlated differently at the individual versus the team level for emotionally laden conflict (*r* = -.58 versus -.72) and team viability (*r* = .79 versus .87). Also, the differences in satisfaction correlates at the individual versus the team level of analysis were smaller for liking (*r* = .66 versus .70) and task interdependence (*r* = .37 versus .40).

Research Question 2: Psychometric Isomorphism and Discriminant Validity. The second research question examined for evidence of discriminant validity by comparing relationships among variables at the individual and the team levels of analysis. Based on the observed scores at the individual level of analysis, six relationships among variables were greater than .75 (emotionally laden and task conflict: *r* = .80, cohesion and satisfaction: *r* = .84, cohesion and viability: *r* = .79, cohesion and liking: *r* = .77, satisfaction and viability: *r* = .85, viability and liking: *r* = .77). Ten correlations at the team level of analysis exceeded the recommended threshold for correlations among distinct variables (emotionally laden and task conflict: *r* = .82, cohesion and satisfaction: *r* = .91, cohesion and viability: *r* = .85, cohesion and liking: *r* = .82 satisfaction and viability: *r* = .89, viability and liking: *r* = ..79, psychological safety and cohesion: *r* = .85, satisfaction and psychological safety: *r* = .82, cohesion and psychological safety: *r* = .85, viability and psychological safety: *r* = .82.). At the team level there were three relationships among variables in which the residual error in comparison to the magnitude of correlation revealed a lack of discriminant validity, (see Table 22.), which can be further understood by examining interitem relationships (Prudon, 2015). As seen in Tables 23 and 24, the interitem correlations among measures of team constructs reveal that modeling individual perceptions resulted in different correlations among measures' items compared to the team level, which modeled the influence of team membership. Therefore, there were differences in the relationships among measures at the individual and the team levels of analysis.

DISCUSSION

The overarching purpose of the current study is to connect measurement theory and theory on team phenomenon operating in a multilevel context. First, using archival data, I applied a multilevel factor analytic framework to measures of common team constructs to examine the internal properties of these measures via an empirical study at the individual and the team levels of analysis. Second, I examined what influences a measure's psychometric properties across levels of analysis and how relationships among measures of team constructs may vary across levels.

Part 1 Discussion: Psychometric Isomorphism in Measures of Team Constructs

To address the lack of clarity on the psychometric properties of common measures of team constructs discussed in the literature review, Hypotheses 1and 2 evaluated the degree of psychometric isomorphism in common measures of team constructs by categorizing the differences in model fit, factor loadings, and residual variance at the individual and the team levels of analysis and the factor structure at the team level of analysis in multidimensional measures. With the exception of the measure of team task interdependence, which was not examined due to a low ICC(1), all measures revealed some degree of metric isomorphism in that model fit estimates led to similar conclusions at both levels. However, the factor loadings were higher and residual variance was lower at the team versus the individual level of analysis across all measures, while estimates of reliability were higher at the team level across all measures. These difference across levels were smaller for team satisfaction; therefore, team satisfaction is best characterized as having revealed strong metric isomorphism while measures of team conflict, cohesion, and psychological safety revealed weak metric isomorphism.

There are two consequences of modeling variance in scores due to individual differences typical in a traditional CFA as opposed to modeling variance due to team membership. First, the measure's estimate of reliability was downwardly biased. Second, differing factor loadings and residual variance across levels means that the relationships among dimensions in the measures of team cohesion and conflict varied at the team versus the individual level. For team cohesion, the correlations increased from the individual to the team level of analysis to the extent that the dimensions were practically indistinguishable at the team level. Concerning the measure of team conflict, the strong correlation between relationship and process at the team level supported a two-factor solution, with task conflict on the second factor. These results are consistent with those of researchers who found simplified factor structures at higher levels of analysis and biased estimates at lower levels of analysis (Dedrick & Greenbaum, 2011; Dyer et al., 2005; Hsu et al., 2015).

Part 2 Discussion: Reasons and Consequences for Varying Degrees of Psychometric Isomorphism

Hypotheses 3a and 3b investigated if characteristics of measures, such as their referent and target, influenced their degree of psychometric isomorphism. Common measures of team constructs in a referent-shift versus a direct consensus model did not show systematic differences across levels of analysis. (See Tables 6 – 8.) Using "team" as a measure's referent versus "I" revealed only small differences in the item's factor loadings between the two models, and not to the extent hypothesized. Therefore, it is unlikely that a measure's referent influenced the degree of psychometric isomorphism when the context of the items was a team.

While measures in the direct consensus model in the current study all used "I" as the referent, they can be characterized further by their target (e.g., team-member, team). Measures using a "team-member" target (i.e., team-member liking and viability) revealed greater measurement variance in model fit, factor loadings, and residual variance than measures using a "team" target (i.e., team satisfaction) from the individual to the team level of analysis. This cross-level variance arose from not modeling meaningful variance due to distinct relationships among team members (i.e., within-person) in a traditional CFA. Therefore, when lower- and higher-level variance is of theoretical interest, it needs to be modeled to accurately assess the psychometric properties of the measure.

Extending beyond measures' characteristics, Hypothesis 4 examined if team-member agreement influenced a measure's degree of psychometric isomorphism. By including teams with low agreement via $r_{WG}(j)$, measurement quality was underestimated and the ability to model the factor structure was constrained by the number of teams with low agreement in a sample. The difference was not as strong when using *ADmd* because splitting the sample based on an average included teams with sufficient team-member agreement. Therefore, low agreement should be estimated with $r_{WG}(j)$, not *ADmd*.

While ICC(1) estimates addressed the overall ability of the sample to measure for a higher-level construct, teams with very low agreement should still be eliminated from the sample to improve the measure's ability to detect variation in the latent factor by reducing the residual variance. Low agreement is problematic because team-member perceptions describing the attributes, beliefs, characteristic, and cognitions of the team should be relatively interchangeable; variation in team-member perceptions can cause residual error. Including teams with very low agreement contradicts current understanding of team consensus constructs and introduced unnecessary error in the measurement of team constructs in the sample. Researchers should remove teams with low $r_{WG}(j)$ before further investigating the respective team construct.

Including teams with low agreement via $r_{WG}(j)$ in analyses with higher-level constructs may lead to misleading results, as the measures' quality is underestimated by introducing error via team-member lack of agreement.

A final consideration regarding psychometric isomorphism is the differentiation among variables across levels of analysis, which was investigated via the research questions. For the first question, there were no systematic differences in the correlations among variables based on their degree of psychometric isomorphism, perhaps because the measures revealed weak to strong metric isomorphism at the individual and the team levels of analysis.

The second question compared the differences in the correlations among variables, in general, at the individual and the team levels of analysis. There were differences in relationships among latent variables at the between-team level, which is problematic as ten relationships among latent variables failed to provide evidence for discriminant validity (i.e., correlations greater than .75; Schmitt et al., 2018). A lack of differentiation among variables can occur for multiple reasons. First, the dynamic team environment makes it difficult to distinguish between measures, as variables continue to influence each other over performance cycles. Second, team members may report on general positive and negative feelings of their team, making it difficult to tease apart theoretically distinct team phenomena. If positive/negative feelings about a team are driving differences among measures of team constructs, then researchers need to refine these measures like researchers did in order to tease apart the difference between positive and negative affect and other individual differences (Weiss, 2002).

Theoretical Implications

The current study contributes to measurement theory by demonstrating how to correctly model consensus constructs based on the theory of the construct in an MCFA framework. By linking theory on team phenomena and Classic Test Theory in a SEM framework, the current study clarifies the appropriate measurement model for common measures of team constructs and gives recommendations for when researchers should adopt alternative models (Crocker & Algina, 1986; Ilgen et al., 2005; Marks et al., 2001).

By linking measurement and theory, this study addressed the concerns of Chen and colleagues (2005) about the lack of understanding regarding how various statistical packages standardize variables at the within and the between levels of analysis in a SEM framework. This clarification was accomplished by detailing best practices for conducting an MCFA and reporting results, conducting analyses in both *R* (via the lavaan package) and *MPlus*, and linking the analysis and interpretation of results to the theory of the construct and the measure's measurement model.

Third, regarding theory on team phenomena, this study questioned if there are substantial differences in relationships among variables at different levels of analysis for measures of team constructs. Differences were found and the high correlations could be a result of problematic measures or construct proliferation not detected because of a misalignment between measurement and theory. Because team variables such as team cohesion, psychological safety, and satisfaction showed strong correlations, team psychological safety and satisfaction may be intertwined, and their strongly correlated relationship at later stages in team development may be a result of the cyclical manner in which team constructs influence each other. However, the story is less clear on team cohesion. Are teams inherently cohesive when they are satisfied or feel psychologically safe to interact? Are teams that want to work together again (i.e., team viability) or that like each other (i.e., team liking) really distinct from teams that cohesively work together?

Researchers need to investigate the extent to which team cohesion is a distinct construct or one that simply taps more broadly into a general positive feeling about the team.

Practical Implications

This study provides researchers with the tools and knowledge to investigate team phenomena in both *R* and *MPlus* by providing MCFA syntax for both software packages (see Appendix B). Consistency in results is vital across statistical software. The current study describes the analytical choices in depth to give researchers the ability to conduct their own analyses when dealing with multilevel data.

While the consequences of misalignment in theory and measurement are evident when evaluating the quality of measures for team constructs, the focus is understanding the true quality of these measures. To examine measures via an MCFA is not always feasible due to the large sample size requirement/recommendation (e.g., ≥ 100 teams and 3 observations per team; B. O. Muthén, 1994). Researchers should initially conduct an MCFA when theoretically necessary during the validation process of measure development. Once the degree of psychometric isomorphism in the measure is established (e.g., configural, metric), the consequences for examining the quality of the measure at the lower level of analysis can be estimated. Therefore, the likelihood of upward or downward biased estimates can be noted, which will limit the need for a large sample size across all studies using multilevel data and allow researchers to move forward in their understanding of the dynamic team environment.

While not a main focus of the current study, the results in the MCFA highlight differences across levels of analysis and measurement models. Researchers can use this knowledge to investigate deviation, shared, and multilevel constructs with more confidence and compare their results to those in the current study. For example, researchers can compare MCFA for measures designed to capture deviation constructs with the results here that focus on measures in a shared construct model. Therefore, the current study contributes to the literature by providing results that can be compared across multilevel factor analytic studies.

Limitations and Future Research

The current study's sample population consisted of student work groups examined over a 15-week time span. While there is no theoretical reason for the factor structure of team variables to vary in university student work groups and professional work groups, it should be noted as a potential limitation since subpopulations can differ. This population was based in the United States. Because cross-cultural differences regarding team conflict and team-related outcomes may exist (Wit et al., 2012), researchers should also investigate (via MCFA) how different cultures view team phenomena such as cohesion, conflict, and psychological safety.

Changes across the stages of team development need to be addressed. As the measures of team constructs in the current study were inherently derived from a reflective perception of people's experiences, the factor structure was examined in established teams. Team emergent states that are more susceptible to status cues, such as team psychological safety (Nembhard & Edmondson, 2006), can be antecedents to other emergent states that may take longer to establish through interactions in project-based work. Differences across stages of team development in how team constructs are measured and evaluated means that strongly correlated variables at later stages of team development may not be theoretically relevant to compare in early stages. Therefore, researchers should disentangle the idea of construct proliferation and the theoretical value of distinct constructs at different stages of team development.

A longitudinal approach in which variation in latent factors is modeled over time to assess discriminant validity represents another research opportunity. Such an examination of constructs is an important piece of measures' validation process. For example, resilience, optimism, and hardiness are highly correlated but have distinct relationships with outcomes as they tap into slightly different aspects of the human condition (Lee et al., 2011). Therefore, researchers need to investigate whether measures of common team constructs tap into a higher-order factor or if they provide meaningful information on the dynamic team environment across the stages of team development.

Conclusion

The current study addressed six overarching questions regarding consequences of misalignment in theory and measurement in common measures of team construsts resulting in six key takeaways. First, current practices were found to be insufficient as most researchers did not fully examine the nature and structure of measures at the team level of analysis. Second, overall, these measures revealed to have metric isomorphism. Third, there was more support for simpler factor structures at the team level for the multidimensional measures of team cohesion and conflict. Fourth, regarding measures' characteristics, there was no evidence that a measure's referent influenced its degree of psychometric isomorphism; however, there was some evidence that a measure's target influences the degree of psychometric isomorphism across levels of analysis. Fifth, findings also suggest that the level of team agreement constrains the ability to accurately model the latent factor. Sixth, relationship among team variables must be examined at the theoretically relevant level of analysis as the relationships vary from the individual to the team level of analysis.

The current study fills important gaps in our knowledge by providing best practices and guidelines for how to evaluate measures of team constructs in a multilevel context and gives practical examples of how to conduct an MCFA (using both *R* and *MPlus*) that aligns the theory of the construct with its measurement. Failure to accurately model the factor structure in measures of common team constructs leads to misleading results in how team variables relate to each other and in the overall quality of the measure. In other words, the Gestalt idiom applies to the current study since the whole (i.e., team) is greater than the sum of its parts (i.e., team members; Islam et al., 2006). This study lays the groundwork for researchers who seek to understand higher-level team phenomena by addressing practical and theoretical considerations and challenges that arise when working with multilevel data.

REFERENCES

Arthur, W., Bell, S. T., & Edwards, B. D. (2007). A longitudinal examination of the comparative criterion-related validity of additive and referent-shift consensus operationalizations of team efficacy. *Organizational Research Methods*, *10*(1), 35–58. https://doi.org/10.1177/1094428106287574

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, *41*(6), 1561–1577. https://doi.org/10.1177/0149206315591075

Bagozzi, R. P., Davis, F. D., & Warshaw, P. R. (1992). Development and test of a theory of technical learning and usage. *Human Relations*, *45*(7), 659–686.

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, *88*(6), 989–1004. https://doi.org/10.1037/0021-9010.88.6.989

Bentler, P. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual*. Sage Publications.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klien & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Jossey-Bass Inc.

Bliese, P. D., Chan, D., & Ployhart, R. E. (2007). Multilevel methods: Future directions in measuremnet, longitudinal analyses, and nonnormal outcomes. *Organizational Research Methods*, *10*(4), 551–563. https://doi.org/10.1177/1094428107301102

Boies, K., & Howell, J. M. (2006). Leader-member exchange in teams: An examination of the interaction between relationship differentiation and mean LMX in explaining team-level outcomes. *Leadership Quarterly*, *17*(3), 246–257. https://doi.org/10.1016/j.leaqua.2006.02.004

Bollen, K. A., & Arminger, G. (1991). Observational Residuals in Factor Analysis and Structural Equation Models. *Sociological Methodology*, *21*, 235–262. https://doi.org/10.2307/270937

Braun, M. T., Kozlowski, S. W. J., Brown, T. A., & DeShon, R. P. (2020). Exploring the dynamic team cohesion–performance and coordination–performance relationships of newly formed teams. *Small Group Research*, *On-line*, 1–30. https://doi.org/10.1177/1046496420907157

Breevaart, K., Bakker, A. B., Demerouti, E., & Hetland, J. (2012). The Measurement of state work engagement: A multilevel factor analytic study. *European Journal of Psychological Assessment*, *28*(4), 305–312. https://doi.org/10.1027/1015-5759/a000111

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). American Psychological Association.

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods*, *2*(1), 49–68.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Can, S., van de Schoot, R., & Hox, J. (2015). Collinear latent variables in multilevel confirmatory factor analysis: A comparison of maximum likelihood and bayesian estimations. *Educational and Psychological Measurement*, *75*(3), 406–427. https://doi.org/10.1177/0013164414547959

Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. *Small Group Research*, *31*(1), 71–88.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of compostition models. *Journal of Applied Psychology*, *83*(2), 234–246. https://doi.org/10.1037/0021-9010.83.2.234

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, G., Bliese, P. D., & Mathieu, J. E. (2005). Conceptual framework and statistical procedures for delineating and testing multilevel theories of homology. *Organizational Research Methods*, *8*(4), 375–409. https://doi.org/10.1177/1094428105280056

Chen, G., Mathieu, J. E., & Bliese, P. D. (2004). A framework for conducting multilevel construct validation. In F. J. Yammarino & F. Dansereau (Eds.), *Multi-level issues: Multilevel issues in organizationoal behavior and processes* (pp. 273–303). Elsevier Ltd.

Cole, M. S., Bedeian, A. G., Hirschfeld, R. R., & Vogel, B. (2011). Dispersion-composition models in multilevel research: A data-analytic framework. *Organizational Research Methods*, *14*(4), 718–734. https://doi.org/10.1177/1094428110389078

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart, and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

D'Haenens, E., van Damme, J., & Onghena, P. (2012). Constructing measures for school process variables: The potential of multilevel confirmatory factor analysis. *Quality and Quantity*, *46*(1), 155–188. https://doi.org/10.1007/s11135-010-9346-4

Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental Health Agencies. *Journal of Emotional and Behavioral Disorders*, *19*(1), 27–40. https://doi.org/10.1177/1063426610365879

Dunlap, W. P., Burke, M. J., & Smith-Crowe, K. (2003). Accurate tests of statistical significance for rWG and average deviation interrater agreement indexes. *Journal of Applied Psychology*, *88*(2), 356–362. https://doi.org/10.1037/0021-9010.88.2.356

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *Leadership Quarterly*, *16*(1), 149–167. https://doi.org/10.1016/j.leaqua.2004.09.009

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

Festinger, L., Back, K., Schachter, S., Kelly, H. H., & Thibaut, J. (1950). *Theory and experiment in social communication*. Research Center for Dynamics Institute for Social Research, University of Michigan. https://isr.umich.edu/wp-content/uploads/historicPublications/Theory_4867_.PDF#page=7

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*(2), 186–192. https://doi.org/10.2307/3172650

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47*(1), 307–338. https://doi.org/10.1146/annurev.psych.47.1.307

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, *33*(6), 987–1015. https://doi.org/10.1177/0149206307308587

Hox, J. J., & Maas, C. J. M. (2001). *The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples*. *8*(2), 154–174. https://doi.org/10.1207/S15328007SEM0802_1

Hsu, H. Y., Kwok, O. man, Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research*, *50*(2), 197–215. https://doi.org/10.1080/00273171.2014.977429

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, F. L. (2017). Conducting Multilevel Confirmatory Factor Analysis Using R. *Working Paper, August*, 1–20. https://doi.org/10.13140/RG.2.2.12391.34724

Huang, F. L., Cornell, D. G., Konold, T., Meyer, J. P., Lacey, A., Nekvasil, E. K., Heilbrun, A.,

& Shukla, K. D. (2015). Multilevel factor structure and concurrent validity of the teacher version of the authoritative school climate survey. *Journal of School Health*, *85*, 843–851.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in Organizations: From Input-Process-Output Models to IMOI Models. *Annual Review of Psychology*, *56*(1), 517–543. https://doi.org/10.1146/annurev.psych.56.091103.070250

Islam, G., Zyphur, M. J., & Beal, D. J. (2006). Can A Whole Be Greater Than The Sum Of Its Parts? A Critical Appraisal Of "Emergence. In *Insper Working Paper* (Issue May 2014). http://www.ibmecsp.edu.br/pesquisa/download.php?recid=3068

Jak, S. (2019). Cross-Level Invariance in Multilevel Factor Models. *Structural Equation Modeling*, *26*(4), 607–622. https://doi.org/10.1080/10705511.2018.1534205

James, L. R., Demaree, R. G., & Wolf, G. (1993). rwg: An Assessment of Within-Group Interrater Agreement. *Journal of Applied Psychology*, *78*(2), 306–309. https://doi.org/10.1037/0021-9010.78.2.306

Jehn, K. A. (1997). A Qualitative Analysis of Conflict Types and Dimensions in Organizational Groups. *Administrative Science Quarterly*, *3*(3), 530–557.

Jehn, K. A., Greer, L., Levine, S., & Szulanski, G. (2008). The effects of conflict types, dimensions, and emergent states on group outcomes. *Group Decision and Negotiation*, *17*(6), 465–495. https://doi.org/10.1007/s10726-008-9107-0

Jehn, K. A., & Mannix, E. A. (2001). The Dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *The Academy of Management Journal*, *44*(2), 238–251. https://www.jstor.org/stable/3069453

Johnson, R. E., Rosen, C. C., & Chang, C.-H. (2011). To aggregate or not to aggregate: Steps for developing and validating higher-order multidimensional constructs. *Journal of Business* *and Psychology*, *26*(3), 241–248.

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, *51*(6), 881–898. https://doi.org/10.1080/00273171.2016.1228042

Kozlowski, S. W. J., & Bell, B. S. (2013). Work groups and teams in organizations: Review update. In N. Schmidt & S. Highhouse (Eds.), *Handbook of Psychology: Vol. 12. Industrial and organizational psychology* (2nd ed., pp. 412–469). Wiley. https://doi.org/10.1002/0471264385.wei1214

Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, *52*(2), 161–193.

Lee, J. E. C., Sudom, K. A., & McCreary, D. R. (2011). Higher-order model of resilience in the Canadian Forces. *Canadian Journal of Behavioural Science*, *43*(3), 222–234. https://doi.org/10.1037/a0024473

Li, F., Duncan, T. E., Harmer, P., Acock, A., & Stoolmiller, M. (1998). Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches. *Structural Equation Modeling*, *5*(3), 294–306. https://doi.org/10.1080/10705519809540106

Loughry, M. L., & Tosi, H. L. (2008). Performance implications of peer monitoring. *Organization Science*, *19*(6), 876–890. https://doi.org/10.1287/orsc.1080.0356

Lubke, G. H., & Dolan, C. V. (2003). Can unequal residual variances across groups mask differences in residual means in the common factor model? *Structural Equation Modelling*, *10*(2), 175–192. https://doi.org/10.1207/S15328007SEM1002

Malhotra, N. K., Lopes, E. L., & Veiga, R. T. (2014). Structural equation modeling with use of

Lisrel: An initial vision. *Revista Brasileira de Marketing*, *13*(2), 28–43. https://doi.org/10.5585/remark.v13i2.2698

Marks, M. A., Mathieu, J. E., & Zaccaro, S. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356–376. https://doi.org/10.5465/AMR.2001.4845785

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size. *Psychological Bulletin*, *103*(3), 391–410. https://doi.org/10.1037/0033-2909.103.3.391

Marsh, H. W., & Grayson, D. (1994). Longitudinal confirmatory factor analysis: Common, time-specific, item-specific, and residual-error components of variance. *Structural Equation Modeling: A Multidisciplinary Journal*, *1*(2), 116–145. https://doi.org/10.1080/10705519409539968

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

Mok, M. (1995). Sample size requirements for 2-level designs in educational research. *Multilevel Modeling Newsletter*, *7*(2), 11–15.

Morgeson, F. P., & Hofmann, D. a. (1999). The structure and function of collective constructions: implications for multilevel research and theory development. *Academy of Management Review*, *24*(2), 249–265. https://doi.org/10.2307/259081

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *1*(15), 210–227. https://doi.org/10.1037/0033-2909.115.2.210

Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41. https://doi.org/10.1177/1088868313501745

Muthén, B. O. (1994). Mutlivariate covariance structure analysis. *Sociological Methods & Research*, *22*(3), 376–398.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén. http://www.ncbi.nlm.nih.gov/pubmed/22210541

Nembhard, I. M., & Edmondson, A. M. Y. C. (2006). Making it safe : The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior*, *27*(June), 941–966. https://doi.org/10.1002/job

O'Neill, T. A., McLarnon, M. J. W., Hoffart, G. C., Woodley, H. J. R., & Allen, N. J. (2018). The structure and function of team conflict state profiles. *Journal of Management*, *44*(2), 811–836. https://doi.org/10.1177/0149206315581662

Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring Clustering in Confirmatory Factor Analysis: Some Consequences for Model Fit and Standardized Parameter Estimates. *Multivariate Behavioral Research*, *49*(6), 518–543. https://doi.org/10.1080/00273171.2014.933762

Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: A critique. *Comprehensive Psychology*, *4*, 1–19. https://doi.org/10.2466/03.cp.4.10

Raykov, T. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences*, *24*(2), 291–293. https://doi.org/10.1016/S0191-8869(97)00159-1

Reise, S. P., Ventura, J., Nuechterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *Journal of Personality and Assessment*, *84*(2), 126–136. https://doi.org/10.1207/s15327752jpa8402

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling*, *3*(4), 369–379. https://doi.org/10.1080/10705519609540052

Roesch, S. C., Aldridge, A. A., Stocking, S. N., Villodas, F., Leung, Q., Bartley, C. E., & Black, L. J. (2010). Multilevel factor analysis and structural equation modeling of daily diary coping data: Modeling trait and state variation. *Multivariate Behavioral Research*, *45*(5), 767–789. https://doi.org/10.1080/00273171.2010.519276

Ryu, E. (2014). Factorial invariance in multilevel confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, *67*(1), 172–194. https://doi.org/10.1111/bmsp.12014

Satorra, A., & Bentler, P. M. (2010). Ensuring possitiveness of the scaled chi-square test statistic. *Psychometrika*, *75*(2), 243–248.

Schaffer, B. S., & Manegold, J. G. (2018). Investigating antecedents of task commitment and task attraction in service learning team projects. *Journal of Education for Business*, *93*(5), 222–232. https://doi.org/10.1080/08832323.2018.1457620

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*(1), 95–113. https://doi.org/10.1177/0013164410387348

Schmitt, T. A., Sass, D. A., Chappelle, W., & Thompson, W. (2018). Selecting the “best” factor structure and moving measurement validation forward: An illustration. *Journal of*

Personality Assessment, *100*(4), 345–362. https://doi.org/10.1080/00223891.2018.1449116

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Shaw, J. D., Zhu, J., Duffy, M. K., Scott, K. L., Shih, H. A., & Susanto, E. (2011). A contingency model of conflict and team effectiveness. *Journal of Applied Psychology*, *96*(2), 391–400. https://doi.org/10.1037/a0021340

Smith-Crowe, K., & Burke, M. J. (2003). Interpreting the statistical significance of observed AD interrater agreement values: Correction to Burke and Dunlap (2002). *Organizational Research Methods*, *6*(1), 129–131. https://doi.org/10.1177/1094428102239428

Somech, A., & Drach-Zahavy, A. (2013). Translating team creativity to innovation implementation: The role of team composition and climate for innovation. *Journal of Management*, *39*(3), 684–708. https://doi.org/10.1177/0149206310394187

Sorra, J. S., & Dyer, N. (2010). Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Services Research*, *10*(1), 1–13. https://doi.org/10.1186/1472-6963-10-199

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. https://doi.org/10.3102/1076998616646200

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, *5*(4), 411–419. https://doi.org/10.1080/10705519809540115

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling*, *23*(6), 777–781. https://doi.org/10.1080/10705511.2016.1217487

Steiger, J. H., & Lind, J. C. (1980). Paper presented at the annual meeting of the Psychometric Society. *Statistically-Based Tests for the Number of Common Factors*.

Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, *32*(1), 29–55. https://doi.org/10.1177/0149206305277792

Tasa, K., Taggar, S., & Seijts, G. H. (2007). The development of collective efficacy in teams: A multilevel and longitudinal perspective. *Journal of Applied Psychology*, *92*(1), 17–27. https://doi.org/10.1037/0021-9010.92.1.17

Tay, L., Woo, S. E., & Vermunt, J. K. (2014). A Conceptual and methodological framework for psychometric isomorphism: Validation of multilevel xonstruct measures. *Organizational Research Methods*, *17*(1), 77–106. https://doi.org/10.1177/1094428113517008

Thomas, J. S., Loignon, A. C., Woehr, D. J., Loughry, M. L., & Ohland, M. W. (2019). Dyadic viability in project teams: the impact of liking, competence, and task interdependence. *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-019-09647-6

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

Van der Vegt, G. S., Emans, B. J. M., & Vliert, E. (2001). Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction. *Personnel Psychology*, *54*(1), 51–69. https://doi.org/10.1111/j.1744-6570.2001.tb00085.x

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, *12*(2), 368–392. https://doi.org/10.1177/1094428107309322

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–69. https://doi.org/10.1177/109442810031002

Wallace, J. C., Edwards, B. D., Paul, J., Burke, M., Christian, M., & Eissa, G. (2016). Change the referent? A meta-analytic investigation of direct and referent-shift consensus models for organizational climate. *Journal of Management*, *42*(4), 838–861. https://doi.org/10.1177/0149206313484520

Weiss, H. M. (2002). Deconstructing job satisfaction. Separating evaluations, beliefs and affective experiences. *Human Resource Management Review*, *12*(2), 173–194. https://doi.org/10.1016/S1053-4822(02)00045-1

Whitton, S. M., & Fletcher, R. B. (2014). The Group Environment Questionnaire: A multilevel confirmatory factor analysis. *Small Group Research*, *45*(1), 68–88. https://doi.org/10.1177/1046496413511121

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). https://doi.org/10.1037/10222-009

Wit, F. R. C. De, Greer, L. L., & Jehn, K. A. (2012). The paradox of intragroup conflict : A meta-analysis. *Journal of Applied Psychology*, *97*(2), 360–390. https://doi.org/10.1037/a0024844

Woehr, D. J., Loignon, A. C., Schmidt, P. B., Loughry, M. L., & Ohland, M. W. (2015). Justifying aggregation with consensus-based constructs: A review and examination of cutoff values for common aggregation indices. *Organizational Research Methods*, *18*(4), 704–737. https://doi.org/10.1177/1094428115582090

Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research*, *45*(3), 420–452. https://doi.org/10.1080/00273171.2010.483378

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165–200.

Zyphur, M. J., Kaplan, S. A., & Christian, M. S. (2008). Assumptions of cross-level measurement and structural invariance in the analysis of multilevel data: Problems and solutions. *Group Dynamics: Theory, Research, and Practice*, *12*(2), 127–140. https://doi.org/10.1037/1089-2699.12.2.127

TABLES

Table 1

Analogy Between Measurement Equivalence and Cross-level Isomorphism Adapted Tay & Colleagues (2014) Framework

Between-Groups Measurement Equivalence	Cross-level Isomorphism
None Groups cannot be compared.	Partial configural isomorphism: Some dimensions hold between levels of analysis. Typically occurs when a simplified factor structure is found at a higher level of analysis reflecting a similar meaning to its lower-level counterpart.
Configural invariance: Pattern of zero and nonzero factor loadings holds between groups	Weak configural isomorphism: Same number of dimensions holds between levels Dimensions are generally shown to be indexed by similar indicators across levels without fixing loading patterns Strong configural isomorphism: Same number of dimensions and the pattern of zero and nonzero factor loadings holds between levels
Metric invariance: Factor loadings are equivalent between groups	Weak metric isomorphism: Relative ordering of factor loadings/item discriminations holds between levels (evidenced by high congruence of the loadings between levels) Strong metric isomorphism: Magnitude of factor loadings/item discriminations holds between levels
Scalar invariance: Indicator thresholds are equivalent between groups	No current models for estimating item thresholds across levels
Invariance in uniqueness	No statistical basis for testing across levels

Note. This framework has been adapted to include partial configural isomorphism. Categories of cross-level isomorphism (i.e., measurement invariance across levels of analysis) to be used when examining isomorphism in a single measure.

Table 2 *(1of 2)*
Model Fit Indices in a Multilevel Context

	Fit Index	Description
Appropriate in a CFA/One-level Factor Model	χ^2	A relative/comparative fit index. χ^2. A test of a null hypothesis examining the likelihood the hypothesized variance/covariance matrix exactly fits the estimated population variance/covariance matrix. Sensitive to sample size, more prone to Type I error than other fit indices, and does not provide information as to the degree of misfit or magnitude. Difficult to compare the χ^2 values due to differences due to model complexity and differences sample size influencing the results. Researchers tend to focus on whether differences between the models are significantly different (e.g., $\alpha < .05$). (See Dunlap et al., 2003; Hu & Bentler, 1999; Satorra & Bentler, 2010)
	CFI	A relative/comparative fit index and a function of χ^2. Provides a normed assessment (ranging from 0 to 1) of the *goodness of fit* between a baseline model – where covariances "between any pairs of variables are set to zero" – and the hypothesized model (Hsu et al., 2015, p. 200). Not sensitive to sample size, and the recommended cutoff value ≥ 0.95. (See P. Bentler, 1990; Hu & Bentler, 1999; Rigdon, 1996; Wu & West, 2010)
	TLI	A relative/comparative fit index and a function of χ^2. Assesses the *goodness of fit* for the proposed model and is sensitive to small sample sizes (Wu & West, 2010). Rewards more parsimonious models by incorporating the degrees of freedom when comparing the χ^2 from the baseline model and the χ^2 from the proposed model. Typically falling between a 0 – 1 range, the recommended cutoff value is ≥ 0.95. (See Hu & Bentler, 1999; Marsh et al., 1988; Tucker & Lewis, 1973)
	RMSEA	An absolute fit index and a function of χ^2 that reflects the *badness of fit* or misfit of the model. Offers researchers the ability to calculate a confidence interval. Takes into account model complexity when examining the discrepancy (i.e., residuals) between the observed covariance/variance matrix and one implied by the proposed model. Sensitive to small sample sizes, and the recommended cutoff value ≤ 0.06. (See Hu & Bentler, 1999; Steiger, 1998, 2016; Steiger & Lind, 1980; Wu & West, 2010)
	SRMR	An absolute fit index. Assesses *badness of fit* or misfit of the model by examining the "average of standardized residuals between the observed and model-implied covariance matrices" (F. F. Chen, 2007, p. 467). Not sensitive to sample sizes, the recommended cutoff for is ≤ 0.08 in general and may be .11 or below when TLI ≥ 0.95. (See P. M. Bentler, 1995; Hu & Bentler, 1999)

Note. The table above reflects the appropriate model fit indices based on the level of analysis (Dyer et al., 2005; Hsu et al., 2015; L. K. Muthén & Muthén, 1998-2017). χ^2 = Chi-squared test, RMSEA = root mean square estimation, CFI = comparative fit index, SRMR = standardized root mean squared residual. W = within-team model. B = between-team model. TLI = Tucker-Lewis Index. The recommended cutoff values do not take into considerations the simplicity/complexity of the proposed model.

Table 2 *Continued (2 of 2)*

Model Fit Indices for Two-level Factor Models

Appropriate in a Two-level Factor Model		Fit Index	Description
	Within-team	χ^2	Adequately detects model misspecification at the within group level, still relies on standards of null hypothesis testing (e.g., $\alpha \leq .05$), and the reduction in sample size (e.g., # of participants - # of groups) was not overly problematic.
	Within-team	CFI	Adequately detects model misspecification at the within group level and the recommended cutoff value remains the same.
	Within-team	TLI	Adequately detects model misspecification at the within group level and performs better than CFI and RMSEA for simple misspecifications, recommended cutoff value remains the same, and the reduction in sample size (e.g., # of participants - # of groups) was not overly problematic.
		RMSEA	Adequately estimates the misfit of the within group portion of the proposed model.
		SRMR–W	Examines the covariance/variance matrices for the within group level of analysis. Adequately estimates the misfit of the within group portion of the proposed model and the recommended cutoff value remains the same (i.e., $\leq .08$).
	Between-team	χ^2	Adequately detects model misspecification at the between group level, still relies on standards of null hypothesis testing (e.g., $\alpha \leq .05$), and χ^2's ability (i.e., empirical power) to detect misfit increases as the ICC(1) value increases.
	Between-team	SRMR–B	Examines the covariance/variance matrices for the between group level of analysis. Adequately detects model misspecification at the between group level and is not sensitive to varying degrees of ICC(1), and the recommended cutoff value remains the same (i.e., $\leq .08$).

Note. The table above reflects the appropriate model fit indices based on the level of analysis (Dyer et al., 2005; Hsu et al., 2015; L. K. Muthén & Muthén, 1998-2017). χ^2 = Chi-squared test, RMSEA = root mean square estimation, CFI = comparative fit index, SRMR = standardized root mean squared residual. W = within-team model. B = between-team model. TLI = Tucker-Lewis Index. The recommended cutoff values do not take into considerations the simplicity/complexity of the proposed model.

Table 3 *Continued (1 of 2)*

Summary of Articles from Literature Review Examining Psychometric Properties of Team Consensus Measures

	# of pubs.	% Reporting inter-rater agreement (range, $\bar{x}$)	% Reporting scale reliability (range, $\bar{x}$)			% Reporting ICC(1) (range, $\bar{x}$)
		$r_{WG(j)}$	α	Agg. α	ω	
Team Conflict	29	41%	100%	*3%*	*0%*	52%
conflict	5	60% (.72 - .98, 87)	*17% (.81 - .97, .91)*	*0%*		*80% (.12 - .55, .32)*
task	*20*	*15% (.68 - .91, .80)*	*100% (.73 - .96, .85)*	*5% (.80)*		*55% (.08 - .41, .21)*
relationship	*20*	*15% (.56 - .96, 82)*	*100% (.72 - .96, .88)*	*0%*		*45% (.10 - .49, .25)*
process	*4*	*25% (.72 - .76, .74)*	*100% (.73 - .93, .75)*	*0%*		*25% (.19 - .25, .21)*
Team Cohesion	7	57%	71%	14%	0%	42%
cohesion	5	*60% (.80 - .90, .85)*	*80% (.80 - .89, .83)*	*20% (.88)*		*40% (.24 - .30, .27)*
interpersonal	5	*0%*	*20% (.81)*	*0%*	0%	*0%*
task attraction	2	*0%*	*50% (.74)*	*0%*		*0%*
task commitment	6	*16% (.90)*	*16% (.93)*	*0%*		*17% (.19)*
Team Psychological Safety	88	16% (.68 - .93, .83)	97% (.39 - .98, .78)	0%	5% (.72 - .88, .80)	38% (.02 - .62, .24)
Team Task Interdependence	11	37% (.37 - .90, .69)	100% (.73 - .92, .82)	0%	0%	50% (.01 - .16, .09)

Note. Pubs. = publications. $\bar{x}$ = mean. α = Chronbach's alpha. Agg. α = aggregated alpha. ω = multilevel composite. Inter-class correlation coefficient (ICC(1)) = the amount of variation between teams versus within teams in a sample. $r_{WG(j)}$ = inter-rater agreement index for each team. Cohesion = refers to when results were on a general latent factor of team cohesion by combining either all three subscales or the task and interpersonal subscales. Cohesion = any use of the items for the measure of team cohesion. Cohesion = results for a general latent factor of team cohesion by combining either all three subscales or the task and interpersonal subscales. If a range was provided, the average of the two numbers was used to calculate the mean. Team Conflict = reporting on any subscale for the measure of team cohesion. Many of the scales were manipulated by removing an item. While every team has a distinct $r_{WG(j)}$ value, only two publications reported a $r_{WG(j)}$ range while the rest reported a singular value which occasionally was specified to be the mean or median of all teams' $r_{WG(j)}$. Conflict refers to when results were reported on a general conflict factor using either all three subscales or a combination of the team relationship and task conflict subscales.

Table 3 *(2 of 2)*

Summary of Articles from Literature Review Examining Psychometric Properties of Team Consensus Measures

	% Reporting to run FA		% Found simplified factor structure	% Reporting information on standardized factor loadings	% Reporting model fit indices				
	CFA	MCFA	(in CFAs)	(range, $\bar{x}$)	χ^2	*CFI*	*TLI*	*RMSEA*	*SRMR*
Team Conflict [1]`	28%	0%	25%	10% (.60 - .97, .80)	10%	14%	14%	14%	14%
Team Cohesion	14%	0%	0%	14% (.57 – 79, .69)	14%	0%	0%	14%	14%
Team Psychological Safety	11%	2%	N/A	11% (.00 - .76, .61)	0%	0%	0%	0%	0%
Team Task Interdependence	18%	0%	N/A	18% (.55 - .84, .71)	0%	0%	0%	0%	0%

Note. CFA = Confirmatory factor analysis. MCFA = Multilevel confirmatory factor analysis. χ^2 = Chi-squared. CFI = comparative fit index. TLI = Tucker-Lewis Index. RMSEA = root mean square error of approximation. SRMR = standardized root mean squared residual. One article conducted a principle components analysis using all three team conflict subscales finding support for a general team conflict factor (Boies & Howell, 2006).

Table 4 *(1of 2)*

Summary of Hypotheses & Research Questions

Topic	Hypotheses (H) & Research Questions (RQ)	F.S.	P.S.	N.S.
Metric Isomorphism	*H1:* Common measures of team constructs (i.e., team cohesion, conflict, psychological safety, task interdependence, satisfaction) metric isomorphism from the individual to the between-team level of analysis.		✓	
	H1a: Team cohesion reveals metric isomorphism from the individual to the between-team level of analysis	✓		
	H1b: Team conflict reveals metric isomorphism from the individual to the between-team level of analysis.	✓		
	H1c: Team psychological safety reveals metric isomorphism from the individual to the between-team level of analysis.	✓		
	H1d: Team satisfaction reveals metric isomorphism from the individual to the between-team level of analysis.	✓		
	H1e: Team task interdependence reveals metric isomorphism from the individual to the between-team level of analysis.			✓
Factor Structure in Multidimensional	*H2:* Measures of team cohesion & conflict have the same factor structure at the individual and between-team level of analysis.			✓
	H2a: The measure of team cohesion reveals a three-factor model (i.e., interpersonal cohesiveness, task attraction, and task commitment) versus a two-factor (i.e., interpersonal and task-oriented cohesion) or one-factor (i.e., general team cohesion) at the between-team level of analysis.			✓
	H2b: The measure of team conflict reveals a three-factor model (i.e., relationship, process, and task conflict) versus a two-factor – combining the process and task or process and relationship dimensions − at the between-team level of analysis.			✓

Note. F.S. = fully supported hypothesis, P.S. = partially supported hypothesis, N.S. = hypothesis not supported.

Table 4 *Continued (2 of 2) Continued*
Summary of Hypotheses & Research Questions

Topic	Hypotheses (H) & Research Questions (RQ)	F.S.	P.S.	N.S.
Measure Characteristics	*H3:* The characteristics of measures of common team constructs relate to their degree of psychometric isomorphism, such that stricter measurement invariance from the individual to the between-team levels of analysis occurs based on the measure’s referent and/or target.		✓	
	H3a: The referent of a measure’s items influences the degree of psychometric isomorphism among measures of common team constructs, such that measures that refer to the self (direct consensus measurement models) reveal stricter measurement invariance from the individual to the between-team levels of analysis than measures that refer to the team (referent-shift consensus measurement models).			✓
	H3b: In direct consensus models, a measure’s target influences the level of psychometric isomorphism such that measures with a target of member-member relationships (e.g., team member liking and viability) have poorer model fit, reduced factor loadings, and greater residual variance at the between-team level versus the within-team and individual level of analysis compared to measures with a team target which will reveal superior indicators of psychometric isomorphism at the individual to the between-team level of analysis.		✓	
Team-member Agreement	*H4:* Team member agreement influences a measure’s level of psychometric isomorphism such that, measures reveal a stricter standard of measurement invariance across levels of analysis when there is greater agreement among members.		✓	
		S.E.	M.E.	N.E.
Research Questions	*RQ1:* Does the degree of measurement invariance across levels of analysis (i.e., psychometric isomorphism) constrain the degree of structural invariance among measures of team cohesion, conflict, interdependence, psychological safety, satisfaction, general member liking, and viability?			✓
	RQ2: Does the level of analysis relate to the degree of differentiation among measures of team cohesion, conflict, interdependence, psychological safety, satisfaction, general member liking, and viability?	✓		

Note. F.S. = fully supported hypothesis, P.S. = partially supported hypothesis, N.S. = hypothesis not supported, S. E. = strong evidence, M.E. = moderate evidence, N.E. = no evidence.

Table 5 *(1 of 2) Continued*
List of Measures

Construct (Citation)	*Subdimension*	Item
Team Cohesion (Carless & De Paola, 2000; Loughry & Tosi, 2008)		
	Task Attraction	Being part of the team allows team members to do enjoyable work.
		Team members get to participate in enjoyable activities.
		Team members like the work that the group does.
	Interpersonal Cohesiveness	Team members like each other.
		Team members get along well.
		Team members enjoy spending time together.
	Task Commitment	Our team is united in trying to reach its goals for performance.
		I'm unhappy with my team's level of commitment to the task. (*RC*)
		Our team members have conflicting aspirations for the team's performance. (*RC*)
Team Conflict (Jehn & Mannix, 2001)		
	Task Conflict	How much conflict of ideas is there in your work group?
		How frequently do you have disagreements within your work group about the task of the project you are working on?
		How often do people in your work group have conflicting opinions about the project you are working on?
	Relationship Conflict	How much relationship tension is there in your work group?
		How often do people get angry while working in your group?
		How much emotional conflict is there in your work group?
	Process Conflict	How often are there disagreements about who should do what in your work group?
		How much conflict is there in your group about task responsibilities?
		How often do you disagree about resource allocation in your work group?
Team Satisfaction (Gladstein, 1984; Van der Vegt et al, 2001)		
		I am satisfied with my present teammates.
		I am pleased with the way my teammates and I work together.
		I am very satisfied with working in this team.

Note. (RC) = Reverse coded items. All measures were administered on a Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Table 5 *(2 of 2) Continued*

List of Measures

Construct (Citation)

Subdimension Item

Team-member Viability (Thomas et al., 2019)

1. I would gladly work with this individual in the future.
2. If I were selecting members for a future work team, I would pick this person.
3. I would avoid working with this person in the future. (*RC*)

Team-member Liking (Thomas et al., 2019)

1. I like this person as an individual.
2. I consider this person to be a friend.
3. I enjoy spending time with this person.

Team Psychological Safety (Edmondson, 1999)

1. If you make a mistake on this team, it is often held against you. (*RC*)
2. Members of this team are able to bring up problems and tough issues.
3. People on this team sometimes reject others for being different. (*RC*).
4. It is safe to take a risk on this team.
5. It is difficult to ask other members of this team for help. (*RC*)
6. No one on this team would deliberately act in a way that undermines my efforts.
7. Working with members of this team, my unique skills and talents are valued and utilized.

Team Task Interdependence (Gladstein, 1984; Van der Vegt et al, 2001)

1. My teammates and I have to obtain information and advice from one another in order to complete our work.
2. I depend on my teammates for the completion of my work.
3. I have a one-person job; I rarely have to check or work with others.
4. I have to work closely with my teammates to do my work properly.
5. In order to complete our work, my teammates and I have to collaborate extensively.

Note. (*RC*) = Reverse coded items. All measures were administered on a Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Table 6

Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Measure of Team Cohesion

	X^2	*df*	CFI	TLI	RMSEA [CI]	SRMR	ω
Step 1: S_T							
1-Factor	12,591.35	27	.87	.83	.12 [.11,.12]	.06	.88
2-Factor	9,337.06	26	.91	.87	.10 [.10, .10]	.06	.88 (*i*) .80 (*t*)
3-Factor	5,972.72	24	.94	.93	.08 [.08 .09]	.04	.88 (*i*), 85 (*ta*), .66 (*tc*)
Step 3: S_W							
1-Factor	15,581.09	27	.88	.84	.13[.13,.13]	.07	.83
2-Factor	11,673.30	26	.91	.88	.11[.11, .11]	.07	.83(*i*), .73(*t*)
3-Factor	7,832.72	24	.94	.91	.10[.10, .10]	.05	.83 (*i*), 80 (*ta*), .56 (*tc*)
Step 4: S_B							
1-Factor	7,390.26	27	.89	.86	.18[.18, .18]	.06	.94
2-Factor	5,041.21	26	.93	.90	.15[.15, .16]	.06	.94(*i*), .89(*t*)
3-Factor	2,962.17	24	.96	.94	.12[.12, .13]	.04	.94 (*i*), .91(*ta*), .79 (*tc*)
Step 5: Multilevel							
3-Factor-w	8,077.13	51	-	-	-	.05 (-*w*)	.84 (*i-w*), .80 (*ta-w*), .56 (*tc-w*)
1-Factor-b						.05 (-*b*)	.94 (-*b*)
3-Factor-w	7,832.18	50	-	-	-	.05 (-*w*)	.84 (*i-w*), .80 (*ta-w*), .56 (*tc-w*)
2-Factor-b						.05 (-*b*)	.98 (*i-b*), .98 (*t-b*)
3-Factor-w	7,617.20	48	-	-	-	.05 (-*w*)	.84 (*i-w*), .80 (*ta-w*), .56 (*tc-w*)
3-Factor-b						.04 (-*b*)	.98 (*i-b*), .98 (*ta-b*), ..99 (*tc-b*)

Note. *n* = 34,400 individuals, 8,361 teams. All factor loadings were modeled with 3-factor structure at the within team level in multilevel analysis. S_T = CFA on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = MCFA on observed scores with team as a grouping factor. 1-Factor = all items loading on a general team cohesion factor. 2-Factor = an interpersonal (*i*) factor and task-oriented (*t*) factor containing the task attraction and commitment items. 3-Factor = an interpersonal (*i*), task-attraction (*ta*), and task-commitment (*tc*) cohesion factors. ω = composite reliability. All fit indices were statistically significant $p < .05$. All models in the MCFA are specified with three factors at the individual level. X^2 = Chi-squared. *df* = degrees of freedom. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. SRMR = Standardized root mean square residual. CI = Confidence interval. ω = composite reliability. -*w* = within-level. -*b* = between-level.

Table 7

Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Measure of Team Conflict

		X^2	*df*	CFI	TLI	RMSEA [CI]	SRMR	ω
Step 1: S_T	*2- F (task, & emotionally laden*)	5408.27	26	.95	.93	.07 [.07, .07]	.04	.83 (*t*), 88 (*e*)
	2-F (disagreements, & relationship)	6567.46	26	.94	.91	.08 [.07, .08]	.04	.87 (*d*), .85 (*r*)
	3-F (task, process, & relationship)	1842.00	24	.98	.97	.04 [.04, .04]	.02	.83 (*t*), .82 (*p*), .85 (*r*)
Step 3: S_W	*2-F (task, & emotionally laden*)	7997.78	26	.95	.92	.08 [.08, .09]	.04	.77 (*t*), 83 (*e*)
	2-F (disagreements, & relationship)	8194.66	26	.94	.92	.08 [.08, .09]	.04	.83(*d*), .77 (*r*)
	3-F (task, process, & relationship)	2429.50	24	.98	.98	.05 [.05, .05]	.02	.77 (*t*), .76 (*p*), .77 (*r*)
Step 4: S_B	*2-F (task, & emotionally laden*)	5178.78	26	.94	.92	.14 [.13, 14]	.05	.91 (*t*), .94 (*e*)
	2-F (disagreements, & relationship)	7641.72	26	.91	.88	.17 [.16, .17]	.05	.92 (*d*), .92 (*r*)
	3-F (task, process, & relationship)	2092.22	24	.98	.96	.09 [.09, .09]	.03	.91 (*t*), .89 (*p*), .92 (*r*)
Step 5: Multilevel	*3-F-w (task, process, & relationship)* *2-F-b (task, & emotionally laden*)	2522.882	50	-	-	-	.03(-*w*) .05 (-*b*)	.77 (*t-w*), .77 (*p-w*), .77 (*r-w*) .99 (*t-b*), .99 (*e-b)*
	3-F-w (task, process, & relationship) *2-F-b (disagreements, & relationship)*	3188.14	50	-	-	-	.03(-*w*) .08 (-*b*)	77 (*t-w*), .76 (*p-w*), .79 (*r-w*) .98 (*d-b*), .99 (*r-b)*
	3-F -w (task, process, & relationship) *3-F-b (task, process, & relationship)*	2341.22	48	-	-	-	02(-*w*) .04 (-*b*)	77 (*t-w*), .76 (*p-w*), .77 (*r-w*) .99 (*t-b*), .98 (*p-b*), .99 (*r-b*)

Note. *n* = 44,379 individuals, 10842 teams. All factor loadings were modeled with 3-factor structure at the within team level in multilevel analysis. S_T = CFA on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = MCFA on observed scores with team as a grouping factor. F = factor structure. E*motionally laden* (*e*) factor with items from the process and relationship subscales. *Disagreements* on team-related activities (*d)* comprised of items from the task and process conflict subscales. 3-Factor = a three-factor solution with *task*, *relationship*, and *process* team conflict as distinct factors. ω = composite reliability. All fit indices were statistically significant $p < .05$. X^2 = Chi-squared. *df* = degrees of freedom. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. SRMR = Standardized root mean square residual. CI = Confidence interval.-*w* = within-level. -*b* = between-level. ω = composite reliability.

Table 8

Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Uni-dimensional Measures of Team Constructs

	X^2	*df*	CFI	TLI	RMSEA [CI]	SRMR	ω
Psychological Safety							
Step 1: S_T	1,617.15	14	.85	.77	.09 [.09, .10]	.05	.76
Step 3: S_W	2383.07	14	.85	.77	.11 [.11, .12]	.06	.71
Step 4: S_B	737.07	14	.91	.86	.13 [.12, .13]	.06	.84
Step 5: Multilevel	2415.34	28	--	--	--	(.07-*wR*, .07-*wM*, .17-*bR*, .18-*bM*)	.73 (*w*), .96 (*b*)
Satisfaction & Task Interdependence – 2-factor model							
Step 1: S_T	1370.28	19	.98	.97	.06 [.06, .07]	.03	.95 (*s*), .73 (*i*)
Step 3: S_W	1216.00	19	.98	.97	.06 [.06, .06]	.03	.93 (*s*), .71 (*i*)
Step 4: S_B	497.69	19	.98	.97	.08 [.07, .09]	.05	.97 (*s*), .78 (*i*)
Step 5: Multilevel	1366.85	38	--	--	--	(.04-*wR*, .04-*wM*, .10-*bR*, .28-*bM*)	.93 (*s-w*), .72 (*i-w*) 1.00 (*s-b*), .89 (*i-w*)
Task Interdependence – 1 factor model							
Step 1: S_T	1281.41	5	.95	.89	.11 [.11, .12]	.04	.74
Step 3: S_W	7568.60	5	.94	.88	.11 [.11, .11]	.04	.72
Step 4: S_B	1531.22	5	.95	.90	.12 [.12, .13]	.04	.78
Step 5: Multilevel	1285.81	10	---	--	--	(.04-*wR*, .04-*wM*, .18-*bR*, 19-*bM*)	.72 (*w*), .90 (*b*)

Note. Team psychological safety MCFA n = 13,341 individuals, 4,054 teams. Team satisfaction (*s*) & task interdependence (*i*) MCFA *n* = 17,045 individuals, 4,054 teams. Team task interdependence MCFA *n* = 20,365 individuals, 4,805 teams. All factor loadings were modeled with 3-factor structure at the within team level in multilevel analysis. S_T = CFA on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = MCFA on observed scores with team as a grouping factor. All fit indices were statistically significant $p < .05$. X^2 = Chi-squared. df = degrees of freedom. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. SRMR = Standardized root mean square residual. CI = Confidence interval. -*w* = within-level. -*b* = between-level. -*R* = results specific to R. -*M* = results specific to MPlus. ω = composite reliability. This current study's measure of team satisfaction was under-identified as a one-factor model. To address this, a CFA and MCFA was conducted specifying a two-factor model with team satisfaction as one latent factor and task interdependence as the second latent factor.

Table 9

Model Fit Indices via Multilevel Confirmatory Factor Analysis and Composite Reliability for the Measures of Team-member Liking and Viability

	X^2	*df*	CFI	TLI	RMSEA [CI]	SRMR (-*wp*, -*bp*, -*bt*)	ω (-*wp*, -*bp*, -*bt*)
Step 1: S_T							
1-Factor	62,067.45	9	.69	.49	.31 [.30, .31]	.09	.89
2-Factor	4,269.91	8	.99	.96	.09 [.08, .09]	.03	.86 (*v*), 90 (*l*)
Step 3: S_{WP}							
1-Factor	53,568.04	9	.85	.76	.28 [.28, .29]	.10	.93
2-Factor	1,104.02	8	1.00	.00	.04 [.04, .05]	.01	.94 (v), 89 (*l*)
Step 4a: S_{BP}							
1-Factor	15,267.67	9	.79	.64	.30 [.29, .30]	.10	.83
2-Factor	1,197.63	8	.98	.97	.09 [.08, .09]	.04	.73 (*v*), .88 (*l*)
Step 4b: S_{BT}							
1-Factor	6,181.87	9	.80	.66	.39 [.38, .40]	.09	.92
2-Factor	846.87	8	.97	.95	.15 [.14, .16]	.04	.88 (*v*), .95 (*l*)
Step 5: Multilevel (three-level CFA)							
1-Factor	4,210.32	26	--	--	--	(.02-*wp*, .06-*bp*, .06-*bt*)	.96 (*wp*), .94 (*bp*), .99 (*bt*)
2-Factor	1,956.68	24	--	--	--	(.02-*wp*, .06-*bp*, .06-*bt*)	.89 (*l-wp*), .88 (*l-bp*), .98 (*l-bt*) .93 (*v-wp*), .74 (*v-bp*), .98 (*v-bt*)

Note. *n* = 74,291 observations, 19,105 individuals, 4,528 teams. CFA = confirmatory factor analysis. MCFA = multilevel CFA accounting for the within-person (*wp*), between-person (*bp*), and between-team (*bt*) levels. All factor loadings were modeled with 3-factor structure at the within team level in multilevel analysis. S_T = CFA on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = MCFA on observed scores with team as a grouping factor. All MCFAs had a two-factor solution at the within- and between-person levels with one and two factors estimated the between-team level. 2-Factor = general team member viability (*v*) and liking (*l*) factor. X^2 = Chi-squared. *df* = degrees of freedom. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. SRMR = Standardized root mean square residual. CI = Confidence interval. ω = composite reliability. All fit indices were statistically significant $p < .05$.

Table 10

Standardized Factor Loadings for the Measure of Team Cohesion Across Level of Analysis & ICC(1)

	Standardized Loadings															ICC(1)
	1-factor					2-factors					3-factors					
Items	S_T	S_W	S_B	Multilevel WT	Multilevel BT	S_T	S_W	S_B	Multilevel WT	Multilevel BT	S_T	S_W	S_B	Multilevel WT	Multilevel BT	
Attract_1	.80	.76	.87	.82	1.00	.83	.79	.90	.82	1.00	.86	.81	.92	.81	1.01	.36
Attract _2	.75	.70	.81	.77	.93	.77	.72	.83	.77	.93	.80	.75	.86	.75	.94	.36
Attract _3	.74	.69	.81	.72	.95	.76	.71	.84	.71	.96	.76	.71	.84	.71	.95	.35
Commit_4	.73	.67	.82	.82	.98	.75	.68	.83	.82	.99	.85	.82	.90	.82	1.00	.43
Commit _5*r*	.51	.41	.66	.55	.95	.53	.43	.67	.54	.97	.62	.54	.75	.54	1.00	.42
Commit _6*r*	.36	.26	.52	.40	.86	.37	.27	.53	.39	.89	.48	.39	.62	.39	.95	.43
Intrprsnl_7	.83	.77	.90	.82	.99	.88	.82	.94	.82	1.00	.88	.82	.94	.82	1.00	.36
Intrprsnl _8	.81	.67	.88	.77	.98	.84	.77	.91	.77	.99	.84	.78	.91	.78	.98	.31
Intrprsnl _9	.83	.41	.89	.80	.95	.84	.79	.89	.80	.94	.84	.79	.89	.79	.94	.28

Note. n = 153,174 individuals, 8,361 teams. 1 factor = all cohesion items loading onto a general team cohesion latent factor. 2 factors = one task-oriented latent factor with items capturing attraction (TaskAttract) and commitment (TaskCommit) to team-related tasks, and the second factor representing the interpersonal cohesiveness (Interpersonal) of the team. All factor loadings were modeled with 3-factor structure at the within team level in multilevel analysis. All factor loadings were modeled with 3-factor structure at the within team level in multilevel analysis. S_T = Confirmatory factor analysis (CFA)on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = Multilevel (CFA) on observed scores with team as a grouping factor. WT = within-team. BT = between-team. ICC(1) = Interclass correlation coefficient. r = reverse coded item.

Table 11

Residual Variance for the Measure of Team Cohesion Across Level of Analysis

Items	1-factor					2-factors					3-factors				
	S_T	S_W	S_B	Multilevel		S_T	S_W	S_B	MCFA		S_T	S_W	S_B	MCFA	
				WT	BT				WT	BT				WT	BT
Attract_1	.35	.43	.24	.32	.01	.30	.38	.20	.33	.00	.26	.34	.15	.33	-.03
Attract _2	.44	.51	.34	.41	.14	.41	.48	.30	.41	.14	.37	.44	.26	.42	.11
Attract _3	.46	.53	.35	.49	.11	.42	.49	.30	.49	.08	.42	.49	.30	.49	.09
Commit_4	.47	.56	.33	.33	.04	.45	.54	.31	.33	.02	.28	.33	.19	.32	.01
Commit _5*r*	.74	.83	.57	.70	.10	.72	.81	.55	.71	.07	.61	.71	.44	.73	.00
Commit _6*r*	.87	.93	.73	.84	.26	.86	.93	.72	.85	.21	.77	.85	.62	.87	.10
Intrprsnl_7	.31	.41	.19	.32	.02	.23	.33	.12	.33	-.01	.23	.33	.12	.33	-.01
Intrprsnl _8	.35	.46	.23	.40	.05	.30	.40	.18	.41	.03	.30	.40	.18	.40	.03
Intrprsnl _9	.31	.40	.21	.37	.10	.30	.38	.20	.36	.12	.30	.38	.21	.36	.11

Note. n = 153,174 individuals, 8,361 teams. 1 factor = all cohesion items loading onto a general team cohesion latent factor. 2 factors = one task-oriented latent factor with items capturing attraction (TaskAttract) and commitment (TaskCommit) to team-related tasks, and the second factor representing the interpersonal cohesiveness (Interpersonal) of the team. All factor loadings were modeled with 3-factor structure at the within-team level in multilevel analysis. S_T = Confirmatory factor analysis (CFA) on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = Multilevel CFA on observed scores with team as a grouping factor. WT = within-team. BT = between-team. r = reverse coded items.

Table 12

Standardized Factor Loadings for the Measure of Team Conflict Across Level of Analysis & ICC(1)

	Standardized Loadings															ICC(1)
	2-factor					2-factor					3-factor					
Items	S_T	S_W	S_B	Multilevel WT	Multilevel BT	S_T	S_W	S_B	Multilevel WT	Multilevel BT	S_T	S_W	S_B	Multilevel WT	Multilevel BT	
Task_1	.78	.72	.87	.72	.97	.70	.64	.78	.74	.91	.78	.71	.87	.72	.97	.24
Task _2	.80	.73	.88	.73	1.00	.77	.70	.86	.72	1.02	.79	.73	.88	.73	1.00	.22
Taskt_3	.79	.73	.88	.73	.99	.71	.65	.79	.76	.93	.79	.72	.88	.73	.99	.23
Process_4	.74	.68	.82	.74	.98	.75	.69	.84	.73	.99	.79	.72	.87	.73	.99	.20
Process_5	.77	.71	.85	.74	.96	.75	.69	.82	.75	.94	.80	.74	.87	.74	.97	.24
Process_6	.70	.65	.78	.69	.97	.72	.66	.81	.68	.98	.73	.68	.82	.68	.98	.16
Relationship_7	.78	.68	.88	.74	.98	.82	.74	.90	.74	.98	.82	.74	.90	.74	.98	.31
Relationship_8	.76	.65	.86	.69	.96	.79	.69	.88	.69	.97	.79	.69	.88	.69	.97	.31
Relationship_9	.76	.66	.87	.73	.98	.81	.73	.90	.73	.99	.81	.73	.90	.73	.99	.28

Note. n = 44,379 individuals, 10842 teams. 2-factor = a two-factor solution. The first 2-factor solution contains an emotionally laden factor with items from the process and relationship subscales and the second factor comprised of items from the task conflict subscale of team conflict. The second 2-factor solution has a factor representing disagreements on team-related activities comprised of items from the task and process conflict subscales and the other factor comprised of items from the task relationship subscale of team conflict. 3-Factor = a three-factor solution with task, relationship, and process team conflict as distinct factors. All factor loadings were modeled with 3-factor structure at the within-team level in multilevel analysis. S_T = Confirmatory factor analysis (CFA) on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix Multilevel = MCFA on observed scores with team as a grouping factor. WT = within-team. BT = between-team. ICC(1) = Interclass correlation coefficient.

Table 13

Residual Variance for the Measure of Team Conflict Across Level of Analysis

Items	2-factors					2-factors					3-factors				
	S_T	S_W	S_B	Multilevel W/in	Multilevel Btwn	S_T	S_W	S_B	Multilevel W/in	Multilevel Btwn	S_T	S_W	S_B	Multilevel W/in	Multilevel Btwn
Task_1	.39	.49	.24	.48	.06	.51	.59	.39	.45	.17	.39	.49	.25	.49	.06
Task _2	.37	.47	.23	.47	.01	.41	.52	.26	.48	-.04	.36	.47	.22	.47	.01
Taskt_3	.37	.47	.22	.47	.03	.50	.59	.38	.43	.13	.37	.48	.23	.47	.03
Process_4	.46	.54	.32	.46	.05	.44	.52	.30	.47	.03	.38	.48	.24	.47	.02
Process_5	.41	.49	.28	.45	.08	.44	.52	.33	.44	.12	.36	.45	.24	.46	.06
Process_6	.52	.58	.40	.53	.07	.49	.57	.35	.54	.03	.47	.54	.33	.54	.05
Relationship_7	.39	.54	.23	.45	.04	.33	.45	.23	.45	.04	.32	.45	.18	.45	.03
Relationship_8	.42	.58	.26	.52	.07	.38	.52	.19	.52	.06	.38	.52	.23	.52	.07
Relationship_9	.42	.57	.25	.46	.04	.34	.46	.24	.47	.02	.34	.47	.19	.47	.02

Note. n = 44,379 individuals, 10842 teams. 2-factor = a two-factor solution. The first 2-factor solution contains an emotionally laden factor with items from the process and relationship subscales and the second factor comprised of items from the task conflict subscale of team conflict. The second 2-factor solution has a factor representing disagreements on team-related activities comprised of items from the task and process conflict subscales and the other factor comprised of items from the task relationship subscale of team conflict. 3-Factor = a three-factor solution with task, relationship, and process team conflict as distinct factors. S_T = Confirmatory factor analysis (CFA) on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = Multilevel CFA on observed scores with team as a grouping factor. WT = within-team. BT = between-team.

Table 14

Standardized Factor Loadings and Residual Variance for the Measure of Team Psychological Safety Across Level of Analysis & ICC(1)

Items	Factor loadings					Residual variances					ICC (1)
	S_T	S_W	S_B	MCFA		S_T	S_W	S_B	MCFA		
				WT	BT				W/in	Btwn	
PsychSafety_1r	.54	.47	.64	.44 (.43)	1.00 (1.01)	.71	.78	.59	.81	.00 (-.20)	.11
PsychSafety_2	.52	.47	.61	.54	.58 (.52)	.73	.78	.63	.71	.68 (.73)	.11
PsychSafety_3r	.56	.51	.63	.47	.98 (1.02)	.69	.74	.61	.78	.04 (-.04)	.10
PsychSafety _4	.64	.60	.71	.64	.75 (.72)	.59	.65	.49	.59	.44 (.48)	.13
PsychSafety _5r	.51	.44	.64	.43	.98 (.99)	.74	.81	.60	.81	.05 (.02)	.10
PsychSafety _6	.53	.49	.63	.52	.84 (.81)	.72	.76	.61	.73	.31(.34)	.08
PsychSafety _7	.68	.66	.73	.68 (.69)	.78 (.75)	.53	.57	.46	.54 (.53)	..39 (.44)	.13

Note. n = 13,373 for individuals, and 3,275 for teams. PsychSafety = Team Psychological safety items. Values in parenthesis reflect output in *R* when MPlus differs. S_T = Confirmatory factor analysis (CFA) on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = Multilevel CFA on observed scores with team as a grouping factor. WT = within-team. BT = between-team. ICC(1) = Interclass correlation coefficient. *r* = reverse-coded items.

Table 15

Standardized Factor Loadings and Residual Variance for the Measure of Team Task Interdependence Across Level of Analysis & ICC(1)

Items	Factor loadings					Residual variances					ICC (1)
	S_T	S_W	S_B	MCFA		S_T	S_W	S_B	MCFA		
				WT	BT				WT	BT	
Interdependence_1	.61	.60	.65	.60	.80	.63	.64	.58	.64	.35	.06
Interdependence_2	.53	.52	.57	.54	.36	.72	.73	.67	.71	.87	.08
Interdependence_3r	.27	.21	.32	.19	.93	.94	.96	.90	.97	.13	.05
Interdependence_4	.79	77	.83	.78.	.96	.38	.41	.31	.39	.09	.06
Interdependence_5	.76	.76	.79	.74	.98	.42	.44	.38	.45	.04	.08

Note. n = 20,365 for individual, and 4,805 for team. Interdependence = Team task interdependence items. S_T = Confirmatory factor analysis (CFA) on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = Multilevel CFA on observed scores with team as a grouping factor. WT = within-team. BT = between-team. ICC(1) = Interclass correlation coefficient. *r* = reverse-coded items.

Table 16

Standardized Factor Loadings and Residual Variance for Measure of Team Satisfaction and Task Interdependence Across Level of Analysis & ICC(1)

Items	Factor loadings					Residual variances					ICC (1)
	S_T	S_W	S_B	MCFA W/in	MCFA Btwn	S_T	S_W	S_B	MCFA W/in	MCFA Btwn	
Satisfaction_1	.92	.90	.96	.90	1.00	.15	.20	.08	.20	.01	.40
Satisfaction_2	.92	.90	.96	.90	.99	.15	.19	.09	.19	.02	.40
Satisfaction_3	.95	.93	.97	.93	1.00	.10	.14	.06	.14	.01	.40
Interdependence_1	.62	.60	.65	.59	1.00	.62	.64	.57	.66	.01	.23
Interdependence_2	.53	.51	.56	.52	.52	.72	.74	.68	.73	.73	.25
Interdependence_3r	.23	.20	.31	.20	.77	.95	.96	.90	.96	.41	.23
Interdependence_4	.78	.76	.82	.78	.96	.40	.43	.33	.40	.09	.24
Interdependence_5	.77	.76	.80	.76	.83	.41	.43	.37	.42	.30	.26

Note. n = 17,045 for individual, and 4,054 for team. S_T = Confirmatory factor analysis (CFA) on observed scores. S_W = CFA on the pooled within covariance matrix. S_B = CFA on the between covariance matrix. Multilevel = Multilevel CFA on observed scores with team as a grouping factor. WT = within-team. BT = between-team. ICC(1) = Interclass correlation coefficient. *r* = reverse-coded items.

Table 17

Standardized Factor Loadings for the Measures of Team-member Viability and Liking Across Level of Analysis & ICC(1)

	1-factor							2-factors							ICC (1)	
Items	S_T	S_{WP}	S_{BP}	S_{BT}	MCFA			S_T	S_{WP}	S_{BP}	S_{BT}	MCFA			BP	BT
					WP	BP	BT					WP	BP	BT		
Viability_1	.94	.94	.91	.97	.96	.91	1.00	.96	.95	.93	.98	.95	.91	1.01	.19	.16
Viability _2	.93	.93	.91	.96	.94	.78	.99	.94	.93	.93	.96	.93	.92	.99	.22	.15
Viability_3r	.59	.83	.34	.62	.85	.02	.91	.60	.83	.36	.63	.84	.11	.93	.49	.11
Liking_4	.72	.69	.69	.83	.82	.77	.97	.82	.82	.78	.88	.82	.74	.98	.40	.16
Liking_5	.67	.66	.65	.75	.84	.88	.93	.86	.84	.85	.92	.84	.87	.96	.46	.20
Liking_6	.73	.71	.70	.81	.89	.91	.98	.92	.89	.90	.96	.89	.91	1.00	.42	.19

Note. n = 75,026 responses, 19,164 individuals, 4,587 teams. MCFA = multilevel confirmatory factor analysis accounting for the within-person (WP), between-person (BP), and between-team (BT) levels simultaneously. Results from variance/covariance matrices via Confirmatory factor analysis: S_{WP} = within-person, S_{BP} = between-person, S_{BT} = between-team. All MCFAs had a two-factor solution at the within- and between-person levels with one or two factors estimated the between-team level. 2-Factor = general team member viability (v) and liking (l) factor. ω = composite reliability. ICC = interclass correlation coefficient. r = reverse-coded items.

Table 18

Residual Variances for the Measures of Team-member Viability and Liking Across Level of Analysis

Items	1-factor							2-factor						
	S_T	S_{WP}	S_{BP}	S_{BT}	MCFA			S_T	S_{WP}	S_{BP}	S_{BT}	MCFA		
					WP	BP	BT					WP	BP	BT
Viability_1	.13	.12	.17	.07	.08	.34	-.01	.09	.10	.13	.04	.09	.17	-.01
Viability _2	.14	.14	.17	.08	.12	.39	.02	.13	.13	.14	.08	.13	.16	.03
Viability_3r	.66	.31	.88	.62	.29	1.00	.17	.64	.31	.87	.61	.30	.99	.14
Liking_4	.49	.53	.53	.32	.32	.42	.05	.33	.33	.39	.22	.33	.45	.05
Liking_5	.55	.57	.58	.44	.29	.22	.14	.26	.30	.29	.16	.30	.25	.09
Liking_6	.47	.50	.51	.34	.20	.17	.03	.16	.21	.19	.07	.21	.18	.00

Note. n = 75,026 responses, 19,164 individuals, 4,587 teams. MCFA = multilevel confirmatory factor analysis accounting for the within-person (WP), between-person (BP), and between-team (BT) levels simultaneously. Results from variance/covariance matrices via Confirmatory factor analysis: S_{WP} = within-person, S_{BP} = between-person, S_{BT} = between-team. All MCFAs had a two-factor solution at the within- and between-person levels with one or two factors estimated the between-team level. 2-Factor = general team member viability (v) and liking (l) factor. ω = composite reliability. ICC = interclass correlation coefficient. r = reverse-coded items.

Table 19

Correlations and Standard Errors (SE) Among Latent Factors at The Individual Level of Analysis

Variables	1	2	3	4	5	6	7
1. Emotion. Conflict	--						
2. Task Conflict	**0.80** (.02)	--					
3. Cohesion	-0.58 (.02)	-0.38 (.01)	--				
4. Satisfaction	-0.57 (.02)	-0.36 (.02)	**0.84** (.02)	--			
5. Task Interdepend.	-0.16 (.01)	-0.05 (.01)	0.41 (.01)	0.37 (.02)	--		
6. Psych. Safety	-0.57 (.02)	-0.38 (.02)	0.73 (.02)	0.63 (.03)	0.30 (.01)	--	
7. Team-member Viability	-0.52 (.02)	-0.35 (.01)	**0.79** (.02)	**0.85** (.02)	0.35 (.01)	0.59 (.02)	--
8. Team-member Liking	-0.34 (.01)	-0.24 (.01)	**0.77** (.01)	0.66 (.02)	0.31 (.01)	0.52 (.01)	**0.77** (.02)

Note. n = 2,332 individuals, 541 teams. Correlations and standard errors in bold reflect a lack of discriminant validity between the respective variables (Bagozzi et al., 1992; Schmitt et al., 2018). Standard error (SE) in parentheses. Variables above better reflect the individual level thus correlations reflect people's general perceptions in a team context regardless of team membership. Emotion. conflict = emotionally laden conflict. Psych. Safety = psychological safety. Task interdepend. = Task interdependence.

Table 20

Correlations Among Latent Factors at the Between-team Level of Analysis via S_B and Multilevel Factor Model

Variables	1	2	3	4	5	6	7
1. Team Emotion. Conflict	--						
2. Team Task Conflict	**.82 (.84)**	--					
3. Team Cohesion	-.72 **(-.88)**	-.47 (-.66)	--				
4. Team Satisfaction	-.73 **(-.84)**	-.46 (-.59)	**.91 (1.03)**	--			
5. Team Task Interdepend.	-.22 (-.11)	-.09 (-.18)	.43 (.38)	.40 (.29)	--		
6. Team Psych. Safety	-.75 **(-.96)**	-.51 (-.74)	**.85 (1.05)**	**.82 (1.04)**	. 35 (.43)	--	
7. Team Viability	-.69 **(-.81)**	-.47 (-.54)	**.87 (.96)**	**.89 (.92)**	.35 (.30)	**.78 (1.05)**	--
8. Team Liking	-.46 (-.72)	-.31 (-.43)	**.82 (.83)**	.70 **(.84)**	.36 (.54)	.64 **(.1.02)**	**.79 (.92)**

Note. n = 2,332 individuals, 541 teams. S_B = between covariance matrix. Multilevel factor model results at the between-team level of analysis in parentheses. Correlations in bold reflect a lack of discriminant validity between the respective variables (Bagozzi et al., 1992; Schmitt et al., 2018). Team emotion. conflict = emotionally laden team conflict. Team psych. Safety = team psychological safety. Team task interdepend. = team task interdependence.

Table 21

Correlations Among Latent Factors at the Within-team Level of Analysis via SW and Multilevel Factor Model

Variables	1	2	3	4	5	6	7
1. Dev. Emotion. Conflict	--						
2. Dev. Task Conflict	**.80 (.79)**	--					
3. Dev. Cohesion	-.45 (-.50)	-.31 (-.33)	--				
4. Dev. Satisfaction	-.42 (-.48)	-.28 (-.30)	**.78 (.81)**	--			
5. Dev. Task Interdepend.	-.12 (-.14)	-.03 (-.03)	.40 (.40)	.35 (.36)	--		
6. Dev. Psych. Safety	-.43 (-.41)	-.27 (-.25)	.63 (.68)	.48 (.55)	.27 (.29)	--	
7. Dev. Team Viability	-.37 (-.43)	-.27 (-.31)	.73 (.77)	**.81 (.83)**	.35 (.35)	.44 (.50)	--
8. Dev. Team Liking	-.23 (-.32)	-.18 (-.23)	.72 (.78)	.63 (.68)	.29 (.30)	.42 (.52)	.74 **(.79)**

Note. $n = 9666$ individuals, 541 teams. Correlations represent relationships among deviations in the current study's measures using team-member deviations from the team mean. Correlations in bold reflect a lack of discriminant validity between the respective variables (Bagozzi et al., 1992; Schmitt et al., 2018). Dev. = deviation within a team. Multilevel factor model results at the within-team level of analysis in parentheses Emotion. conflict = emotionally laden conflict. Team psych. Safety = psychological safety. Team task interdepend. = team task interdependence.

Table 22

Correlations and Standard Errors (SE) Among Latent Factors at The Between-Team Level of Analysis Via S_B

Variables	1	2	3	4	5	6	7
1. Team Emotion. Conflict	--						
2. Team Task Conflict	**.82** (.05)	--					
3. Team Cohesion	-.72 (.05)	-.47 (.04)	--				
4. Team Satisfaction	-.73 (.06)	-.47 (.05)	**.91 (.07)**	--			
5. Team Task Interdepend.	-.22 (.03)	-.09 (.02)	.43 (.03)	.40 (.04)	--		
6. Team Psych. Safety	-**.75** (.06)	-.51 (.05)	**.85** (.07)	**.81** (.08)	.35 (.03)	--	
7. Team Viability	-.60 **(.05)**	-.46 (.04)	**.87** (.05)	**.89 (.07)**	.35 (.03)	**.78** (.06)	--
8. Team Liking	-.46 (.03)	-.32 (.03)	**.83** (.04)	.71 (.05)	.36 (.03)	.64 (.05)	**.79** (.04)

Note. $n = 9666$ individuals, 541 teams. Correlations and standard errors in bold reflect a lack of discriminant validity between the respective variables (Bagozzi et al., 1992; Schmitt et al., 2018). Standard error (SE) in parentheses. Team emotion. conflict = emotionally laden team conflict. Team psych. Safety = team psychological safety. Team task interdepend. = team task interdependence.

Table 23 (1 of 6)

Inter-item Correlations Among Observed Variables at The Individual Level of Analysis

Team variables	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. Coh1_TA	4.02	0.83									
2. Coh2_TA	3.84	0.9	.69*								
3. Coh3_TA	4.08	0.78	.66*	.58*							
4. Coh4_IC	4.21	0.71	.63*	.57*	.61*						
5. Coh5_IC	4.31	0.71	.60*	.53*	.57*	.75*					
6. Coh6_IC	4.01	0.79	.66*	.66*	.58*	.73*	.66*				
7. Coh7_TC	4.26	0.82	.62*	.52*	.60*	.61*	.63*	.57*			
8. Coh8_TCr	4.08	1.07	.45*	.37*	.45*	.45*	.47*	.43*	.56*		
9. Coh9_TCr	3.88	1.08	.28*	.22*	.30*	.31*	.35*	.27*	.41*	.53*	
10. Con1_TC	1.7	0.79	-.20*	-.18*	-.21*	-.21*	-.23*	-.21*	-.20*	-.22*	-.23*
11. Con2_TC	1.58	0.73	-.26*	-.23*	-.27*	-.26*	-.29*	-.24*	-.30*	-.29*	-.29*
12. Con3_TC	1.75	0.78	-.20*	-.19*	-.23*	-.21*	-.24*	-.21*	-.21*	-.20*	-.22*
13. Con4_RC	1.4	0.72	-.35*	-.29*	-.35*	-.41*	-.45*	-.36*	-.39*	-.42*	-.34*
14. Con5_RC	1.36	0.67	-.33*	-.29*	-.36*	-.36*	-.40*	-.33*	-.37*	-.38*	-.29*
15. Con6_RC	1.3	0.64	-.30*	-.24*	-.29*	-.33*	-.38*	-.28*	-.33*	-.33*	-.30*
16. Con7_PC	1.47	0.7	-.29*	-.24*	-.31*	-.31*	-.34*	-.28*	-.35*	-.34*	-.28*
17. Con8_PC	1.52	0.78	-.36*	-.30*	-.34*	-.37*	-.41*	-.35*	-.43*	-.44*	-.34*
18. Con9_PC	1.46	0.7	-.27*	-.23*	-.29*	-.29*	-.34*	-.28*	-.34*	-.34*	-.28*

Note. $n = 9666$ individuals, 541 teams. Coh = Cohesion, TA = task attraction, IC = Interpersonal cohesiveness, TC = Task cohesiveness. Con = Conflict, TC = Task conflict, RC = Relationship conflict, PC = Process conflict. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. *M* and *SD* are used to represent mean and standard deviation, respectively. ** indicates $p < .05$. * indicates $p < .01$.

Table 23 *(2 Of 6) Continued*

Inter-item Correlations Among Observed Variables at The Individual Level of Analysis

Team variables	*M*	*SD*	1	2	3	4	5	6	7	8	9
19. Satisfaction1	4.29	0.88	**.62***	**.53***	**.57***	**.63***	**.64***	**.62***	**.68***	**.60***	.40*
20. Satisfaction2	4.25	0.89	**.61***	**.53***	**.58***	**.61***	**.63***	**.60***	**.68***	**.58***	.37*
21. Satisfaction3	4.22	0.93	**.63***	**.56***	**.59***	**.62***	**.63***	**.62***	**.67***	**.60***	.38*
22. Interdepend.1	3.86	0.92	.27*	.25*	.22*	.26*	.26*	.25*	.26*	.21*	.12*
23. Interdepend.2	2.83	1.22	.12*	.09*	.07*	.12*	.12*	.11*	.13*	.11*	0.02
24. Interdepend.3r	3.78	0.91	.11*	.08*	.13*	.13*	.15*	.13*	.15*	.16*	.17*
25. Interdepend.4	3.4	1.05	.23*	.20*	.17*	.21*	.21*	.23*	.24*	.18*	.10*
26. Interdepend.5	3.75	0.99	.29*	.23*	.22*	.25*	.25*	.27*	.31*	.21*	.11*
27. Psych. Safety1r	6.19	1.26	.22*	.20*	.26*	.27*	.30*	.24*	.23*	.27*	.22*
28. Psych. Safety2	5.35	1.52	.26*	.23*	.26*	.29*	.30*	.26*	.33*	.27*	.23*
29. Psych. Safety3r	6.54	0.99	.22*	.18*	.24*	.29*	.32*	.24*	.26*	.26*	.22*
30. Psych. Safety4	5.39	1.42	.33*	.31*	.33*	.38*	.39*	.35*	.36*	.33*	.27*
31. Psych. Safety5r	6.12	1.37	.38*	.34*	.37*	.39*	.41*	.37*	.45*	.42*	.27*
32. Psych. Safety6	5.88	1.61	.25*	.19*	.25*	.30*	.29*	.23*	.29*	.28*	.26*
33. Psych. Safety7	5.97	1.17	.41*	.38*	.43*	.45*	.45*	.42*	.43*	.36*	.30*
34. Viability1	4.29	0.72	**.59***	**.53***	**.54***	**.61***	**.60***	**.61***	**.61***	**.54***	.34*
35. Viability2	4.14	0.79	**.59***	**.53***	**.54***	**.59***	**.59***	**.60***	**.60***	**.52***	.32*
36. Viability3r	4.14	1.03	.28*	.22*	.23*	.32*	.33*	.28*	.33*	.39*	.33*
37. Liking4	4.36	0.65	**.52***	.49*	.49*	**.63***	**.57***	**.59***	.48*	.37*	.25*
38. Liking5	3.95	0.81	**.54***	**.55***	.48*	**.59***	.49*	.64*	.44*	.30*	.17*
39. Liking6	4.07	0.74	**.57***	**.55***	**.50***	**.62***	**.53***	**.69***	.47*	.36*	.21*

Note. n = 9666 individuals, 541 teams. Interdepend. = Task interdependence. Psych. Safety = Psychological safety. . Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. *M* and *SD* are used to represent mean and standard deviation, respectively. ** indicates p < .05. * indicates p < .01.

Table 23 *(3 of 6) Continued*

Inter-item Correlations Among Observed Variables at The Individual Level of Analysis

Variable	10	11	12	13	14	15	16	17	18
10. Con1_TC									
11. Con2_TC	.60*								
12. Con3_TC	.65*	.61*							
13. Con4_RC	.48*	.51*	.47*						
14. Con5_RC	.42*	.52*	.45*	.64*					
15. Con6_RC	.45*	.50*	.46*	.68*	.66*				
16. Con7_PC	.48*	.55*	.47*	.53*	.52*	.52*			
17. Con8_PC	.43*	.53*	.43*	.61*	.57*	.54*	.66*		
18. Con9_PC	.47*	.56*	.48*	.53*	.50*	.52*	.59*	.58*	
19. Satisfaction1	-.25*	-.31*	-.24*	-.48*	-.40*	-.39*	-.38*	-.46*	-.38*
20. Satisfaction2	-.24*	-.31*	-.26*	-.46*	-.40*	-.37*	-.37*	-.46*	-.38*
21. Satisfaction3	-.24*	-.31*	-.23*	-.45*	-.38*	-.35*	-.35*	-.45*	-.36*
22. Interdepend.1	-.07*	-.10*	-.07*	-.11*	-.11*	-.11*	-.10*	-.14*	-.11*
23. Interdepend.2	-0.02	-0.03	-0.04	-.06*	-.05*	-0.03	-0.04	-.09*	-.05*
24. Interdepend.3r	.04**	0.01	0.03	-.07*	-.08*	-.07*	-.08*	-.09*	-.07*
25. Interdepend.4	-0.01	-0.04	-0.01	-.06*	-.08*	-.08*	-.04**	-.10*	-.08*
26. Interdepend.5	0.01	-0.04	-0.01	-.08*	-.10*	-.06*	-.08*	-.14*	-.08*

Note. n = 9666 individuals, 541 teams. Con = Conflict, TC = Task conflict, RC = Relationship conflict, PC = Process conflict. Interdepend. = Task interdependence. . Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 23 *(4 of 6) Continued*
Inter-item Correlations Among Observed Variables at The Individual Level of Analysis

Variable	10	11	12	13	14	15	16	17	18
27. Psych. Safety1r	-.24*	-.28*	-.24*	-.30*	-.32*	-.30*	-.31*	-.30*	-.28*
28. Psych. Safety2	-.05*	-.11*	-0.03	-.17*	-.14*	-.14*	-.14*	-.17*	-.11*
29. Psych. Safety3r	-.23*	-.26*	-.19*	-.33*	-.30*	-.32*	-.28*	-.28*	-.26*
30. Psych. Safety4	-.13*	-.18*	-.13*	-.24*	-.25*	-.23*	-.18*	-.24*	-.19*
31. Psych. Safety5r	-.18*	-.25*	-.17*	-.34*	-.32*	-.28*	-.31*	-.35*	-.30*
32. Psych. Safety6	-.12*	-.15*	-.09*	-.25*	-.22*	-.23*	-.18*	-.20*	-.18*
33. Psych. Safety7	-.16*	-.21*	-.17*	-.31*	-.30*	-.27*	-.23*	-.24*	-.22*
34. Viability1	-.25*	-.31*	-.24*	-.43*	-.37*	-.32*	-.36*	-.44*	-.35*
35. Viability2	-.24*	-.29*	-.24*	-.40*	-.35*	-.30*	-.33*	-.42*	-.31*
36. Viability3r	-.11*	-.16*	-.11*	-.27*	-.22*	-.21*	-.20*	-.26*	-.22*
37. Liking4	-.17*	-.20*	-.17*	-.30*	-.26*	-.23*	-.22*	-.29*	-.24*
38. Liking5	-.15*	-.16*	-.17*	-.25*	-.20*	-.16*	-.18*	-.25*	-.18*
39. Liking6	-.16*	-.18*	-.17*	-.28*	-.23*	-.19*	-.21*	-.27*	-.21*

Note. n = 9666 individuals, 541 teams. Psych. Safety = Psychological safety. . Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates p < .05. * indicates p < .01.

Table 23 *(5 of 6) Continued*

Inter-item Correlations Among Observed Variables at the Individual Level of Analysis

Variable	19	20	21	22	23	24	25	26
19. Satisfaction1								
20. Satisfaction2	.87*							
21. Satisfaction3	.89*	.89*						
22. Interdepend.1	.28*	.28*	.27*					
23. Interdepend.2	.17*	.16*	.16*	.38*				
24. Interdepend.3r	.12*	.10*	.11*	.14*	-0.01			
25. Interdepend.4	.22*	.21*	.23*	.42*	.41*	.20*		
26. Interdepend.5	.27*	.27*	.28*	.48*	.31*	.23*	.62*	
27. Psych. Safety1r	.27*	.25*	.27*	.07*	0	.12*	.05*	.06*
28. Psych. Safety2	.29*	.29*	.30*	.15*	.08*	.14*	.11*	.13*
29. Psych. Safety3r	.28*	.26*	.25*	.13*	0.04	.15*	0.04	.08*
30. Psych. Safety4	.35*	.35*	.36*	.14*	.09*	.14*	.13*	.16*
31. Psych. Safety5r	.47*	.47*	.46*	.21*	.07*	.16*	.16*	.19*
32. Psych. Safety6	.28*	.27*	.28*	.12*	.05*	.14*	.07*	.10*
33. Psych. Safety7	.41*	.42*	.42*	.20*	0.02	.15*	.13*	.18*
34. Viability1	**.77***	**.73***	**.77***	.27*	.16*	.10*	.22*	.26*
35. Viability2	**.74***	**.71***	**.75***	.26*	.16*	.08*	.22*	.25*
36. Viability3r	.40*	.38*	.39*	.12*	0.04	.18*	.12*	.15*
37. Liking4	**.56***	**.56***	**.57***	.22*	.08*	.12*	.18*	.22*
38. Liking5	**.52***	**.52***	**.54***	.19*	.08*	.07*	.18*	.22*
39. Liking6	**.57***	**.57***	**.59***	.23*	.10*	.08*	.20*	.24*

Note. $n = 9666$ individuals, 541 teams. Interdepend. = Task interdependence. Psych. Safety = Psychological safety. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 23 *(6 of 6) Continued*

Inter-item Correlations Among Observed Variables at the Individual Level of Analysis

Variable	27	28	29	30	31	32	33	34	35	36	37	38
27. Psych. Safety1r												
28. Psych. Safety2	.14*											
29. Psych. Safety3r	.47*	.18*										
30. Psych. Safety4	.31*	.42*	.29*									
31. Psych. Safety5r	.37*	.23*	.42*	.29*								
32. Psych. Safety6	.23*	.36*	.27*	.34*	.25*							
33. Psych. Safety7	.32*	.37*	.33*	.46*	.37*	.44*						
34. Viability1	.26*	.27*	.29*	.33*	.43*	.25*	.39*					
35. Viability2	.24*	.26*	.25*	.32*	.41*	.23*	.39*	.89*				
36. Viability3r	.24*	.20*	.22*	.25*	.30*	.24*	.24*	.42*	.37*			
37. Liking4	.22*	.26*	.25*	.34*	.33*	.26*	.40*	.67*	.65*	.28*		
38. Liking5	.16*	.21*	.18*	.27*	.28*	.19*	.35*	.62*	.61*	.17*	.71*	
39. Liking6	.18*	.23*	.19*	.31*	.32*	.21*	.37*	.68*	.66*	.22*	.75*	.83*

Note. $n = 9666$ individuals, 541 teams. Psych. Safety = Psychological safety. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 24 *(1 of 6)*

Inter-item Correlations Among Latent Variables at the Between-team Level of Analysis via S_B

Team variables	1	2	3	4	5	6	7	8	9
1. Coh1_TA									
2. Coh2_TA	.79*								
3. Coh3_TA	.77*	.68*							
4. Coh4_IC	.75*	.69*	.69*						
5. Coh5_IC	.73*	.65*	.71*	.85*					
6. Coh6_IC	.79*	.76*	.68*	.84*	.79*				
7. Coh7_TC	.73*	.62*	.72*	.74*	.78*	.69*			
8. Coh8_TCr	.57*	.47*	.62*	.60*	.65*	.56*	.70*		
9. Coh9_TCr	.40*	.32*	.44*	.46*	.52*	.42*	.56*	.67*	
10. Con1_TC	-.27*	-.28*	-.29*	-.30*	-.37*	-.29*	-.28*	-.30*	-.34*
11. Con2_TC	-.36*	-.32*	-.39*	-.39*	-.48*	-.35*	-.42*	-.44*	-.46*
12. Con3_TC	-.28*	-.27*	-.32*	-.29*	-.40*	-.31*	-.31*	-.32*	-.34*
13. Con4_RC	**-.53***	-.47*	**-.54***	**-.60***	**-.67***	**-.54***	**-.59***	**-.60***	**-.51***
14. Con5_RC	-.48*	-.43*	**-.54***	**-.52***	**-.61***	**-.46***	**-.55***	**-.57***	-.48*
15. Con6_RC	-.46*	-.39*	-.48*	**-.51***	**-.59***	-.44*	**-.51***	**-.53***	-.49*
16. Con7_PC	-.43*	-.38*	-.48*	-.48*	**-.54***	-.43*	**-.52***	**-.51***	-.48*
17. Con8_PC	-.50*	-.43*	-.53*	-.55*	**-.62***	-.48*	**-.59***	**-.61***	**-.52***
18. Con9_PC	-.41*	-.37*	-.44*	-.45*	**-.51***	-.39*	-.49*	**-.51***	-.47*

Note. $n = 9666$ individuals, 541 teams. Coh = Cohesion, TA = task attraction, IC = Interpersonal cohesiveness, TC = Task cohesiveness. Con = Conflict, TC = Task conflict, RC = Relationship conflict, PC = Process conflict. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 24 *(2 of 6) Continued*

Inter-item Correlations Among Latent Variables at the Between-team Level of Analysis via S_B

Variable	1	2	3	4	5	6	7	8	9
19. Satisfaction1	**.74***	**.63***	**.72***	**.79***	**.81***	**.72***	**.80***	**.74***	**.58***
20. Satisfaction2	**.74***	**.63***	**.74***	**.77***	**.80***	**.71***	**.81***	**.74***	**.56***
21. Satisfaction3	**.75***	**.65***	**.74***	**.77***	**.79***	**.72***	**.78***	**.74***	**.56***
22. Interdepend.1	.35*	.33*	.30*	.36*	.38*	.32*	.37*	.30*	.23*
23. Interdepend.2	.12*	.05*	.10*	.14*	.12*	.11*	.17*	.14*	.06*
24. Interdepend.3r	.21*	.18*	.20*	.21*	.21*	.19*	.28*	.28*	.26*
25. Interdepend.4	.28*	.23*	.18*	.26*	.23*	.28*	.28*	.22*	.20*
26. Interdepend.5	.33*	.25*	.24*	.30*	.27*	.32*	.36*	.24*	.20*
27. Psych. Safety1r	.39*	.33*	.40*	.45*	.50*	.40*	.43*	.46*	.44*
28. Psych. Safety2	.41*	.35*	.39*	.45*	.45*	.40*	.47*	.43*	.40*
29. Psych. Safety3r	.30*	.27*	.28*	.42*	.47*	.35*	.38*	.38*	.36*
30. Psych. Safety4	**.52***	.46*	.47*	**.58***	**.57***	**.54***	**.50***	.48*	.45*
31. Psych. Safety5r	**.55***	.47*	**.52***	**.58***	**.59***	**.53***	**.63***	**.58***	.47*
32. Psych. Safety6	.40*	.32*	.38*	.46*	.46*	.37*	.48*	.47*	.45*
33. Psych. Safety7	**.56***	**.51***	**.57***	**.62***	**.62***	**.58***	**.59***	**.51***	.45*
34. Viability1	**.70***	**.62***	**.69***	**.76***	**.76***	**.72***	**.74***	**.68***	**.50***
35. Viability2	**.71***	**.62***	**.69***	**.75***	**.75***	**.73***	**.73***	**.65***	.47*
36. Viability3r	.41*	.33*	.41*	**.51***	**.51***	.43*	.49*	**.56***	**.53***
37. Liking4	**.67***	**.61***	**.59***	**.77***	**.70***	**.74***	**.62***	**.52***	.38*
38. Liking5	**.68***	**.69***	**.57***	**.71***	**.62***	**.78***	**.54***	.41*	.25*
39. Liking6	**.71***	**.69***	**.61***	**.74***	**.67***	**.82***	**.59***	.48*	.31*

Note. n = 9666 individuals, 541 teams. Interdepend. = Task interdependence. Psych. Safety = Psychological safety. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 24 *(3 of 6) Continued*

Inter-item Correlations Among Latent Variables at the Between-team Level of Analysis via S_B

Variable	10	11	12	13	14	15	16	17	18
10. Con1_TC									
11. Con2_TC	.72*								
12. Con3_TC	.77*	.75*							
13. Con4_RC	.59*	.65*	.59*						
14. Con5_RC	.54*	.68*	.60*	.79*					
15. Con6_RC	.57*	.67*	.57*	.80*	.80*				
16. Con7_PC	.58*	.70*	.58*	.68*	.69*	.70*			
17. Con8_PC	.50*	.61*	.50*	.75*	.72*	.69*	.76*		
18. Con9_PC	.58*	.68*	.58*	.65*	.65*	.65*	.71*	.68*	
19. Satisfaction1	-.35*	-.47*	-.34*	**-.69***	**-.60***	**-.59***	**-.57***	**-.63***	**-.53***
20. Satisfaction2	-.33*	-.47*	-.36*	**-.66***	**-.60***	**-.57***	**-.56***	**-.63***	**-.54***
21. Satisfaction3	-.34*	-.46*	-.35*	**-.67***	**-.59***	**-.55***	**-.54***	**-.62***	**-.52***
22. Interdepend.1	-.13*	-.20*	-.13*	-.22*	-.20*	-.20*	-.23*	-.24*	-.20*
23. Interdepend.2	-.12*	-.12*	-.14*	-.12*	-.12*	-.12*	-.12*	-.16*	-.13*
24. Interdepend.3r	.09*	0.02	.07*	-.11*	-.12*	-.06*	-.08*	-.15*	-.10*
25. Interdepend.4	-0.04	-.08*	-.07*	-.12*	-.13*	-.11*	-.11*	-.14*	-.12*
26. Interdepend.5	.05**	-0.04	-0.01	-.13*	-.12*	-.09*	-.14*	-.16*	-.11*

Note. n = 9666 individuals, 541 teams. Con = Conflict, TC = Task conflict, RC = Relationship conflict, PC = Process conflict. Interdepend. = Task interdependence. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 24 *(4 of 6) Continued*

Inter-item Correlations Among Observed Variables at the Between-team Level of Analysis via S_B

Variable	10	11	12	13	14	15	16	17	18
27. Psych. Safety1r	-.38*	-.46*	-.41*	**-.52***	**-.53***	**-.50***	**-.50***	**-.50***	-.48*
28. Psych. Safety2	-.12*	-.22*	-.09*	-.30*	-.26*	-.26*	-.29*	-.30*	-.20*
29. Psych. Safety3r	-.36*	-.43*	-.31*	-.49*	-.44*	-.51*	-.43*	-.45*	-.39*
30. Psych. Safety4	-.28*	-.37*	-.28*	-.46*	-.41*	-.46*	-.40*	-.43*	-.37*
31. Psych. Safety5r	-.23*	-.34*	-.24*	**-.51***	-.46*	-.44*	-.47*	**-.50***	-.41*
32. Psych. Safety6	-.19*	-.30*	-.19*	-.44*	-.38*	-.41*	-.37*	-.39*	-.35*
33. Psych. Safety7	-.24*	-.38*	-.29*	-.49*	-.46*	-.44*	-.41*	-.44*	-.38*
34. Viability1	-.35*	-.45*	-.36*	**-.64***	**-.56***	**-.54***	**-.54***	**-.62***	**-.51***
35. Viability2	-.36*	-.43*	-.38*	**-.63***	**-.55***	**-.52***	**-.51***	**-.59***	-.48*
36. Viability3r	-.16*	-.26*	-.15*	-.44*	-.37*	-.36*	-.36*	-.43*	-.36*
37. Liking4	-.25*	-.31*	-.26*	-.46*	-.42*	-.40*	-.39*	-.46*	-.38*
38. Liking5	-.23*	-.23*	-.24*	-.38*	-.30*	-.29*	-.29*	-.36*	-.29*
39. Liking6	-.25*	-.29*	-.25*	-.44*	-.36*	-.34*	-.35*	-.40*	-.33*

Note. n = 9666 individuals, 541 teams. Psych. Safety = Psychological safety. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates p < .05. * indicates p < .01.

Table 24 *(5 of 6) Continued*

Inter-item Correlations Among Observed Variables at the Between-team Level of Analysis via S_B

Variable	19	20	21	22	23	24	25	26
19. Satisfaction1								
20. Satisfaction2	.93*							
21. Satisfaction3	.94*	.94*						
22. Interdepend.1	.39*	.39*	.37*					
23. Interdepend.2	.15*	.17*	.16*	.38*				
24. Interdepend.3r	.23*	.21*	.23*	.18*	-.07*			
25. Interdepend.4	.26*	.25*	.26*	.46*	.46*	.21*		
26. Interdepend.5	.29*	.31*	.30*	.50*	.34*	.30*	.67*	
27. Psych. Safety1r	.49*	.47*	.49*	.20*	.14*	.15*	.12*	.10*
28. Psych. Safety2	.46*	.46*	.46*	.20*	.08*	.23*	.13*	.15*
29. Psych. Safety3r	.43*	.39*	.40*	.21*	.10*	.20*	0.02	.07*
30. Psych. Safety4	**.56***	**.55***	**.56***	.27*	.14*	.17*	.20*	.23*
31. Psych. Safety5r	**.65***	**.65***	**.63***	.32*	.12*	.25*	.19*	.27*
32. Psych. Safety6	.49*	.46*	.47*	.23*	.11*	.25*	.12*	.18*
33. Psych. Safety7	**.60***	**.60***	**.59***	.28*	.02	.22*	.14*	.20*
34. Viability1	**.85***	**.82***	**.84***	.33*	.13*	.22*	.24*	.26*
35. Viability2	**.83***	**.80***	**.82***	.32*	.15*	.19*	.24*	.26*
36. Viability3r	**.57***	**.55***	**.57***	.23*	.08*	.25*	.17*	.17*
37. Liking4	**.69***	**.67***	**.68***	.33*	.06*	.23*	.23*	.28*
38. Liking5	**.60***	**.58***	**.60***	.26*	.07*	.16*	.24*	.28*
39. Liking6	**.66***	**.65***	**.66***	.29*	.07*	.18*	.24*	.28*

Note. $n = 9666$ individuals, 541 teams. Interdepend. = Task interdependence. Psych. Safety = Psychological safety. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 24 *(6 of 6) Continued*

Inter-item Correlations Among Observed Variables at the Between-team Level of Analysis via S_B

Variable	27	28	29	30	31	32	33	34	35	36	37	38
27. Psych. Safety1r												
28. Psych. Safety2	.25*											
29. Psych. Safety3r	.57*	.26*										
30. Psych. Safety4	.45*	.54*	.41*									
31. Psych. Safety5r	.48*	.39*	.52*	.46*								
32. Psych. Safety6	.39*	.47*	.39*	.45*	.43*							
33. Psych. Safety7	.42*	.48*	.37*	.54*	.50*	.53*						
34. Viability1	**.50***	.39*	.46*	**.53***	**.60***	.45*	.57*					
35. Viability2	.48*	.38*	.42*	**.51***	**.59***	.42*	.57*	.93*				
36. Viability3r	.39*	.34*	.34*	.44*	.49*	.41*	.36*	.59*	.53*			
37. Liking4	.40*	.35*	.39*	**.51***	**.50***	.40*	**.54***	**.76***	**.74***	.46*		
38. Liking5	.31*	.27*	.27*	.44*	.42*	.27*	.49*	**.68***	**.69***	.32*	.79*	
39. Liking6	.34*	.31*	.30*	.49*	.48*	.33*	**.51***	**.74***	**.74***	.37*	.84*	.90*

Note. n = 9666 individuals, 541 teams. Psych. Safety = Psychological safety. Inter-item correlations ≥ in bold. Inter-item correlations from the same measure in triangles. ** indicates $p < .05$. * indicates $p < .01$.

Table 25

Multigroup Analysis for Measure of Team Cohesion via $S_{B:}$ High & Low $r_{WG(j)}$ & AD median

	$r_{WG(j)}$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	7362.41 (*54*)	7441.61 (*62*)	7442.01 (*74*)	10948.66 (*79*)	6936.40 (*54*)	7441.61 (*62*)	7113.79 (*70*)	7452.88 (*79*)
CFI	.89	.89	.89	.84	.89	.89	.89	.89
TLI	.85	.87	.89	.85	.86	.87	.89	.90
RMSEA	.18	17	.16	.18	.18	.17	.16	.15
CI	[.18, .18]	[.17, .17]	[.16, .16]	[.18, .18]	[.17, .18]	[.17, .17]	[.16, .16]	[.15 .15]
SRMR	.06	.06	.06	.17	.06	.06	.06	.06
CFI Dif.		.00*	.00*	.00*		.00	.00	.00*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG(j)}$ = team-member agreement index for multi-item measures. n = 211 teams for low $r_{WG(j)}$, 8,150 teams for high $r_{WG(j)}$. n = 4,398 teams for low *ADmd*, 3,963 for high *ADmd*. Data frame was split by teams with $r_{WG(j)} > .67$ as the high agreement group and split by teams with average deviation from the median > .35 as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. * = $p < .001$.

Table 26

Multigroup Analysis for Measure of Team Conflict via S_B: High & Low $r_{WG(j)}$ & AD median

	$r_{WG(j)}$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	5320.86 (*52*	5554.25 (*59*)	4019.05 (*66*)	7503.52 (*75*)	4524.78 (*52*)	5088.28 (*59*)	5088.57 (*66*)	8961.57 (*75*)
CFI	.94	.93	.94	.89	.93	.93	.93	.87
TLI	.91	.92	.93	.89	.91	.91	.92	.87
RMSEA	.14	.13	.12	.14	.13	.12	.12	.15
CI	[.13, .14]	[.13, .13]	[.12, .12]	[.15, .16]	[.12, .13]	[.12, .12]	[.12, .12]	[.15, .15]
SRMR	.05	.05	.04	.13	.05	.07	.07	.09
CFI Dif.		.01*	Failed	05*		.00*	.00*	.06*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG(j)}$ = team-member agreement index for multi-item measures. n = 176 teams for low $r_{WG(j)}$, 10,666 teams for high $r_{WG(j)}$ n = 5,618 teams for low *ADmd*, 5,224 for high *ADmd*. Data frame was split by teams with $r_{WG(j)} > .72$ as the high agreement group and split by teams with average deviation from the median > .22 as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. Failed = Model failed to identify. * = $p < .001$.

Table 27

Multigroup Analysis for Measure of Team Psychological Safety via S_B: High & Low $r_{WG}(j)$ & AD median

	$r_{WG}(j)$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	698.12 (*28*)	740.32 (*34*)	740.56 (*40*)	2266.91 (*43*)	712.91 (*28*)	739.25 (*34*)	739.28 (*40*)	1440.87 (*47*)
CFI	.87	.86	.87	.57	.88	.88	.88	.76
TLI	.81	.83	.86	.62	.83	.85	.88	.79
RMSEA	.12	.11	.10	.17	.12	.11	.10	.14
CI	[.11, .13]	[.11, .12]	[.10, .11]	[.16, .18]	[.11, .13]	[.11, .12]	[.10, .11]	[13, .14]
SRMR	.05	.06	.06	.15	.05	.06	.06	.11
CFI Dif.		.01*	.00*	.30*		.00*	.00*	.12*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG}(j)$ = team-member agreement index for multi-item measures. n = 1442 teams for low $r_{WG}(j)$, 1833 teams for high $r_{WG}(j)$ n = 1,638 teams for low *ADmd*, 1,637 for high *ADmd*. Data frame was split by teams with $r_{WG}(j) > .67$ as the high agreement group and split by teams with average deviation from the median > .33 as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. Failed = Model failed to identify. * = $p < .001$.

Table 28

Multigroup Analysis for Measure of Team Satisfaction via S_B: High & Low $r_{WG}(j)$ & AD median

	$r_{WG}(j)$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	506.59 (*38*)	524.23 (*44*)	524.28 (*50*)	2951.03 (*58*)	491.72 (*38*)	508.85 (*44*)	533.87 (*50*)	785.20 (*58*)
CFI	.98	.98	.98	.87	.98	.98	.98	.97
TLI	.97	.97	.98	.97	.97	.97	.98	.97
RMSEA	.08	.07	.07	.16	.08	.07	.07	.08
CI	[.07, .08]	[.07, .08]	[.06, .07]	[.15, .16]	[.07, .08]	[.07, .08]	[.06, .07]	[.07, .08]
SRMR	.05	.05	.05	.16	.05	.05	.05	.05
CFI Dif.		.00*	.00	.11*		.00*	.00	.01*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG}(j)$ = team-member agreement index for multi-item measures. n = 561 teams for low $r_{WG}(j)$, 3,493 teams for high $r_{WG}(j)$. n = 2,126 teams for low *ADmd* and 1,928 for high *ADmd*. Data frame was split by teams with $r_{WG}(j) > .67$ as the high agreement group and split by teams with average deviation from the median > .38 as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. Failed = Model failed to identify. * = $p < .001$.

Table 29

Multigroup Analysis for the Measure of Team Task Interdependence via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median (ADmd)

	$r_{WG}(j)$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	396.01 (*10*)	449.52 (*14*)	449.54 (*18*)	2591.23 (*23*)	332.83 (*10*)	341.80 (*14*)	341.82 (*18*)	599.89 (*23*)
CFI	.94	.94	.94	.63	.95	.95	.95	.90
TLI	.89	.91	.93	.68	.89	.92	.94	.92
RMSEA	.13	.11	.10	.22	.12	.11	.09	.11
CI	[.12, .14]	[.11, .12]	[.09, .11]	[.21, .22]	[.11, .13]	[.10, .12]	[.08, .10]	[.10, .12]
SRMR	.04	.05	.05	.20	.04	.04	.04	.06
CFI Dif.		.00*	.00*	.31*		.00*	.00	.05*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG}(j)$ = team-member agreement index for multi-item measures. n = 678 teams for low $r_{WG}(j)$, 4,127 teams for high $r_{WG}(j)$. n = 2,676 teams for low *ADmd*, 2,129 for high *ADmd*. Data frame was split by teams with $r_{WG}(j) > .67$ as the high agreement group and split by teams with $ADmd > .67$ as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. Failed = Model failed to identify. * = $p < .001$.

Table 30

Multigroup Analysis for the Measure of Team Viability via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median (ADmd)

	$r_{WG}(j)j$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	772.62 (*16*)	1054.26 (*20*)	1055.16 (*24*)	2764.15 (*30*)	832.26 (*16*)	1161.66 (*20*)	1162.08 (*24*)	2223.83 (*30*)
CFI	.98	.97	.97	.91	.97	.96	.96	.93
TLI	.95	.95	.96	.91	.95	.95	.96	.93
RMSEA	.15	.15	.14	.20	.15	.16	.15	.18
CI	[.13, .15]	[.14, .16]	[.13, .14]	[.19, .21]	[.14, .16]	[.15, .17]	[.14, .15]	[.17, .19]
SRMR	.03	.08	.08	.12	.04	.11	.11	.10
CFI Dif.		.01*	.01*	.07*		.01*	.01*	.04*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG}(j)$ = team-member agreement index for multi-item measures. n = 641 teams for low $r_{WG}(j)$, 3,887 teams for high $r_{WG}(j)$. n = 2,313 teams for low *ADmd*, 2,215 for high *ADmd*. Data frame was split by teams with $r_{WG}(j) > .67$ as the high agreement group and split by teams with $ADmd > .47$ as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. Failed = Model failed to identify. * = $p < .001$.

Table 31

Multigroup Analysis for the Measure of Team Liking via S_B: High & Low $r_{WG}(j)$ & Average Deviation Median (ADmd)

	$r_{WG}(j)$				*AD median*			
	Configural	Metric	Scalar	Strict	Configural	Metric	Scalar	Strict
X^2 (*df*)	978.14 (*16*)	1165.41 (*20*)	891.93 (*24*)	1915.70 (*30*)	852.96 (*16*)	873.62 (*20*)	874.65 (*24*)	978.36 (*30*)
CFI	.97	.96	.97	.94	.97	.97	.97	.97
TLI	.94	.94	.96	.94	.95	.96	.97	.97
RMSEA	.16	.16	.13	.17	.15	.14	.13	.12
CI	[.15, .17]	[.15, .17]	[.12, .13]	[.16, .17]	[.14, 16]	[.13, .15]	[.12, .13]	[.11, .13]
SRMR	.05	.05	.04	.09	.04	.04	.04	.04
CFI Dif.		.01*	.00	.00*		.00*	.00*	.00*

Note. S_B = between-team covariance matrix. *ADmd* = Average deviation from the median on scores on a multi-item measure. $r_{WG}(j)$ = team-member agreement index for multi-item measures. n = 121 teams for low $r_{WG}(j)$, 4,407 teams for high $r_{WG}(j)$. n = 2,421 teams for low *ADmd*, 2,107 for high *ADmd*. Data frame was split by teams with $r_{WG}(j) > .67$ as the high agreement group and split by teams with $ADmd > .43$ as the high agreement group. X^2 = Chi-squared. *df* = degrees of freedom. *CFI* = Comparative fit index. *TLI* = Tucker-Lewis index. *RMSEA* = Root mean square error of approximation. *SRMR* = Standardized root mean square residual. *CI* = Confidence interval for *RMSEA*. *CFI Dif.* = difference in CFI from configural model. Failed = Model failed to identify. * = $p < .001$.

FIGURES

Table 32

Participant Characteristics in Data Set by Measure

Measures	# of teams	# of members (*female, male, other*)	Race/ethnicity: Asian	Black/African American	Hispanic/ Latino	Native American	Other	White/ Caucasian	% of # reporting 18 – 25
Team cohesion	8,361	34,400 (*5,284, 9,393, 70*)	1,339	621	1,184	35	411	6,303	91% of 4,148
Team conflict	10,901	44,379 (*7,125, 10,632, 514*)	1,673	824	1,444	46	514	6,971	62% of 441
Team psych. safety	3,275	13,341 (*1,923, 2,170, 17*)	356	215	413	11	144	1.928	90% of 1,545
Team satisfaction	4,054	17,045 (*3,019, 4,054, 32*)	404	265	466	11	162	2,521	93% of 2,403
Team task interdepend.	4,805	20,365 (*3,429, 4,280, 38*)	558	324	561	11	210	2,742	93% of 2,525
Member lik. and viability	19,105	74,852 (*2,726, 3,577, 22*)	510	295	479	14	220	2,882	92% of 2,571

Note. Team psych. safety = team psychological safety. Team task interdepend. = team task interdependence. Member lik. and viability = team-member liking and viability. The totals on gender and race/ethnicity do not equal the total number of participants per sample and are based upon what was reported. The number of participants reporting to be between 18 to 25 years of age are a portion of the total sample for each measure and specified above.

Differences in Modeling the Between- and Within-team Levels versus the Individual Level of Analysis	**Types of Constructs Typically Operating at Respective Levels of Analysis**
Between-team Models the influence of the team on member's scores by examining how team membership influences patterns of responses among team members.	Between-team Team Emergent States & Processes (e.g., team cohesion, conflict, satisfaction, collective efficacy, and diversity; Waller et al., 2016; Woehr et al., 2015).
Within-team Models the influence of disagreement/differences among team members by examining the deviation in scores within a team. In essence removing the influence of team membership on people's scores.	Within-team Subgroups and dispersion constructs (e.g., distributed collaboration and differences in perceptions of team conflict; Loignon et al, 2019; Carton & Cummings, 2012).
Individual/Person Models the specific influence of the team member on their own scores; does not take into account a grouping factor.	Individual/Person Individual Differences (e.g., personality, and self-efficacy; Raykov, 1998; Tasa, Taggar, & Seijts, 2007).

Figure 1. Description of what each level of analysis takes into consideration when reflecting the influence of a latent construct and examples of constructs that primarily operate at those respective level of analysis.

Differences in Modeling the Between-team, Between-person, and Within-person Levels versus the Individual Level of Analysis	**Types of Constructs Typically Operating at Respective Levels of Analysis**
Between-team Models the influence of the team on member's scores by examining how team membership influences patterns of responses among team members.	Between-team Team Emergent States & Processes (e.g., Overall team viability and team-member liking across the team).
Between-person In a three-level model containing within person ratings and a grouping factor (e.g., team membership), this level is conceptualized as modeling the influence of differences/deviations between team-members across events or targets (e.g., people, time points, events).	Between-person Constructs related to the differences in people's general perception within a team/group across time points, events, or targets (e.g., team members' general tendency in how they view the viability of their relationships with other members or the extent to which they typically like other team members in their team).
Within-person Models the influence of differences in people's own scores when there are multiple ratings done by each person while removing the influence of the grouping factor(s). In a three-level model this level accounts for the influence of two grouping factors (e.g., individual, team) and models the variation in peoples' scores due to differences across time, target, or time point in a team/group).	Within-person Changes in people's perceptions over time or differences in people's perceptions across people in a team/group. Constructs are specific to time points or differing targets (e.g., diary studies, event-based constructs, within-person coping strategies, differences in member's perceptions of specific relationships with other team members; Breevaart et al., 2012; Roesch et al., 2010)
Repeated Observations Investigates distinct factor structures by comparing differences in models based on people's scores at one time point or target without accounting for a grouping factor (e.g., person). This is done via multi-group analysis in which each grouping of observations (e.g., time point, target, event) is examined for differences in factor structures based on the grouping.	Repeated Observations Examines for distinct constructs based on life experiences, time points, or target. Constructs are specific to an understanding of an event, a time point, or target (e.g., understanding trauma after a traumatic experience, measuring resilience from people who have *bounced back* from a life set back)

Figure 2. Description of what each level of analysis takes into consideration when reflecting the influence of a latent construct with examples of constructs that primarily operate at those respective level of analysis.

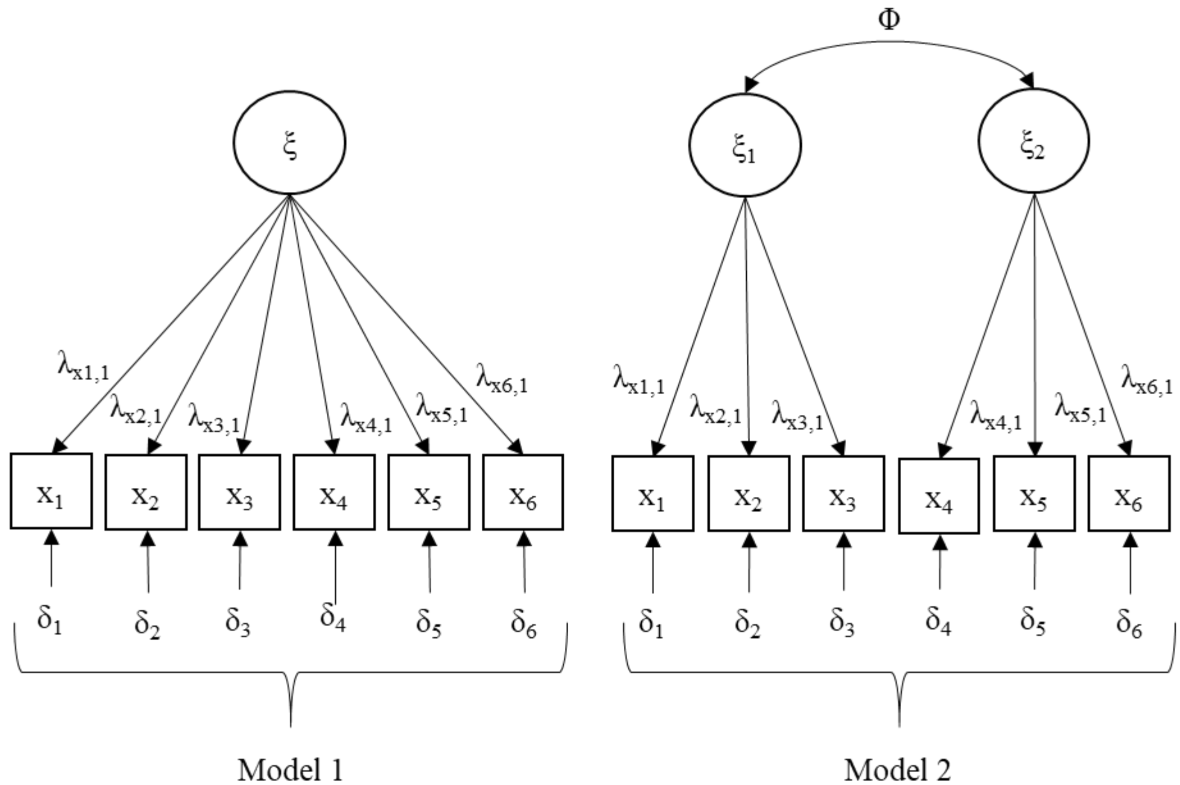

Figure 3. Measurement model for one factor and two factor solution. Φ = (i.e., Phi) covariance matrix of latent factors. ξ = latent factor. *x* = observed variable. λ = (i.e., lambda) factor loadings. $\lambda_{X1,1}$ - $\lambda_{X6,1}$ represent the respective factor loadings on each of the indicators/observed variables ($x_1 - x_6$); while δ_1 - δ_6 represent the unique and error variance (i.e., residual variance) associated with each observed variable in a CFA.

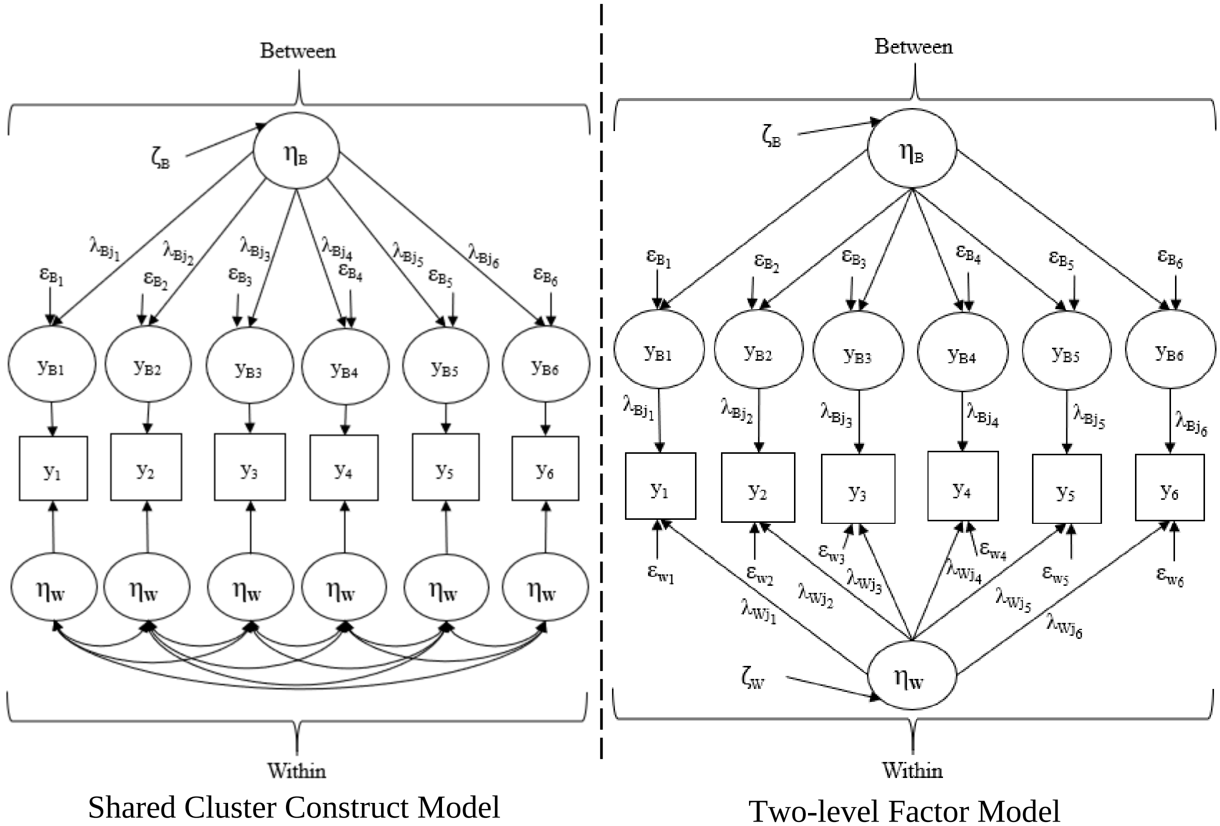

Figure 4. Measurement models of multilevel factor models at the between and within group level of analysis. The shared cluster construct model specifies the lower level as residual variance, whereas, the two-level factor model specifies latent factors at the within- and between-team levels. ζ = (i.e., Zeta) indicates the residual variance associated with a level of analysis. η = (i.e., eta) the higher levels of analysis as a latent factor. λ = (i.e., lambda) factor loadings. ε = (i.e., varepsilon) residual variance for each indicator mean or observed variable. **B** = between group level of analysis. **W** = within group level of analysis. $y_{B1} – y_{B9}$ = group means for each indicator. $y_1 – y_9$ = observed variable for each indicator.

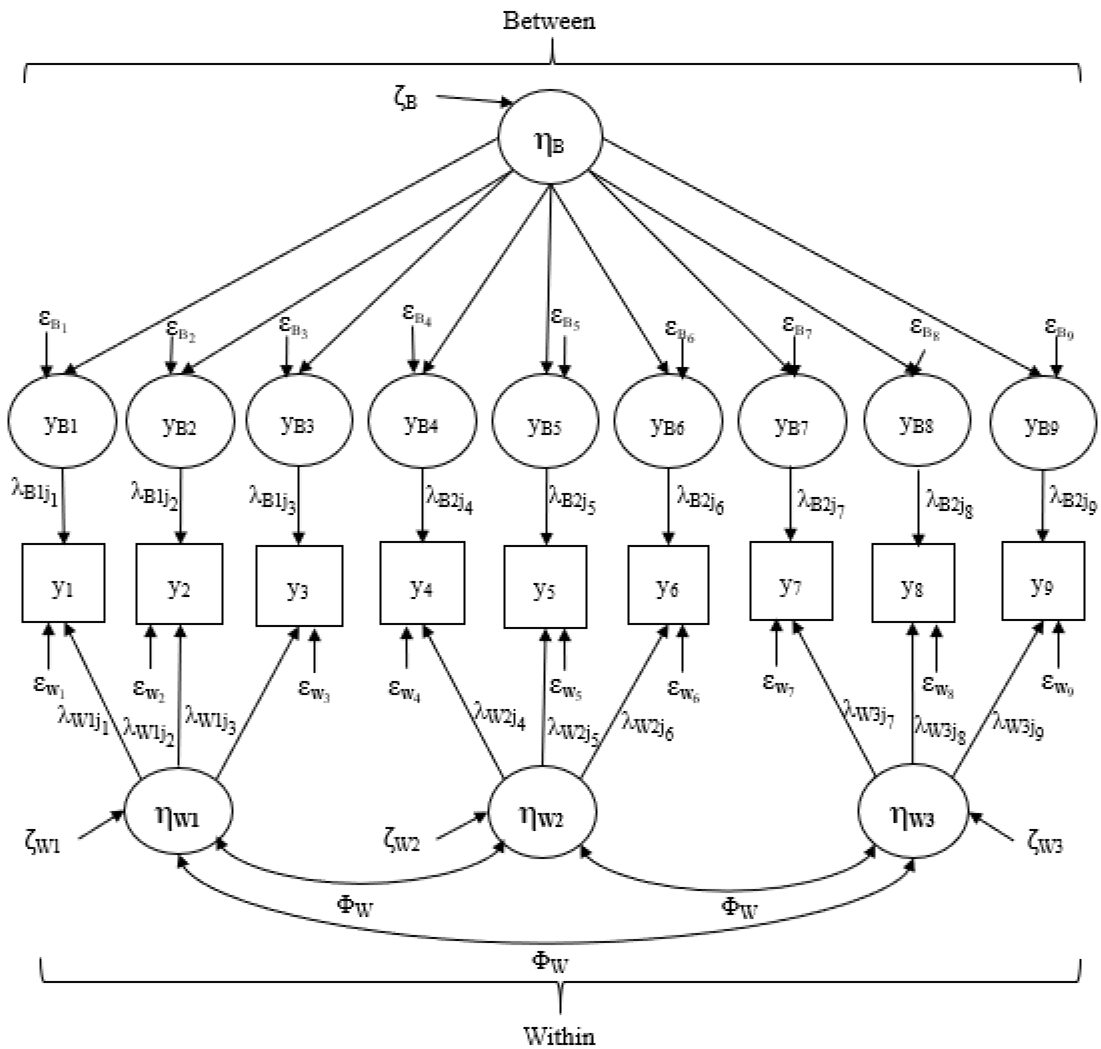

Figure 5. Measurement model of a multilevel factor revealing a general latent factor at the between group level and 3 factors at the within group level of analysis. Φ = (i.e., Phi) covariance matrix of latent factors. ζ = (i.e., Zeta) indicates the residual variance associated with a level of analysis. η = (i.e., eta) the higher levels of analysis as a latent factor. λ = (i.e., lambda) factor loadings. ε = (i.e., varepsilon) residual variance for each indicator mean or observed variable. **B** = between group level of analysis. **W** = within group level of analysis. $y_{B1} – y_{B9}$ = group means for each indicator. $y_1 – y_9$ = observed variable for each indicator.

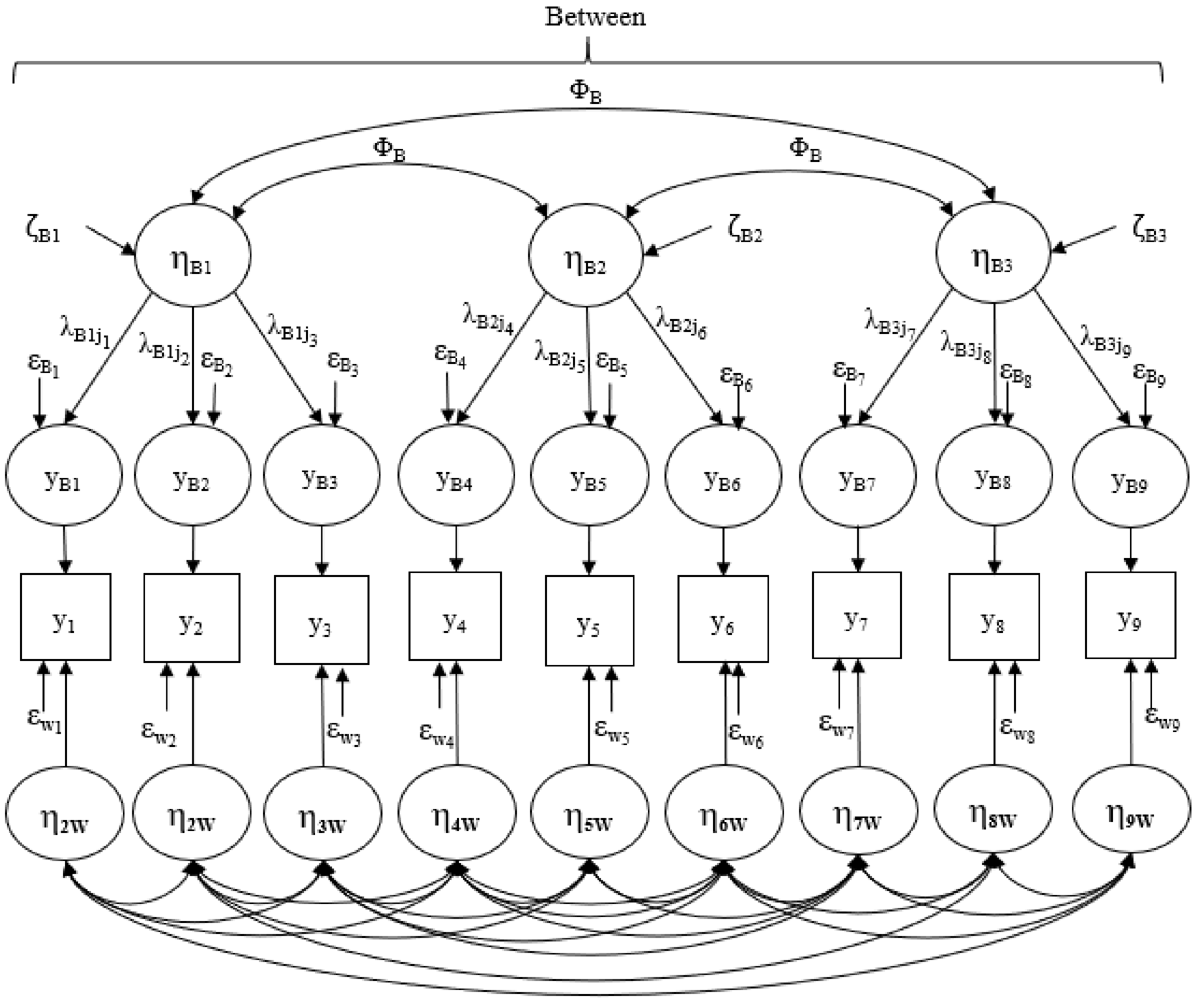

Figure 6. Measurement model of a shared construct with three factors at the higher level and the lower level is modeled as residual variance. Φ = (i.e., Phi) covariance matrix of latent factors. ζ = (i.e., Zeta) indicates the residual variance associated with a level of analysis. η = (i.e., eta) the higher levels of analysis as a latent factor. λ = (i.e., lambda) factor loadings. .ε = (i.e., varepsilon) residual variance for each indicator mean or observed variable. **B** = between group level of analysis. **W** = within group level of analysis. $y_{B1} - y_{B9}$ = group means for each indicator. $y_1 - y_9$ = observed variable for each indicator.

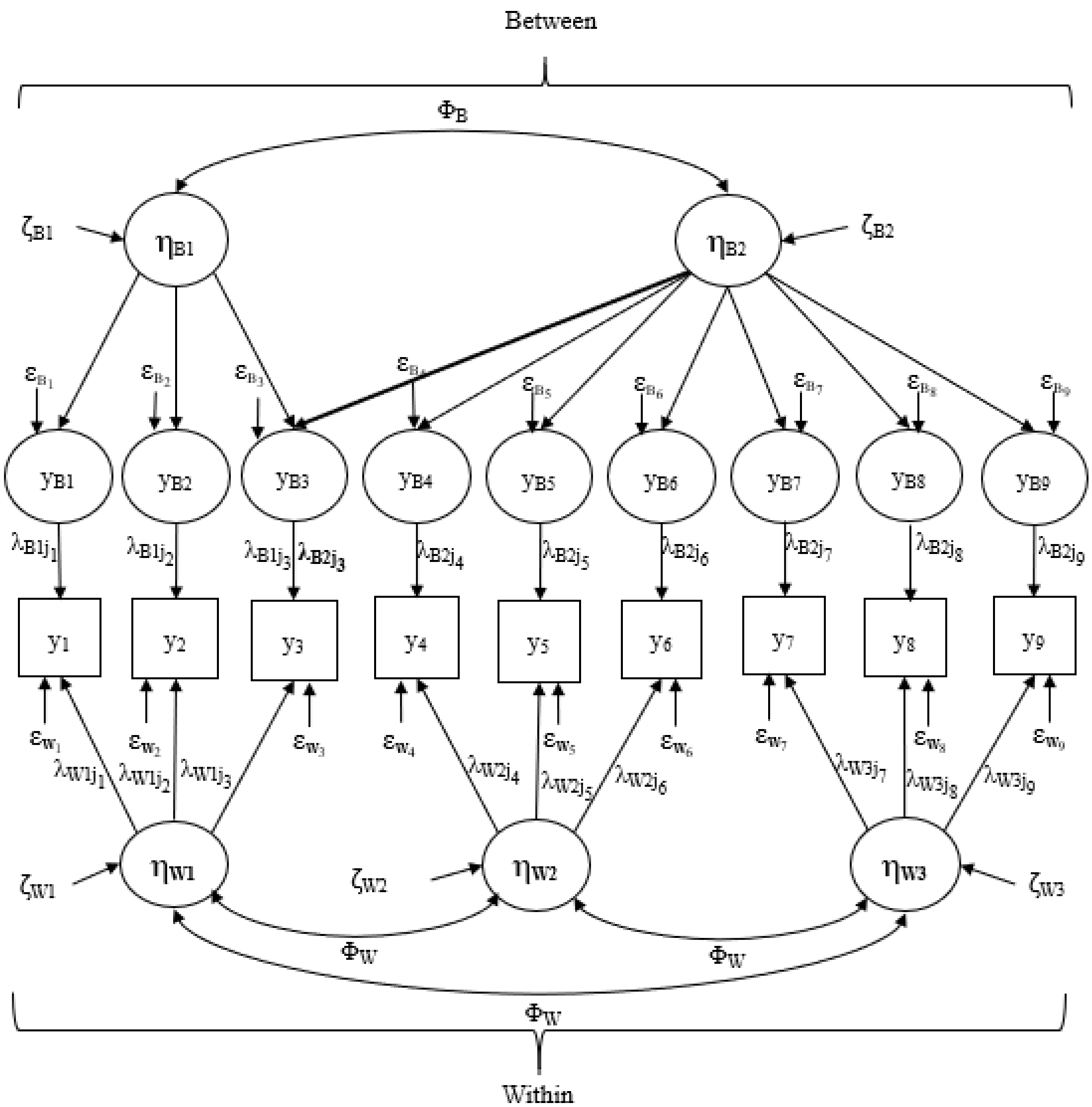

Figure 7. Measurement model of a multilevel factor revealing partial configural isomorphism with a simplified factor structure at the higher level. This model reflects both the between- and within- level of analysis with 2-factors at the team level, 3-factors at the within level, and cross-loading at the team level for an indicator (i.e., y_4) in bold. Φ = (i.e., Phi) covariance matrix of latent factors. ζ = (i.e., Zeta) indicates the residual variance associated with a level of analysis. η = (i.e., eta) the higher levels of analysis as a latent factor. λ = (i.e., lambda) factor loadings. ε = (i.e., varepsilon) residual variance for each indicator mean or observed variable. **B** = between group level of analysis. **W** = within group level of analysis. y_{B1} – y_{B9} = group means for each indicator. y_1 – y_9 = observed variable for each indicator.

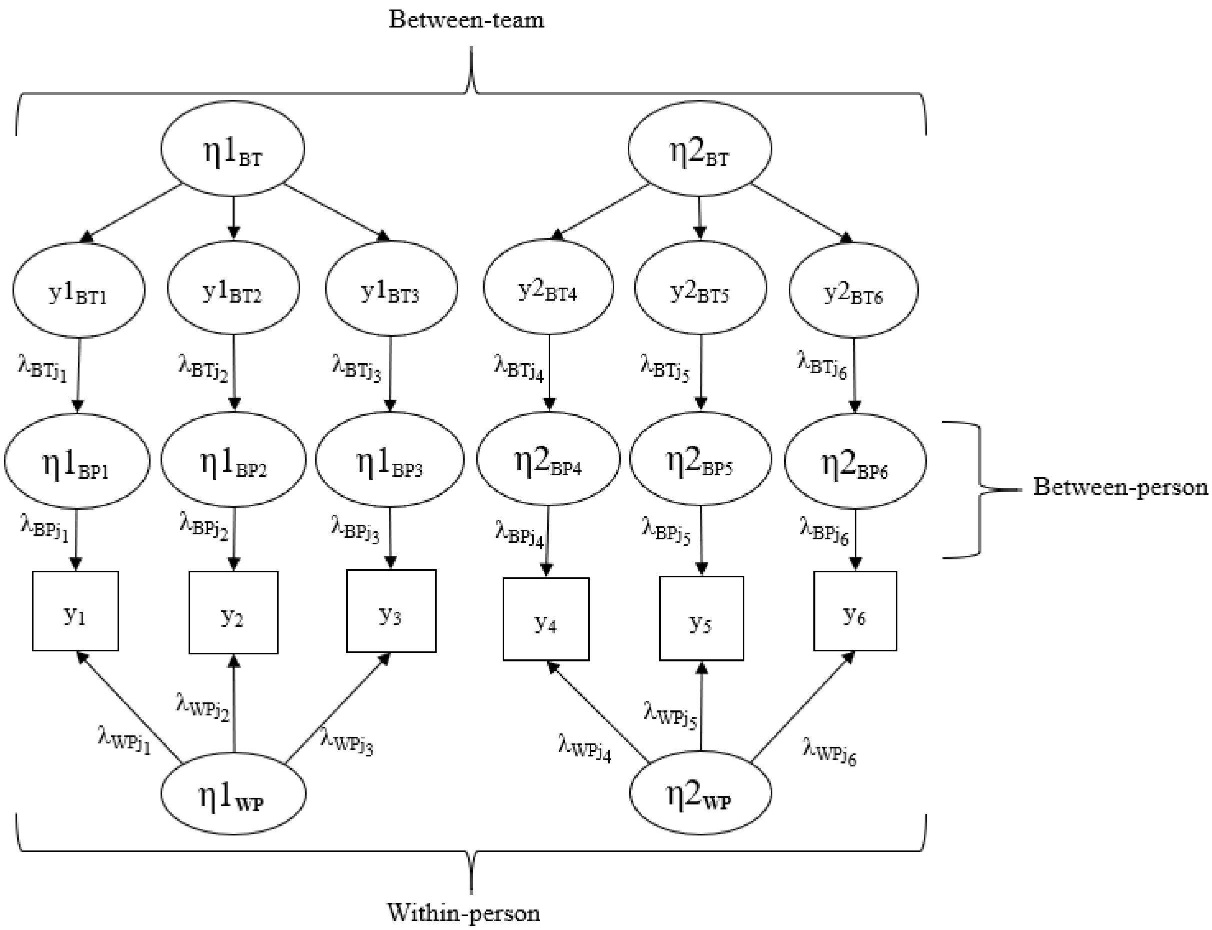

Figure 9. Three-level multilevel factor model revealing at a minimum strong configural isomorphism. This model reflects within-person, between-person, and between-team levels of analysis with two factors at the within-person and between-person levels. η = (i.e., eta) latent factor. λ = (i.e., lambda) factor loadings. **BT** = between team level of analysis. **BP** = between person level of analysis. **W** = within group level of analysis. $y_{B1} – y_{B9}$ = means for each indicator at the respective level of analysis. $y_1 – y_9$ = observed variable for each indicator.

APPENDIX A

To categorize the degree of psychometric isomorphism (e.g., partial configural, strong configural, weak metric, and strong metric isomorphism) for each measure of a common team construct, the following aspects of a multilevel confirmatory factor analysis (MCFA) is assessed and interpreted: ICC(1), model fit, factor variances and loadings, and residual variance (Dyer et al., 2005; Geldhof et al., 2014; Tay et al., 2014). Each psychometric assessment of these measures and their indicators provides evidence for which level of analysis the construct primarily operates and how/if the measure operates at different levels. The current study examined well-established measures of constructs; therefore, the focus is on confirmatory as opposed to exploratory factor analytic techniques (See Dedrick & Greenbaum, 2011 for a review). The following section briefly reviews the equations associated with CFA and MCFA, compares CFA and MCFA techniques, and provides an in-depth discussion detailing the different parts of an MCFA in relation to measures of consensus constructs.

CFA & MCFA Equations

For single level of analysis, a measure's quality is examined within the factor analytic framework via an EFA and a CFA (Bryant & Yarnold, 1995; Crocker & Algina, 1986). In a multilevel context, FA is conducted in a structural equation modeling framework via an MEFA and a MCFA to inspect a measure's properties across levels of analysis (e.g., individual, within-team, and between-team; B. O. Muthén, 1994). For clarity purposes, standard notations are used for a CFA following Bryant & Yarnold (1995) and MCFA notations are derived from Muthén & Muthén's work (B. O. Muthén, 1994; L. K. Muthén & Muthén, 1998-2017). (See Figure 3 and 4.)

A MCFA allows researchers to compare levels of analysis by assessing a measure's latent factor(s)'s overall model or component parts, evaluate the degree of psychometric isomorphism, examine for evidence of convergent validity, and establish reliability estimates. This is achieved by decomposing the variance in observed scores. The above explanations and related figures (Figures 3 – 9) can also be expressed in equation form. Specifically, the observed scores ($\mathbf{x}$) in a CFA is a result of the ability of indicators ($\boldsymbol{\lambda}$, factor loadings) to assess a construct ($\boldsymbol{\xi}$, latent factor) plus any random or measurement error ($\boldsymbol{\delta}$, residual error):

$$\mathbf{x} = \boldsymbol{\lambda}\boldsymbol{\xi} + \boldsymbol{\delta}$$

In an MCFA, the variance in observed scores is broken down further by accounting for influences from the within ($\boldsymbol{\eta}_{\mathbf{W}}$) and between ($\boldsymbol{\eta}_{\mathbf{B}}$) team level of analysis. In other words, observed scores are explained by the amount of deviation within a team and the influence of team membership:

$$\mathbf{x} = \boldsymbol{\eta}_{\mathbf{W}} + \boldsymbol{\eta}_{\mathbf{B}}$$

Breaking down this equation further, we see how the factor loadings and residual error associated with each level of analysis influence the observed scores. Specifically, observed scores are the result of the indicators' ability to capture within ($\boldsymbol{\lambda}_{\mathbf{W}}\boldsymbol{\xi}_{\mathbf{W}}$) and between ($\boldsymbol{\lambda}_{\mathbf{B}}\boldsymbol{\xi}_{\mathbf{B}}$) level variance associated with a latent factor is estimated, and the residual error at each respective level of analysis ($\boldsymbol{\varepsilon}_{\mathbf{W}}$ and $\boldsymbol{\varepsilon}_{\mathbf{B}}$):

$$\mathbf{x} = (\boldsymbol{\lambda}_{\mathbf{W}}\boldsymbol{\xi}_{\mathbf{W}} + \boldsymbol{\varepsilon}_{\mathbf{W}}) + (\boldsymbol{\lambda}_{\mathbf{B}}\boldsymbol{\xi}_{\mathbf{B}} + \boldsymbol{\varepsilon}_{\mathbf{B}})$$

As seen in the formulas above, all sources of variance are not accounted for if the grouping factor (e.g., team membership) in the data is not modeled in an FA (see Li et al., 1998; B. O. Muthén, 1994; Stapleton et al., 2016 for a more indepth review of equations). Therefore, measures of team consensus constructs which aggregate members scores in teams need to be analyzed multilevel factor analytic framework.

Theory and measurement intertwine as this study seeks to uncover if the pattern of people's responses due to team membership (i.e., between-team level) explains variance differently in a latent variable's factor structure and loadings than people's responses in general regardless of team membership (i.e., individual level). The within-team level of analysis is not of theoretical interest for comparison in common measures of team constructs because, as shown in Figure 1, the within-team level models deviations from the average of team members' responses which is different from modeling people's general tendency regardless of what team in which they are a member (i.e., individual level).

Comparing Figure 3 and 4 highlights the differences in modeling the variation at different levels of analysis and based on the theoretical development of the construct. As discussed previously, Figure 3 reflects a traditional CFA model in which no grouping factor is modeled as an influence of variation on observed scores. Alternatively, Figure 4 represents two models in which the examination of variation among indicators is due to the within- and between-team level of analysis for a single factor model. The first model is a shared cluster construct model in which the latent factor is modeled at the between level and variation at the within level is modeled as residual error. There are six indicators for the lower level of analysis representing the observed scores of the measure's items as influenced by team membership and individual differences (i.e., $y_1 – y_6$) and six indicators for the between-team level of analysis representing the latent mean of the measure's items as influenced by team membership ($y_{B1} – y_{B6}$). For team consensus constructs, a shared cluster construct model is appropriate as it models differences among team members as part of error. (See Stapleton et al., 2016 for a review of MCFA models).

In the second model − a two-level factor model, the within (η_W) and between (η_B) level of analysis become latent factors. The within level portion (η_W) is similar to a typical CFA model but reflects the influence of the deviation within a team on the observed variables (y_1 - y_6). The between level factor (η_B) represents the influence of team membership on indicator latent means (y_{B1} - y_{B6}) which, in turn, influences the observed scores (y_1 - y_6). The two-level factor model in Figure 4 models a between-team latent factor (e.g., team construct) and a within-team factor (e.g., deviation construct) simultaneously.

CFA & MCFA Results

As made evident in the equations above, misalignment in theory and measurement is problematic in teams' research when a measure's ability to capture a construct is evaluated based on team-member perceptions rather than how team membership influences the patterns of people's responses because team constructs are meant to capture an aspect of the team not people's perceptions in a team. Understanding the cross-level differences in measures' psychometric properties is examined via model fit, factor loadings, and residual variances in a factor analytic framework.

Model Fit. When evaluating measures of team constructs collected with data from individuals, researchers use factor analysis to assess the overall ability of the measure to capture team phenomenon via model fit indices; however, model fit is not typically assessed at the between-team level. Specifically, at the individual level via a CFA researchers use a variety of model fit indices assessing the hypothesized relationships between the observed and latent variables (Hsu et al., 2015; Marsh et al., 1988; Schreiber et al., 2006). In other words, model fit indices reflect the extent the proposed model fits the observations in the data. These indices (e.g., X^2, *TLI, RMSEA, CFI, SRMR-W,* and *SRMR-B*) provide insight as to the number of factors (i.e., latent variables) and helps researchers compare theoretically relevant alternative measurement models of latent variables (P. Bentler, 1990; Satorra & Bentler, 2010; Steiger, 1998; Tucker & Lewis, 1973). While a variety of model fit indices are appropriate in a CFA, Hsu and colleagues (2015) find common fit indices are not sensitive to model misspecification at the within-group and between-group levels. Therefore, it is not appropriate to report model fit indices in a MCFA that are typically reported in a CFA for measures designed to capture multilevel phenomena. Table 2 reflects the level of analysis, appropriate model fit indices, and index description as relevant to a MCFA.

When comparing alternative measurement models, the model with the best fit psychometrically is assumed to more accurately capture a latent variable (i.e., construct). For example, in a CFA framework Figure 3 reflects two potential models based on six different indicators at a single level of analysis. The first model has all indicators (x_1 - x_6) being influenced by one latent factor (ξ); whereas, the second model has two distinct factors (ξ_1 & ξ_2) explaining the shared variance for three indicators each (x_1 - x_3 and x_4 - x_6, respectively). Model fit indices help researchers uncover which factor model more accurately reflects the construct's factor structure based on the observations in their data. When the data contains a theoretically relevant grouping factor (e.g., within person, team, or organization), higher levels of analysis are modeled in an MFA framework (Stapleton et al., 2016). In this scenario, model fit indices via an MFA are examined at each level of analysis (e.g., individual, within-team, between-team) and compared with a model that includes both the influence of the lower level (e.g., within-team) and higher level (e.g., between-team; See Figure 4).

Factor Loadings and Variances. Factor loadings provide insight regarding the ability of a measure's item to capture a construct and evidence for convergent validity. Specifically, the pattern and magnitude of factor loadings clarify the ability of items to capture constructs at different levels of analysis while the relationship between the indicators and a latent factor(s) relates to convergent validity (Asparouhov et al., 2015; Gerbing & Anderson, 1988; Jak, 2019; Reise et al., 2005). The next section covers what factor loadings are, their role in understanding team phenomenon, and how they relate to convergent validity in a multilevel context.

The factor loadings (lambda, λ) of a measure's indicators reflect the strength of a relationship between an indicator and a latent factor (F. B. Bryant & Yarnold, 1995; Crocker & Algina, 1986; Dyer et al., 2005). In Figure 3, $\lambda_{X1,1}$ - $\lambda_{X6,1}$ represent the respective factor loadings on each of the indicators/observed variables ($x_1 - x_6$) in a CFA. In a MCFA, factor loadings are also calculated to reflect the influence of the within (e.g., λ_{Wj1} - λ_{Wj6}) and between (e.g., λ_{Bj1} - λ_{Bj6}) level of analysis (See Figure 4). Specifically, the degree to which the $\eta_{\mathbf{W}}$ influences observed scores is estimated via λ_{Wj1} - λ_{Wji}; while the influence of $\eta_{\mathbf{B}}$ on the latent means of the observed scores is estimated via λ_{Bj1} - λ_{Bji} (see Figure 4 – 6). The factor loadings associated with each level of analysis are examined via their respective factor matrix.

A factor matrix (Lambda, Λ) contains the pattern of factor loadings (i.e., the "matrix of coefficients regressed from the latent factor to observed variables" in a CFA; Byrne et al., 1989, p. 457). In a MCFA, a factor matrix is created for each level of analysis in which the individual level factor matrix (Λ_I) is derived from the factor loadings from a CFA (e.g., $\lambda_{X1,1}$ - $\lambda_{X6,1}$; see Figure 2), within matrix (Λ_W) contains the factor loadings influenced by the deviation in scores within a team (e.g., λ_{Wj1} - λ_{Wj6}), and the between matrix (Λ_B) represents the factor loadings influenced by aspects of team membership (e.g., λ_{Bj1} - λ_{Bj6}; see Figure 3). If the factor loadings are consistent across levels, then the level of analysis does not influence an indicator's ability to capture a latent factor. More specifically, comparing the factor matrices at different levels of analysis reveals the degree to which factor loadings (λ) can vary across levels and how this differs among indicators. Variation across levels occurs when the strength of an indicator(s) is driven by one level of analysis over another or due to spurious effects (D'Haenens et al., 2012; B. O. Muthén, 1994; Ryu, 2014; Stapleton et al., 2016). In other words, the structure, pattern, and magnitude/strength of the factor loadings can vary as a function of the level of analysis and is examined via the factor matrices (e.g., Λ_I, Λ_W, and Λ_B).

For measures of team consensus constructs, understanding how a measure's items functions in a multilevel context is established by comparing the factor loadings across levels of analysis via examining the factor matrices at the individual (Λ_I) and between-team (Λ_B) levels as there is not theoretical reason for the loadings to be relevant at the within-team level. In other words, these measures inherently require a degree of consistency in scores among team members, not deviation in scores within the team. Therefore, the ability of a measure's items to capture team phenomenon via factor loadings at the between-team level of analysis is evaluated and compared to what is found at the individual level. This comparison will shed light on any potential consequences of misalignment in measurement and theory when drawing conclusions about item quality using data collected from individuals for team phenomena.

Differences in factor loadings at the individual and between-team level of analysis are categorized by the degree of psychometric isomorphism (See Table 1). The less stringent the standard, the greater chance for potential consequences of misalignment in measurement and theory. *Partial configural isomorphism* is mainly examined via model fit indices and is further investigated by examining the factor matrix for potential cross-loadings at the between-team level that do not occur at the individual level (See Figure 4). In this scenario, an indicator may tap two related constructs at a higher but not a lower level of analysis. Strong configural isomorphism indicates that the factor structure is consistent across levels and the measure's items adequately capture relevant aspects of the latent factor. Evidence of *weak metric isomorphism* is established by comparing the relative ordering of factor loadings via examining the Λ_I and Λ_B; while, *strong metric isomorphism* is established by examining if the magnitude of these factor loadings are consistent in the Λ_I and Λ_B (B. O. Muthén, 1994; Tay et al., 2014).

Specifically, a measure of a common team construct has weak metric isomorphism when the factor matrices (i.e., Λ_I and Λ_B) reveal the indicators have the same rank order (i.e., relative ordering) from least to greatest factor loadings at the individual and between-team levels. *Strong metric invariance* occurs when a measure reveals the same relative order and magnitude of factor loadings at the individual $\lambda_{X1,1}$ - $\lambda_{Xi,i}$) and between-team (λ_{Bj1} - λ_{Bji}) level of analysis. In other words, the degree of psychometric isomorphism in these measures of common team constructs is reflected in the consistency of their factor loadings (i.e., structure/pattern, magnitude, and relative ordering) at the individual and team levels. It is important to note that theoretically a measure can capture higher-level phenomenon adequately with weaker factor loadings at lower levels of analysis (D'Haenens et al., 2012; Sorra & Dyer, 2010). Therefore, the magnitude of the factor loadings may not be consistent across levels of analysis in measures of common team constructs.

This means that even if factor loadings are in the acceptable range in a CFA, the strength of these indicators/items to capture team level phenomenon is not accurately assessed unless the between-team level − in which the construct primarily operates − is taken into account. This occurs because the between-team level factor loadings reflect how the latent factor (i.e., η_B**)** influences scores on the indicator latent means (e.g., $y_{B1} – y_{B9}$) as opposed to the individual level in which the factor loadings reflect the relationship between the observed scores and latent factor

(B. O. Muthén, 1994). However, there are important norms to keep in mind when evaluating the magnitude of factor loadings discussed below.

Regarding their magnitude, researchers typically retain indicators (i.e., keep the item in the final measure) in an EFA with a factor loading between .30 and 1. While there are other things to consider when assessing a measure's indicator theoretically and psychometrically, indicators with a .30 factor loading are assumed to weakly tap a latent factor while .70 reveals a strong association with a factor (Schmitt & Sass, 2011). In the current study, it would be problematic if factor loadings dropped below .30 at the team level of analysis but not unexpected if they increase at the team level as I am examined team level constructs.

In addition to indicating the degree of psychometric isomorphism, factor loadings provide evidence for convergent reliability. Specifically, Gerbing and Anderson (1988) recommend examining the factor matrix among indicators in a CFA as it provides more stringent test of convergent validity. Specifically, to establish convergent validity each indicator should load onto a specific factor, not multiple factors. This remains true in a multilevel context. Therefore, it is vital the relationship between observed and latent variables at the level of analysis the construct theoretically operates is psychometrically evaluated. In other words, we don't know how valid our measures are if we do not examine the factor matrix at the appropriate level of analysis.

For example, Figure 7 reveals a simplified factor structure at the between-team level of analysis with two factors ($\eta_{\mathbf{B1}}$ and $\eta_{\mathbf{B2}}$) influencing the latent mean of one indicator (y_{B3}). As some cross-loading is likely to occur in related constructs (e.g., subdimensions of team cohesion or conflict), Asparouhov, Muthén, & Morin (2015) argue that this is not adding 'noise' in measurement but rather provides information as to how a construct influences an indicator. Therefore, if the cross-loading becomes too great (i.e., ≥ .30 on a factor loading for more than one distinct factor) then the indicator would likely be removed in an EFA. That is because, an indicator needs to clearly identify the construct of interest without significantly tapping into a distinct but related construct. While cross-loading items is typically addressed in an EFA, cross-loading in a CFA framework can also be addressed by examining the correlations among measures' indicators (Prudon, 2015).

Residual Variance. Another important aspect in examining the quality of a measure is residual variance. Residuals are the "element-wise difference between observed and model-implied covariance matrix" and estimated at the individual, within-team, and between-team level in a MCFA (see Figure 2; Kim et al., 2016, p. 887). In other words, residual covariances in a factor analysis highlight the discrepancy in the observed and estimated model. As with the current study, the ability of a CFA & MCFA to fully decompose sources of variance not related to a latent factor (i.e., residual variance) is limited without a longitudinal approach (Marsh & Grayson, 1994). This is because without a longitudinal approach, the two components of residual error (i.e., random and item specific measurement error) cannot be untangled (Lubke & Dolan, 2003). Regardless, even when investigated at a single time point this 'noise' in measurement can be of substantive interest in consensus models as residual covariance differs across levels of analysis and influences a measure's reliability in a multilevel context (Geldhof et al., 2014).

In a CFA, residuals (delta, δ) refer to variance due to both item specific and random measurement error that is not associated with the latent factor at the level of data collection (See Figure 2; Crocker & Algina, 1986; Marsh & Grayson, 1994). In other words, residuals in a CFA reflect the difference between the observed value and the variation not associated with the latent factor for each indicator. In a MCFA, residuals (varepsilon, ε) at the within and between level of analysis (ε_W, and ε_B, respectively) reflect the amount of variation not associated with that specific level of analysis (See Figures 4 – 8). At the within-team level, residuals (ε_W) represent the variation in observed scores not explained by the within-team latent factor (η_W). At the between-team level, residuals (ε_B) represent the variation in indicators not explained by their respective latent means (y_B). Additionally, MCFA assesses the residual variance associated with the individual (Zeta, ζ), the within-team (ζ_W), and between-team (ζ_B) levels of analysis. In MCFA, residuals for each indicator and level of analysis should be reported along with the factor loadings as residual variance provides information as to systematic error variance regarding the level of analysis and is essential in calculating the measure's composite reliability (Geldhof et al., 2014; Kim et al., 2016).

Residuals provide two important pieces of information in the current study. First, factor models can be incorrectly specified without properly accounting for residual variance at higher levels of analysis (Lubke & Dolan, 2003; Meredith, 1993). In this scenario, the factor models based on a CFA would be incorrect if the residuals varied from the individual to between level of analysis because variance associated with the lower level is minimized at the higher level (van Mierlo et al., 2009). In other words, by examining residual variance across levels of analysis researchers can uncover sources of systematic variance associated with a specific level of analysis.

Second, reliability estimates using Chronbach's alpha leads to false conclusions regarding the measures' quality if they operate in multilevel context. A measure's reliability must be estimated at the level the construct is hypothesized to operate (e.g., between-team for team consensus constructs) not at the level in which data is collected (e.g., individual). This is problematic as previous research estimating the reliability of team consensus constructs often do not account for differences in residual variance from the individual to the between-team level of analysis (Edmondson, 1999; Jehn & Mannix, 2001; Loughry & Tosi, 2008; Van der Vegt et al., 2001). Specifically, previous research likely underestimates the quality of the measure when evaluating these measures at the individual level (Geldhof et al., 2014). This is important because residual variance at different levels of analysis provides insight as to how model misspecification at various levels can occur and informs the measure's estimate of reliability (Bollen & Arminger, 1991; Geldhof et al., 2014). In summary, the residual variance and composite reliability will differ at the individual and between level of analysis due to how the variance is modeled.

APPENDIX B

MCFA Cohesion Syntax in R

Step 1: CFA of sample total covariance matrix at the individual level

1 factor individual level

```
library(lavaan)

library(semTools)


onefactor <- '

Cohesion =~  Coh1_TA  + Coh2_TA  +  Coh3_TA  + Coh4_IC  + Coh5_IC  + Coh6_IC

  + Coh7_TC  + Coh8_TCr  + Coh9_TCr '


#fiml - full information liklihood; mlr = ""MLR" for maximum likelihood esti
mation with robust 'Huber-White' standard errors and a scaled test statistic
which is asymptotically equivalent to the Yuan-Bentler T2-star test statisti
c

fit1 <- sem(onefactor, data = Data_T2_Coh2, missing = "fiml", estimator = "m
lr")


summary(fit1, fit.measures=TRUE,rsquare=TRUE,standardized=TRUE)


# Obtain Omega

reliability(fit1)

= "tree")
```

2 factor individual level

```
library(lavaan)

library(semTools)


#Two factor cohesion combining task attraction and commitment onto one task-
oriented factor

twofactor <- '

Task=~Coh1_TA  + Coh2_TA  +  Coh3_TA  + Coh7_TC  + Coh8_TCr  +    Coh9_TCr
```

```
Interpersonal=~Coh4_IC  + Coh5_IC  + Coh6_IC'


#fiml - full information liklihood; mlr = ""MLR" for maximum likelihood esti
mation with robust 'Huber-White' standard errors and a scaled test statistic
which is asymptotically equivalent to the Yuan-Bentler T2-star test statisti
c
fit2 <- sem(twofactor, data = Data_T2_Coh2, missing = "fiml", estimator = "m
lr")


summary(fit2, fit.measures=TRUE,rsquare=TRUE,standardized=TRUE)


lavInspect(fit2,"cor.lv")


# Obtain Omega
reliability(fit2)
```

3 factor individual level

```
library(lavaan)
library(semTools)


#Three factor model
threefactor <- '
TaskAttract=~Coh1_TA  + Coh2_TA  +  Coh3_TA
TaskCommit=~Coh7_TC  + Coh8_TCr  +    Coh9_TCr
Interpersonal=~Coh4_IC  + Coh5_IC  + Coh6_IC'


#fiml - full information liklihood; mlr = ""MLR" for maximum likelihood esti
mation with robust 'Huber-White' standard errors and a scaled test statistic
which is asymptotically equivalent to the Yuan-Bentler T2-star test statisti
c
fit3 <- sem(threefactor, data = Data_T2_Coh2, missing = "fiml", estimator =
"mlr")


summary(fit3, fit.measures=TRUE,rsquare=TRUE,standardized=TRUE)


lavInspect(fit3,"cor.lv")


# Obtain Omega.
```

```
reliability(fit3)
```

Step 2: Estimate between-group level variation

Lavaan's code has the same results as MPlus in the multilevel factor analysis in Step 5.

```
library(multilevel)

mult.icc(Data_T2_Coh[, c("Coh1_TA","Coh2_TA", "Coh3_TA","Coh4_IC","Coh5_IC",
"Coh6_IC","Coh7_TC","Coh8_TCr", "Coh9_TCr")], Data_T2_Coh$TeamIDUnique)
```

Step 3: Within-group factor structure

St = Individual-level covariance matrix Sw = Within-covariance matrix Sb = Between-covariance matrix #### Calculate means of items within group Single-level CFA model is tested, this time using the covariance matrix (SPW) based on individual-level scores, adjusted for their respective group means

```
library(dplyr)

#Create group means of Items

Data_T2_Coh_CFA <- Data_T2_Coh %>%

  group_by(TeamIDUnique) %>%

  summarise(Coh1_TA_TM = mean(Coh1_TA, na.rm = TRUE)) %>%

  ungroup()


Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh1_TA_TM
")],

                 by=c("TeamIDUnique"),all.x =TRUE)


Data_T2_Coh_CFA <- Data_T2_Coh %>%

  group_by(TeamIDUnique) %>%

  summarise(Coh2_TA_TM = mean(Coh2_TA, na.rm = TRUE)) %>%

  ungroup()


Data_T2_Coh<-merge(Data_T2_Coh, Data_T2_Coh_CFA[,c("TeamIDUnique","Coh2_TA_T
M")],

                 by=c("TeamIDUnique"),all.x =TRUE)



Data_T2_Coh_CFA <- Data_T2_Coh %>%

  group_by(TeamIDUnique) %>%
```

```
  summarise(Coh3_TA_TM = mean(Coh3_TA, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh, Data_T2_Coh_CFA[,c("TeamIDUnique","Coh3_TA_T
M")],
                 by=c("TeamIDUnique"),all.x =TRUE)

Data_T2_Coh_CFA <- Data_T2_Coh %>%
  group_by(TeamIDUnique) %>%
  summarise(Coh4_IC_TM = mean(Coh4_IC, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh4_IC_TM
")],
                 by=c("TeamIDUnique"),all.x =TRUE)

 Data_T2_Coh_CFA <- Data_T2_Coh %>%
  group_by(TeamIDUnique) %>%
  summarise(Coh5_IC_TM = mean(Coh5_IC, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh5_IC_TM
")],
                 by=c("TeamIDUnique"),all.x =TRUE)

 Data_T2_Coh_CFA <- Data_T2_Coh %>%
  group_by(TeamIDUnique) %>%
  summarise(Coh6_IC_TM = mean(Coh6_IC, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh6_IC_TM
")],
                 by=c("TeamIDUnique"),all.x =TRUE)

 Data_T2_Coh_CFA <- Data_T2_Coh %>%
  group_by(TeamIDUnique) %>%
```

```
  summarise(Coh7_TC_TM = mean(Coh7_TC, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh7_TC_TM
")],
                 by=c("TeamIDUnique"),all.x =TRUE)

 Data_T2_Coh_CFA <- Data_T2_Coh %>%
  group_by(TeamIDUnique) %>%
  summarise(Coh8_TCr_TM = mean(Coh8_TCr, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh8_TCr_T
M")],
                 by=c("TeamIDUnique"),all.x =TRUE)

Data_T2_Coh_CFA <- Data_T2_Coh %>%
  group_by(TeamIDUnique) %>%
  summarise(Coh9_TCr_TM = mean(Coh9_TCr, na.rm = TRUE)) %>%
  ungroup()

Data_T2_Coh<-merge(Data_T2_Coh,Data_T2_Coh_CFA[,c("TeamIDUnique","Coh9_TCr_T
M")],
                 by=c("TeamIDUnique"),all.x =TRUE)

#Adjust ind scores by mean of team
Data_T2_Coh$Coh1_TA_FA <-  Data_T2_Coh$Coh1_TA - Data_T2_Coh$Coh1_TA_TM
Data_T2_Coh$Coh2_TA_FA <-  Data_T2_Coh$Coh2_TA - Data_T2_Coh$Coh2_TA_TM
Data_T2_Coh$Coh3_TA_FA <-  Data_T2_Coh$Coh3_TA - Data_T2_Coh$Coh3_TA_TM
Data_T2_Coh$Coh4_IC_FA <-  Data_T2_Coh$Coh4_IC - Data_T2_Coh$Coh4_IC_TM
Data_T2_Coh$Coh5_IC_FA <-  Data_T2_Coh$Coh5_IC - Data_T2_Coh$Coh5_IC_TM
Data_T2_Coh$Coh6_IC_FA <-  Data_T2_Coh$Coh6_IC - Data_T2_Coh$Coh6_IC_TM
Data_T2_Coh$Coh7_TC_FA <-  Data_T2_Coh$Coh7_TC - Data_T2_Coh$Coh7_TC_TM
Data_T2_Coh$Coh8_TCr_FA <-  Data_T2_Coh$Coh8_TCr - Data_T2_Coh$Coh8_TCr_TM
Data_T2_Coh$Coh9_TCr_FA <-  Data_T2_Coh$Coh9_TCr - Data_T2_Coh$Coh9_TCr_TM
```

```
#remove dataset

rm(Data_T2_Coh_CFA)
```

Covariance Matrices

A variance–covariance matrix is then created and its values corrected to reflect division by the appropriate denominator. (The typical devisor for the covariance matrix is N-1, but for the current purposes it should be N-G, where G=the number of groups. Thus, each element of the matrix needs to be transformed by multiplying by N-1 and then dividing by N-G.)

```
#Subset for Ind CFA

Coh_Ind_CFA <- subset(Data_T2_Coh, select = c("Coh1_TA","Coh2_TA", "Coh3_TA"
,"Coh4_IC","Coh5_IC","Coh6_IC","Coh7_TC","Coh8_TCr",   "Coh9_TCr"))


#Subset for within group analysis

Coh_WI_CFA<-Data_T2_Coh [ , c( "Coh1_TA_FA","Coh2_TA_FA", "Coh3_TA_FA","Coh4
_IC_FA","Coh5_IC_FA", "Coh6_IC_FA","Coh7_TC_FA","Coh8_TCr_FA",   "Coh9_TCr_F
A")]


#Subset for between group analysis

Coh_BW_CFA<-Data_T2_Coh [ , c("Coh1_TA_TM","Coh2_TA_TM", "Coh3_TA_TM","Coh4_
IC_TM","Coh5_IC_TM", "Coh6_IC_TM","Coh7_TC_TM","Coh8_TCr_TM",   "Coh9_TCr_TM
")]


#Count number of observations for GrpSize to know how many groups there are

##Subset data

Grp<-Data_T2_Coh [ , c( "TeamIDUnique", "Coh1_TA_FA","Coh2_TA_FA", "Coh3_TA_
FA","Coh4_IC_FA","Coh5_IC_FA", "Coh6_IC_FA","Coh7_TC_FA","Coh8_TCr_FA",   "C
oh9_TCr_FA")]

#Remove duplicate observations so only on observation per group

GrpSize<-Grp[!duplicated(Grp[,c('TeamIDUnique')]),]


#matrices

##Creates covariance matrix for individual-level

St_Ind <- cov(Coh_Ind_CFA)

##Creates covariance matrix for within analysis

###variance-covariance matrix is then created and its valuescorrected to ref
lect division by the appropriate denominator. (The typical devisor for the c
ovariance matrix is N-1, but for the current purposes it should be N-G, wher
```

```
e G=the number of groups. Thus,each element of the matrix needs to be transf
ormed by multiplying by N-1 and then dividing by N-G.)

  Cov <- cov(Coh_WI_CFA)

###Corrects covariance matrix

Sw_Grp<- (Cov*(34400-1))/(34400 - 8361)


##Between covariance matrix - first obtaining the variance-covariance matrix
of the group means. This matrix must also be corrected to reflect the approp
riate denominator or divisor. To do this, one should multiple the elements o
f the matrix by the default divisor (N-1) and then divide the appropriate di
visor, in this case, the between-group level, G-1 (where G=the number of gro
ups). This corrected matrix is then used to assess the between-group factor
structure

  Cov <- cov(Coh_BW_CFA)

Sb_Grp <- (Cov*(34400-1))/(8361-1)
```

Combine matrices

This is essentially telling lavaan to compare two groups by combining the within and between covariance matrices. However, these are just one group, we are treating it like 2. The combined.n syntax lets you specify the sample size for each covariance matrix.

```
#Combine covariance matrices

combined.cov <- list(within = Sw_Grp, between = Sb_Grp)

#Specify the sample size for each matrix

combined.n <- list(within = 34499 - 8460, between = 8460)
```

Specify Within Model - 1 factor

```
library(lavaan)

library(semTools)


Cohesion.model <- '

#latent variables

Cohesion=~ Coh1_TA_FA + Coh2_TA_FA + Coh3_TA_FA  +Coh4_IC_FA + Coh5_IC_FA  +
Coh6_IC_FA  +Coh7_TC_FA + Coh8_TCr_FA + Coh9_TCr_FA'


#regressions
```

```
#The covariance structure created to capture within group variance from devi
ation scores is used here
fitw1 <- sem(Cohesion.model, sample.cov = Sw_Grp,
           sample.nobs = 34400)


summary(fitw1, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)


# Obtain Omega
reliability(fitw1)
```

Specify Within Model - 2 factor

```
Cohesion.model <- '
#latent variables
Task=~ Coh1_TA_FA + Coh2_TA_FA + Coh3_TA_FA +  Coh7_TC_FA + Coh8_TCr_FA + Co
h9_TCr_FA
Interpersonal=~Coh4_IC_FA + Coh5_IC_A + Coh6_IC_FA


#
#correlated residuals
Interpersonal ~ Task
 '
#The covariance structure created to capture within group variance from devi
ation scores is used here
fitw2 <- sem(Cohesion.model, sample.cov = Sw_Grp,
           sample.nobs = 34400)


summary(fitw2, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)


# Obtain Omega
reliability(fitw2)
```

Specify Within Model - 3 factor

```
Cohesion.model <- '
#latent variables
```

```
TaskAttract=~ Coh1_TA_FA + Coh2_TA_FA + Coh3_TA_FA
TaskCommit=~Coh7_TC_FA + Coh8_TCr_FA + Coh9_TCr_FA
Interpersonal=~Coh4_IC_FA + Coh5_IC_FA + Coh6_IC_FA



#regressions


#correlated residuals
TaskAttract ~ Interpersonal
Interpersonal ~ TaskCommit
TaskCommit ~ TaskAttract
 '
#The covariance structure created to capture within group variance from devi
ation scores is used here
fitW3 <- sem(Cohesion.model, sample.cov = Sw_Grp,
           sample.nobs = 34400)


summary(fitW3, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)


# Obtain Omega.
reliability(fitW3)
```

Step 4: Between-group factor structure

St = Individual-level covariance matrix Sw = Within-covariance matrix Sb = Between-covariance matrix ##### Specify Between Model - 1 factor

```
Cohesion.model2 <- '
#latent variables
Cohesion=~ Coh1_TA_TM + Coh2_TA_TM + Coh3_TA_TM + Coh4_IC_TM + Coh5_IC_TM +
Coh6_IC_TM + Coh7_TC_TM + Coh8_TCr_TM + Coh9_TCr_TM


#regressions


 '
```

```
fitB1 <- sem(Cohesion.model2, sample.cov = Sb_Grp,
           sample.nobs = 8361)
summary(fitB1, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)

lavInspect(fitB1,"cor.lv")

# Obtain Omega
reliability(fitB1)
```

2 factor – between group

```
Cohesion.model2 <- '
#latent variables
Task=~ Coh1_TA_TM + Coh2_TA_TM+  Coh3_TA_TM + Coh7_TC_TM + Coh8_TCr_TM + Coh
9_TCr_TM
Interpersonal=~Coh4_IC_TM + Coh5_IC_TM + Coh6_IC_TM

#regressions

#correlated residuals
Interpersonal ~ Task
 '
fitB2 <- sem(Cohesion.model2, sample.cov = Sb_Grp,
           sample.nobs = 8361)


summary(fitB2, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)

lavInspect(fitB2,"cor.lv")

# Obtain Omega
reliability(fitB2)
```

3 factor - between group

```
Cohesion3.model <- '
#latent variables
```

```
TaskAttract=~ Coh1_TA_TM + Coh2_TA_TM + Coh3_TA_TM
Interpersonal=~Coh4_IC_TM + Coh5_IC_TM + Coh6_IC_TM
TaskCommit=~Coh7_TC_TM + Coh8_TCr_TM + Coh9_TCr_TM

#regressions

#correlated residuals
Interpersonal ~ TaskCommit
TaskCommit ~ TaskAttract
TaskAttract ~ Interpersonal
 '
fitB3 <- sem(Cohesion3.model, sample.cov = Sb_Grp,
           sample.nobs = 8361)

summary(fitB3, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)

lavInspect(fitB3,"cor.lv")

# Obtain Omegase two formulas assume that the model-implied covariance matri
x explains item relationships perfectly. The residuals are subject to sampli
ng error. The third formula use observed covariance matrix instead of model-
implied covariance matrix to calculate the observed total variance. This for
mula is the most conservative method in calculating coefficient omega.
reliability(fitB3)
```

Step 5: MCFA

2 LEVELS, 3_1 FACTOR

Examining for a general team cohesion factor at higher level of analysis

```
library(lavaan)
library(semTools)

twolevel3_1factor <- '
level: 1
  Coh_TA_W =~ Coh1_TA  + Coh2_TA  + Coh3_TA
  Coh_IC_W =~ Coh4_IC  + Coh5_IC  + Coh6_IC
```

```
  Coh_TC_W =~ Coh7_TC  + Coh8_TCr  + Coh9_TCr
level: 2
  Coh_B =~ Coh1_TA  + Coh2_TA  + Coh3_TA + Coh4_IC  + Coh5_IC  + Coh6_IC + C
oh7_TC  + Coh8_TCr  + Coh9_TCr   '


#fiml - full information liklihood; mlr = ""MLR" for maximum likelihood esti
mation with robust 'Huber-White' standard errors and a scaled test statistic
which is asymptotically equivalent to the Yuan-Bentler T2-star test statisti
c. This is why
fit3_1 <- sem(twolevel3_1factor, data = Data_T2_Coh2, cluster = 'TeamIDUniqu
e', missing = "fiml", estimator = "mlr")


summary(fit3_1, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)


#ICC identical to Mplus, Bliese's package is not
lavInspect(fit3_1, "icc")
lavInspect(fit3_1,"cor.lv")
# Obtain Omega
reliability(fit3_1)
```

2-level 2-factor conflict

```
twolevel3_2factor <- '
level: 1
  Coh_TA_W =~ Coh1_TA  + Coh2_TA  + Coh3_TA
  Coh_IC_W =~ Coh4_IC  + Coh5_IC  + Coh6_IC
  Coh_TC_W =~ Coh7_TC  + Coh8_TCr  + Coh9_TCr
level: 2
  Coh_TO_B =~ Coh1_TA  + Coh2_TA  + Coh3_TA + Coh7_TC  + Coh8_TCr  + Coh9_TC
r
  Coh_IC_B =~  Coh4_IC  + Coh5_IC  + Coh6_IC
'
#fiml - full information liklihood; mlr = ""MLR" for maximum likelihood esti
mation with robust 'Huber-White' standard errors and a scaled test statistic
which is asymptotically equivalent to the Yuan-Bentler T2-star test statisti
c. This is why
fit3_2 <- sem(twolevel3_2factor, data = Data_T2_Coh2, cluster = 'TeamIDUniqu
e', missing = "fiml", estimator = "mlr")
```

```
#SRMR_between does not use the identical formula as MPlus. See documentation
in MPlus for equation differences.

summary(fit3_2, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)


#ICC identical to Mplus, Bliese's package is not

lavInspect(fit3_2, "icc")


lavInspect(fit3_2,"cor.lv")


# Obtain Omegase two formulas assume that the model-implied covariance matri
x explains item relationships perfectly. The residuals are subject to sampli
ng error. The third formula use observed covariance matrix instead of model-
implied covariance matrix to calculate the observed total variance. This for
mula is the most conservative method in calculating coefficient omega.

reliability(fit3_2)
```

2-level 3-factor conflict

```
twolevel3factor <- '

level: 1

  Coh_TA_W =~ Coh1_TA  + Coh2_TA  + Coh3_TA

  Coh_IC_W =~ Coh4_IC  + Coh5_IC  + Coh6_IC

  Coh_TC_W =~ Coh7_TC  + Coh8_TCr  + Coh9_TCr

level: 2

  Coh_TA_B =~ Coh1_TA  + Coh2_TA  + Coh3_TA

  Coh_IC_B =~ Coh4_IC  + Coh5_IC  + Coh6_IC

  Coh_TC_B =~ Coh7_TC  + Coh8_TCr  + Coh9_TCr

'

#fiml - full information liklihood; mlr = ""MLR" for maximum likelihood esti
mation with robust ‘Huber-White’ standard errors and a scaled test statistic
which is asymptotically equivalent to the Yuan-Bentler T2-star test statisti
c. This is why

fit3_3 <- sem(twolevel3factor, data = Data_T2_Coh2, cluster = 'TeamIDUnique'
, missing = "fiml", estimator = "mlr")


#ICC identical to Mplus

lavInspect(fit3_3, "icc")
```

```
lavInspect(fit3_3,"cor.lv")

summary(fit3_3, fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)

# Obtain Omega

reliability(fit3_3)
```

MCFA Step 5 in MPlus

```
TITLE: MCFA Cohesion;

DATA: FILE IS MPlus_Cohesion.csv;
```

```
VARIABLE: NAMES ARE
        Coh1 Coh2 Coh3 Coh4 Coh5 Coh6 Coh7 Coh8 Coh9 TmNm;
    USEVARIABLES ARE Coh1-Coh9;
    CLUSTER = TmNm;

ANALYSIS:
    TYPE = TWOLEVEL;
   !Note: with missing data estimator=mlr is used to obtain robust estimates
(Yuan & !Bentler, 2000), if non-robust estimates are desired use estimator=m
l;
 ! Missing data estimation is now the default in Version 5 and higher;
    ESTIMATOR IS MLR;
    H1iterations = 10000; ! This allows the model to
    ! converge as the default is 1000 iterations
MODEL:
      %WITHIN%
      CohTAw by Coh1 Coh2 Coh3;
      CohICw by Coh4 Coh5 Coh6;
      CohTCw by Coh7 Coh8 Coh9;

      %BETWEEN%
      Cohb by Coh1 Coh2 Coh3 Coh4 Coh5 Coh6 Coh7 Coh8 Coh9;

OUTPUT: STDYX;  !YX is for continuous variables
        sampstat; !will display sample means, variances,
        !covariances and correlations for continuous variables
        !MPlus uses full info max liklihood
```