6-14-2021 10:30 AM

Finding Teams that Fight Fair: Exploring Trajectories of Team Conflict Over Time

Natasha E. Ouslis, *The University of Western Ontario*

Supervisor: Allen, Natalie J., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology

Abstract

Disagreements are a reality for teams. Yet how and when teams experience conflict may impact their chances of success. We know relatively little about how team conflict emerges over time, especially for project-based teams. Disagreements over *personal* topics, logistics, and contributions have been consistently damaging to team performance (De Dreu & Weingart, 2003; O'Neill, Allen, & Hastings, 2013). The implications of *task-based conflict* over time, however, are inconsistent and poorly understood. To resolve these questions, I conducted three studies examining how conflict developed over the lifetimes of 272 engineering design project teams. Study 1 explored the measurement and patterns of dynamic team conflict. Conflict can be consistently measured over time; I found two classes of teams following different conflict trajectories. In Study 2, I examined whether personality and demographic traits influence team conflict over time and explored how conflict affects performance. Members' demographic characteristics and personality traits related to their individual conflict perceptions. Accelerating relationship conflict predicted poorer team-rated performance, whereas extraversion and conscientiousness predicted better team-rated performance. In Study 3, I used faultlines to predict conflict paths and team performance. Teams with demographic faultlines saw relationship conflict increase more quickly over time. This in turn predicted lower performance. Personality faultlines had no relation to conflict or performance. Taken together, this set of studies uses new team input methods and finds that clusters of teams explain the conflict-success connection. These results help us understand conflict as it happens: from the moment teams work together to when they complete their projects.

Keywords

Team conflict, longitudinal, project teams, engineering design, innovation, performance, demographics, personality, faultlines.

Summary for Lay Audience

Most, if not all, teams disagree. Yet some kinds of team conflict may affect teams' performance differently. We know relatively little about how team conflict changes over time, especially for project-based teams. Conflict over personal topics, logistics, and team members contributions are consistently harmful for team performance (De Dreu & Weingart, 2003; O'Neill, Allen, & Hastings, 2013). It is not well understood if, and when, task-based conflict is helpful. To resolve these questions, I conducted three studies examining how conflict developed over the lifetimes of 272 engineering design project teams. In Study 1, I showed that conflict can be consistently measured over time. I found two classes of teams that have different conflict patterns. In Study 2, I found that team members' personality and demographic traits influence their ratings of team conflict. The more that teams' relationship conflict increased over time, the poorer their performance was. However, teams with higher average extraversion and conscientiousness had better team-rated performance. In Study 3, I found that teams with stronger rifts between members on demographic traits saw relationship conflict increase more quickly over time; this relationship conflict predicted poorer performance. This set of studies compares many team and member inputs, and clusters of teams, to explain the conflict-success connection. These results help us understand conflict as it happens: from the moment teams work together to when they complete their projects.

Acknowledgements

It takes a village to raise a scholar. I appreciate all the supporters who have helped and challenged me over my many years of education. First, I thank my supervisor, Dr. Natalie Allen. Thank you for your support, understanding, flexibility, and questions. To all my examiners and committee members – Professors Adair, Olivera, Robertson, Benson, and Tremblay – thank you. Your feedback is crucial for us to learn and develop.

To all my collaborators and past supervisors, I am eternally grateful for the opportunity to work with you. Thank you for giving me the chance to expand my views and to present our work. I am incredibly lucky to have learned from such amazing lab mates, professors, graduate students, and undergraduate students across both alma maters. Thank you as well to the administrative professionals who helped me navigate these schools and complete my degrees; if I had to keep track of everything myself, this would either be extremely late or never finished!

To my friends and family, thank you for bringing balance into my life. To my friends and colleagues, scattered across the world: thank you for making time for me and for always being open to collaboration. To my family – Zoran, Sofia, Sebastian, Chris, Frances, Louie, Lana, Matthew, Lucas, Elizabeth, and Marcia Sr. – thank you for teaching me leadership, self-confidence, and community. To Fatima, thank you for welcoming me into your life so quickly and so openly. To Maya and Marmalade, thank you for always comforting me – without words.

To Brian, thank you for everything. We have grown and changed so much over more than seven years together. Yet your love is constant. You challenge me, you believe in me, and you set a great example through your actions. I'm so excited for the next steps in the personal and professional lives we've built. I love you.

Table of Contents

List of Tables

List of Figures

List of Appendices

INTRODUCTION

Teams are ubiquitous in organizations. Across knowledge work, professional services, and manufacturing, more employees are working in teams than have before (Bikfalvi, Jäger, & Lay, 2014; Salas, Cooke, & Rosen, 2008). Many, if not all, of these teams experience conflict. This is not surprising, as team members debate ideas, may take disagreements personally, and sometimes put forth less effort than other members do (Jehn, 1995). Though team conflict can influence decision making, creativity, performance, and satisfaction (e.g., De Dreu & Weingart, 2003), researchers know less about the *dynamics* of team conflict over time. The overall goal of this project is to explore team conflict trajectories, understand how they differ within and across teams, and identify the antecedents and outcomes of these dynamic conflict paths.

Teams can develop better ideas or become more efficient through process gains (Huffmeier & Hertel, 2011). Process gains happen when team members perform better together than they could alone. When team member interaction speeds up group tasks and increases the quality of teams' output, a group has benefitted from process gains. Researchers found that in some cases, group members have higher motivation by working together, which improves their performance (Weber & Hertel, 2007). Unfortunately, teams can perform worse than the sum of their parts (i.e., their individual members) through process losses. Process losses happen when groups perform below their potential due to the speed and/or quality trade-offs associated with team member interaction (e.g., Miner, 1984). Not surprisingly, perhaps, groups can experience significant process losses from team conflict (Jehn, 1995). Team researchers find there are at least three forms, or types, of conflict: task conflict, in which team members disagree on task-related matters, relationship conflict, in which team members experience personality clashes and personal attacks, and process conflict, in which members disagree on how to complete their tasks. These three conflict types are each related to lower individual satisfaction, intention to stay in the group, and group member liking (De Dreu & Weingart, 2003; De Wit, Greer, & Jehn, 2012). Yet task conflict can also *benefit* teams in non-routine tasks (Jehn, 1995), decision-making contexts (O'Neill, Allen, & Hastings, 2013), and if the team has a high level of psychological safety (Bradley, Postlethwaite, Klotz, Hamdani, & Brown, 2012). These results indicate that task conflict can be beneficial, in certain situations. Unlike task conflict, relationship and process conflict show consistent negative relations with team performance across multiple meta-analyses (De Dreu & Weingart, 2003; De Wit et al., 2012; O'Neill et al., 2013).

These mixed relations between task conflict and team outcomes are puzzling. They create uncertainty for researchers and practitioners about the role of task conflict in the team lifecycle. Inconsistent results, such as those above, make theory development and practical application difficult. Fortunately, there are many possible reasons for these mixed findings. Moderators such as task type, seniority level in the organization, teamwork setting, and measurement methods may account for these inconsistent results. However, these moderators are not consistently supported across research reviews; as may be expected, some variability between studies also remains unexplained (De Dreu & Weingart, 2003; De Wit et al., 2012; O'Neill et al., 2013). In one meta-analysis, researchers tested the interaction between task and relationship conflict. The presence of other conflict (i.e., of the relationship variety) may change the relation between task conflict and important outcomes in teams. Thus, future research could examine the co-occurrence of conflict types to understand why task conflict has such mixed results.

If we only use meta-analyses to find and test these explanations, however, we will have a limited set of tools for resolving inconsistencies in the task conflict literature. Some potential moderators of the conflict-performance correlation may not be reported, or they may be confounded with other variables. Other challenges with the team level meta-analytic approach exist. For example, another limitation is that systematic reviews test all variables at the team level; this is appropriate but incomplete. Researchers usually refer to conflict at the team level, yet analyses at this level treat members as interchangeable. Overall team conflict scores may not reflect conflict between subgroups in the team. Subgroups, comprised of at least two members, may be how team conflict actually starts (Humphrey & Aime, 2014). However, team and other multi-level researchers use measures of agreement to justify team-level analyses (Woehr, Loignon, Schmidt, Loughry, & Ohland, 2015). Focusing on high agreement within the team and low member-to-member variability may be masking the true processes happening within a group. This means low agreement scores and high variation within a team may reflect accurate, but varying, perceptions across team members.

Team members can experience conflict one-on-one at first, as one team member disagrees with another (Amason & Schweiger, 1994). Whereas conflict can start as a dyadic disagreement, it can spread to other team members through observation and by team members taking sides in the conflict (Jehn, Rispens, Jonsen, & Greer, 2013; Peterson & Behfar, 2003). A single-level research approach (Hammond, McClelland, & Mumpower, 1980), in which the team mean represents members' perceptions, assumes the mean reflects the true feeling each team member has. The process of averaging individual perceptions into one team value treats within-team differences as error (Kristjansson, Kircher, & Webb, 2007). By averaging individual team members' responses, however, researchers may lose valuable and substantive variability from each team member's perspective. By considering individual differences in team constructs, researchers can analyze individual differences within teams and variability between teams (Chan, 1998a).

The static nature of most team process research is another limitation to fully understanding team conflict relations. Team variables, especially process-oriented constructs, may develop and impact team outcomes over time. Yet team research has often collected process variables only once and at the same time as other variables (i.e., cross-sectionally; Humphrey & Aime, 2014). Thus, the inputs for any meta-analysis reflect one point in time for most studies; these moments may differ across studies. This makes direct comparisons difficult and less accurate. The scattered design and timing of team conflict measures in meta-analytic reviews arguably results in less consistency within the team lifecycle and in pressures that the team faces externally (e.g., project deadlines). For example, De Wit and colleagues (2012) theorized that top management teams had positive task conflict-performance relations, because they may avoid relationship conflict when task-based disagreements happen. This assumes a dynamic process whereby task conflict may emerge earlier in the team's lifecycle. This early task conflict may predict higher levels of relationship conflict in some team types, yet not in others. As many researchers have recommended (e.g., Humphrey & Aime, 2014; Mathieu, Hollenbeck, van Knippenberg, & Ilgen, 2017), collecting measures of team conflict over time can help researchers understand when each conflict type emerges and how conflict impacts team outcomes.

Recent Methodological Developments to Improve Conflict Research

Relations Between Conflict Types

In their 2003 meta-analysis, De Dreu and Weingart found the task conflict-performance relation was less negative when task and relationship conflict were weakly correlated, compared to when they were strongly correlated. Specifically, in some of the 14 studies that they examined, teams had high task and relationship conflict and this situation was related to poorer team performance. Other teams had low task and relationship conflict, resulting in better performance for those teams. In other studies, task and relationship conflict did not show substantial overlap; task conflict had little relation to performance there. The average correlation between relationship and task conflict in this meta-analysis was strong at .54 (De Dreu & Weingart, 2003). This value is not strong enough to suggest task and relationship conflict are the same construct (Carlson & Herdman, 2012). However, the results were consistent nearly a decade later. Researchers replicated this strong effect in a more recent meta-analysis with 116 studies (De Wit et al., 2012). This confirms the finding that task and relationship conflict co-occur in some cases, yet not in others.

One study using latent profile analysis demonstrated that indeed, commensurate levels of task and relationship conflict occur in some teams, whereas other teams show differences between conflict levels for these two types (O'Neill, McLarnon, Hoffart, Woodley, & Allen, 2018). This design allowed researchers to test in one study what was previously done in a meta-analysis. Across three samples, the authors found a four-profile solution in which relationship and process conflict co-occurred across all profiles. In some profiles, task conflict scores were much higher than process and relationship conflict scores. In other profiles, all three conflict levels were high. Team profiles with high task and low process/relationship conflict showed the highest team performance, potency, and conflict management. This finding suggests that task disagreements may help performance when the team has low process and relationship conflict, but not when personal conflict is high. By analyzing all team conflict types at once, researchers can uncover new insights that explain the inconsistent patterns and contradictory results in the task conflict literature.

Multilevel Investigations

Team members can differ from one another on many characteristics. However, traditional team process research gathers data at the individual level to average at the team level. Using this model, researchers expect high agreement as it reflects a shared construct among team members (Cole, Bedeian, Hirschfeld, & Vogel, 2011). By considering how individual team members differ in their perceptions of team processes, researchers can deepen our understanding of conflict emergence, development, and resolution. Researchers take a configural approach in much team input research, yet this approach is featured less often in team process research. For example, two team composition meta-analyses found that input variables calculated by the variability, minimum, and maximum of the team members' characteristics showed practical incremental validity above the mean in predicting team performance (Barrick, Stewart, Neubert, & Mount, 1998; Bell, 2007). Variance in some members' personality traits also interacts with relationship conflict to predict member satisfaction and desire to remain with the team (Tekleab & Quigley, 2014). Thus, individual variables can predict team processes and outcomes better with configural approaches, as these new methods consider individual differences in team members, than with an approach based exclusively on averaging member inputs.

Researchers can use configural approaches to build on existing team research by considering multiple characteristics of team members at once. Faultlines are rifts within a team from member attributes that create subgroups (Meyer & Glenz, 2013). These rifts are negatively related to team performance, satisfaction, cohesion, task conflict, and relationship conflict (Thatcher & Patel, 2011). Perceived faultlines based on member evaluations are also related to conflict between subgroups; conflict is worse when perceptions reflect objective demographic differences (Greer & Jehn, 2007). Rifts on demographics, abilities, and personality show unique patterns with team functioning, cohesion, and conflict (Molleman, 2005). However, other studies find more complex relations between these rifts and conflict. For example, task, relationship, and process conflict in one study were related to lower team performance in groups without faultlines and higher performance in groups with gender or culture rifts (Hideg & Adair, 2010). These analytic approaches help researchers and practitioners understand how conflict emerges and how it relates to each team member's unique set of characteristics.

Finally, new approaches for aggregating task conflict itself help to unite traditional averaging approaches with newer methods that take variability or dispersion into account. One study measured positive skewness in task conflict within teams to better reflect differences in members' conflict perceptions (Sinha, Janardhanan, Greer, Conlon, & Edwards, 2016). In this study, positively skewed task disagreements were related to higher task performance. This relation was mediated by reflective communication. Positive skewness in this conflict type means a small number of members perceived high task conflict and most members perceived low conflict. For example, two people who voice their disagreement by suggesting new ideas to each other may feel there is high task conflict. To the rest of the team, this conversation may not have received much attention; they may rate task conflict as low. This skewed task conflict helped team performance because it motivated team members to communicate about the way they work together. This study connects newer team conflict methods with the potential origins of team conflict: disagreements between two members (Humphrey & Aime, 2014; Korsgaard, Jeong, Mahony, & Pitariu, 2008). If researchers took a multilevel approach, they could combine differences *within* teams and differences *between* teams to better understand conflict.

Dynamic, Longitudinal Research

By conducting both multilevel and dynamic research, researchers can capture multi- and cross-level effects that happen in teams over time (Humphrey & Aime, 2014). This integrative type of research can help theorists and practitioners better understand how team conflict operates. Longitudinal research designs, where researchers take multiple measures of the same construct, can add predictive power to analyses and help theorists create more specific psychological theories. For example, trajectories of job satisfaction explain 43% of employee turnover variance, whereas cross-sectional predictors explained only 5% (Liu, Mitchell, Lee, Holton, & Hinkin, 2012). As well, declining commitment trajectories better predicted intentions to quit and actual quitting behaviour within the next nine months than one-time measures of commitment (Bentein, Vandenberghe, Vandenberg, & Stinglhamber, 2005). Even heritability estimates benefit from longitudinal designs: electroencephalography (EEG) patterns over time, measured with the amplitude of electric signals in the brain and analyzed with latent growth modeling (LGM), better reflect heritability or genetic influences than EEG signals measured at one time point (Carlson & Iacono, 2006). These examples illustrate the value of longitudinal research in explaining results across subfields in psychology.

Several studies have investigated team conflict over time. In one longitudinal study, better-performing teams had low but increasing process conflict, low relationship conflict that increased before project deadlines, and medium task conflict at the midpoint of their group interactions (Jehn & Mannix, 2001). When investigating within-team change patterns, researchers found that teams with decreasing task conflict in the second half of their project have higher satisfaction at the end of their project (Li & Roe, 2012). As well, teams with relationship disagreements that decrease over either period in the project (i.e., Time 1 to 2 or Time 2 to 3) show higher satisfaction. This suggests, predictably, that teams that resolve their personal conflicts are happier with their team members. Teams with less process conflict after the midpoint of their time together have higher satisfaction later on. These patterns either reflect true differences in group processes over time, or they are the result of overinterpreting small differences in a limited sample of teams. The results above have not been replicated to date; therefore, these results may be sample-dependent. It is worth noting that the average trajectory in Li and Roe's (2012) analysis represented a small percentage of all teams in the sample. This means most teams had different trajectories than the average. Thus, traditional longitudinal growth methods, that fit all teams into one average trajectory, may not accurately reflect most teams' change over time based on previous research. However, analytic approaches to investigate longitudinal research are proliferating.

When modeling both task and relationship conflict over time, relationship conflict was significantly higher in teams where task conflict and perceived team performance were low (e.g., Guenter, van Emmerik, Schreurs, Kyupers, van Iterson, & Notelaers, 2016). In this study, there were no differences in the relationship conflict trajectory between high and low task conflict under high perceived performance conditions. However, relationship conflict trajectories diverged when perceived performance was low. In low perceived performance situations, low task conflict predicted decreasing relationship conflicts; high task conflict predicted accelerating relationship conflicts. In a longitudinal study of three conflict types, process conflict early in the team's time together predicted higher conflict of all types in the future (Greer, Jehn, & Mannix, 2008). Taken together, these studies suggest that team conflict types influence each other and predict performance and satisfaction over time.

Interrelations between Conflict Types

The studies above help us understand how team conflict unfolds. However, there are some limitations in the studies' analytic methods that restrict the conclusions we can draw from these data. All but one study was conducted exclusively at the team level; thus, researchers were not theorizing, or empirically testing, conflict relations at the dual emergent level (Humphrey & Aime, 2014). To my knowledge, no study has established measurement invariance across time, as recommended by Chan (1998a), to ensure comparable means that are suitable for modeling change. Completing this step would provide more assurance that conflict type scales are measuring the same underlying concepts at different stages of team tenure.

Few studies have used latent growth or multilevel modeling, which are preferred and more robust methods for analyzing longitudinal data (Chan, 1998a; Collins, Gibson, Quigley, & Parker, 2016; Misangyi, LePine, Algina, & Goeddeke Jr., 2006). Compared to multilevel modeling techniques, repeated measures analysis of variance and repeated measures regression can result in higher error rates and misspecify the proportion of between-group variance (Misangyi et al., 2006). Some studies use longitudinal modeling with manifest variables, which can potentially contaminate true conflict trajectories with measurement error variance if there is low interitem reliability for conflict scales. These issues can be resolved by representing each time point as a latent variable with questionnaire items as indicators (Chan, 1998a). This measurement invariance process and LGM approach can address the analytic challenges in the studies above.

Recent research integrates dyads within teams, team-level analyses, and longitudinal data collection to test more sophisticated relations between task and relationship conflict. Over 8 weeks, researchers found that early relationship conflict between any dyadic pairs in the team was related to lower subsequent information exchange, even after researchers controlled for task conflict (Humphrey, Aime, Cushenberry, Hill, & Fairchild, 2017). Conversely, information exchange promoted later task conflict. This is concerning for teams that experience relationship conflict early; this type of disagreement could hinder further productive discussions and reduce the information shared between group members. Other researchers find that individual, dyadic, and subgroup conflict occurs more frequently than team-level conflict in which members agree about the level of conflict they perceive (Shah, Peterson, Jones, & Ferguson, 2020). Interestingly, the same researchers found task conflict measured at the individual and dyadic levels help team performance, yet *team-level* conflict scores *negatively* predict performance. This provides further evidence that within-team variations in conflict are meaningful and do not simply reflect error.

Combining these analytic approaches to *effectively* test theory and measurement over time requires careful design, control, and high power. Reaching adequate statistical power of 80% or higher in multilevel and small group (i.e., 3-5 members) research requires large sample sizes (Chen, Bliese, & Mathieu, 2005; Mathieu, Aguinis, Culpepper, & Chen, 2012). The sample sizes typically found in published journal articles (i.e., approximately 40-60 teams) may not be enough to test these complex relations. To establish whether conflict types cause team performance, we must examine alternative explanations. Variables such as team size, meeting frequency, deadlines, and task type (De Dreu & Weingart, 2003; De Wit et al., 2012; O'Neill et al., 2013) may also explain the link between conflict and performance. Models of team change over time suggest that the team's temporal midpoint marks a shift in team processes (Gersick, 1988; 1991). Thus, there is considerable value in studying project-based teams with clear work stages to model team process changes over time. Ideally, researchers should investigate this by gathering longitudinal data from these teams, while ensuring a consistent size and frequency of teamwork, as teams complete similar tasks under the same deadline structure.

Study 1 Hypotheses

The purpose of Study 1 was to establish the properties of team conflict over time. However, trajectories for relationship and task conflict may be different within and/or across teams. As relationship and process conflict show similar patterns with outcome variables (De Wit et al., 2012; O'Neill et al., 2013) and move in tandem within published profile analyses (O'Neill et al., 2018), I expect relationship and process conflict trajectories to look similar. However, task conflict is distinct from other conflict types due to its inconsistent relation to team outcomes (O'Neill et al., 2013), its differential pattern in profile analyses (O'Neill et al., 2018), and its role as a trigger of other forms of conflict (Amason & Schweiger, 1994; Peterson & Behfar, 2003). Therefore, I expect task conflict will be less correlated to relationship and process conflict than relationship and process conflict will be correlated with each other.

H1: Process and relationship conflict will be more strongly intercorrelated than these two conflict types and task conflict.

To model conflict over time and test predictions related to team performance, I must first establish measurement invariance across time. Team conflict measures, of the sort used in this set of studies, have been used for nearly two decades (e.g., De Wit et al., 2012; O'Neill et al., 2013). I expect these team conflict measures will show measurement invariance.

H2a: Conflict measures will display strong measurement invariance across time points at the individual level.

H2b: Conflict measures will display strong measurement invariance across time points at the team level.

Next, I will characterize individual and team conflict trajectories over time. One theory of team change over time takes its name from evolution (Gould & Eldredge, 1986). This theory -- punctuated equilibrium theory (Gersick, 1991) -- has received empirical support in studies examining key team constructs including conflict (Okhuysen & Waller, 2002) and cohesion (Michinov & Michinov, 2007). It is a model for predicting the shape of team trajectories, one of the major goals of this set of studies. This theory posits that teams typically exist in a state of equilibrium that is punctuated by events that disrupt the team's normal functioning (Gersick, 1989). The midpoint of a project team is a common temporal milestone; at this time, the team's tasks or processes may need to shift. For example, teams may stop generating ideas and begin to implement them (Humphrey & Aime, 2014). The midpoint is easier to keep track of in some conditions than in others: teams who started short tasks on an easy-to-remember time (e.g., 3:00pm or 12:30am), for example, could perceive the midpoint of their tasks more easily than teams starting on atypical times (e.g., at 5:47pm; Labianca, Moon, & Watt, 2005). This midpoint change may be reflected in significant slopes of conflict over teams' time together. This means team conflict may change over time, instead of staying at a consistent level throughout.

Some studies find that task-focused strategies were helpful to teams at their midpoint (Woolley, 1998). Other studies show the midpoint of a team was more disruptive when teams were told to focus on time management, information sharing, or elaboration (Okhuysen & Waller, 2002). Thus, I expect teams will show changes in team conflict levels before and after the temporal midpoint. In this proposed sample, project teams have a defined shift in their work when their first small project has been completed and they receive external feedback on their performance. This happens near the temporal midpoint and is reflected in the timing of data collected for this project. Thus, I expect that time will explain the variability in team conflict scores.

H3: Time will explain variance in team conflict scores.

It is possible, however, that not all teams will follow this same pattern. I expect some teams to have different trajectories than other teams across team conflict types due to unique team interactions. In addition, team members do not always perceive conflict uniformly (e.g., Sinha et al., 2016). Therefore, I expect team members' conflict trajectories to differ as in the study of intrateam longitudinal conflict conducted by Li and Roe (2012).

H4: Multiple classes of team member conflict trajectories will fit the individual conflict data better than a one-class solution.

H5: Multiple classes of team conflict trajectories will fit the team conflict data better than a one-class solution.

Study 2 Hypotheses

Building on meta-analytic findings (De Dreu & Weingart, 2003; De Wit et al., 2012; O'Neill et al., 2013), I expect that conflict trajectories will predict team performance. Task conflict can become problematic if it escalates into relationship and/or process conflict (e.g., Jehn, 1997; Wang, Jing, & Klossek, 2007). I expect that positive slopes or accelerating trajectories of relationship conflict will negatively relate to performance. However, task conflict can be helpful in decision-making groups (O'Neill et al., 2013). Teams in this study have many decision-making tasks: they may benefit from the idea generation and exploration that comes from task conflict among team members. Thus, I expect that a higher intercept for task conflict will relate positively to team performance. Yet as teams work together past the midpoint of their projects, they may focus on production and efficiency over idea generation and divergent thinking. This suggests increases in task conflict may be detrimental after the team's midpoint. Therefore, I expect that changes in the level of task conflict over time will not predict better team outcomes.

There are many dimensions of team performance scores. Meta-analyses of conflict and team performance relations (e.g., De Dreu & Weingart, 2003; De Wit et al., 2012) find substantial heterogeneity across studies, suggesting that measurement or contextual factors may explain when conflict helps or hinders performance. The relationship between task conflict and performance was more positive for top management team performance than for teams lower in the organizational hierarchy (De Wit et al., 2012). Task conflict was also more positively related to financial and decision-making performance compared to overall team performance and more negatively related to field team performance rather than performance measured in a classroom or laboratory setting.

Team research uses multiple sources of performance, including team-rated, expert-rated, and objective performance measures. In a recent meta-analysis, process conflict was more negatively related to team-rated performance than to supervisor/expert ratings and objective ratings (O'Neill et al., 2013). The same research found that relationship conflict was more negatively related to team-rated performance than to expert and objective performance ratings. These differences may reflect common method variance between team ratings of conflict and performance. However, common method variance may not fully explain these differing effects. Other workplace research finds that self-evaluations are more closely related to some psychological states such as emotional intelligence than others' evaluations (Joseph, Jin, Newman, & O'Boyle, 2015) and that workplace experiences such as burnout have unique mediating pathways that predict self-rated performance but not other-rated job performance (Parker & Kulik, 1995). Thus, conflict types may relate differently to team performance when measured through team members' ratings than when measured through outside evaluators' ratings. For this reason, conflict scores may be more strongly related to member-rated team performance than to other-rated team performance.

H6: Relationship and task conflict scores will predict team performance.

H7: Relationship conflict slopes, but not task conflict slopes, will predict team performance.

The hypotheses presented above describe team conflict's intercepts and trajectories as they relate to team outcomes. To understand how these conflict states emerge, I will explore how team conflict relates to teams' personality and demographic composition variables. Many researchers have studied the impact of team demographics on conflict and performance. To date, at least four meta-analyses have tested the connection between team member demographic characteristics and team performance (Bell, Villado, Lukasik, Belau, & Briggs, 2011; Horwitz & Horwitz, 2007; Van Dijk, Van Engen, & Van Knippenberg, 2012; Wei, Liu, & Chen, 2015). Contrary to popular belief (Eagly, 2016), many types of demographic group diversity have negative (e.g., gender and race) or null (e.g., age) relations with team performance. Thus, demographic characteristics may relate to team performance directly. Meta-analyses have investigated whether a link between demographic diversity and conflict exists (e.g., De Wit & Greer, 2008). Whereas the connection between task conflict and demographic diversity is weak or nonexistent, demographic traits may relate to personal disagreements such as relationship conflict. Mohammed and Angell (2004) found that team gender diversity was related to higher relationship conflict. In addition to the hypotheses below, exploratory analyses will examine the relation between individual-level demographic characteristics and outcomes (i.e., team conflict and performance).

H8: Demographic diversity (e.g., gender and ethnicity composition) on the team will negatively predict relationship conflict at the team level.

H9: Demographic diversity (e.g., gender and ethnicity composition) on the team will negatively predict performance at the team level.

Existing theory and empirical research support two roles of personality composition: 1) using team personality as a moderator that can change how group processes affect performance, and 2) using personality as an input factor that affects the level of team conflict (Driskell, Hogan, & Salas, 1987). For example, a team's average score on a particular personality trait, such as openness to experience or emotional stability, can interact with task conflict to predict higher team performance (Bradley et al., 2012). In addition, team members higher in openness to experience reported lower relationship conflict, but showed no differences in task conflict (Gallo, 2017). The opposite relation was found in a study of dyads; individuals with higher openness reported more relationship conflict in their dyads than individuals with lower openness (Bono et al., 2002). Members of teams with higher emotionality levels perceived higher relationship conflict than members of groups with low average emotionality (Bolger &

Zuckerman, 1995). People higher in emotionality reported more relationship conflict (Bono, Boles, Judge, & Lauver, 2002; Neuman, Wagner, & Christiansen, 1999). Thus, I expect that team emotionality will relate to higher relationship conflict intercepts and slopes in teams, and personality may directly predict team performance. Higher team openness may relate to higher task conflict, as team members may share their divergent views and be more willing to engage with others' ideas (Aeron & Pathak, 2017). Some research underscores the importance of team openness norms for successfully using conflict to improve performance (e.g., Amason & Mooney, 1999; De Dreu & Weingart, 2003; Esquivel & Kleiner, 1996). Thus, I expect team-level openness will be related to team performance and task conflict.

Team agreeableness levels are strong predictors of team performance compared to other personality factors in single studies (e.g., Bradley, Baur, Banford, & Postlethwaite, 2013). In meta-analytic research (Peeters, Van Tuijl, Rutte, & Reymen, 2006), agreeableness and conscientiousness were the only of five personality traits to predict team performance via their elevation (i.e., their score) and their variability. There has been little research on how individual agreeableness and conscientiousness relate to perceptions of team conflict. Individuals' agreeableness moderated the impact of conflict between individuals, such that workers higher in agreeableness might be more negatively affected by conflict than workers lower in agreeableness (Ilies, Johnson, Judge, & Keeney, 2011). Yet individuals who are more extraverted, conscientious, and agreeable perceived lower relationship conflict in dyads (Bono et al., 2002; Neuman et al., 1999). Some research finds a significant relation between agreeableness and dyadic conflict (Asendorpf & Wilpers, 1998; Bono et al., 2002; Graziano, Jensen-Campbell, & Hair, 1996). Individuals may be more affected by conflict over time, though they might perceive lower conflict in the moment. Of course, much existing research on personality and conflict is not conducted in teams. This limits our ability to use background research on personality and conflict in teams to make specific predictions for these studies. Team member agreeableness is related to lower relationship and task conflict (Gallo, 2017). Recognizing this, I expect individuals with higher agreeableness will perceive lower relationship conflict in their teams.

H10a: Individual personality traits will predict team member ratings of conflict.

H10b: Team aggregated personality traits will predict team conflict scores.

H11a: Individual personality traits will predict performance at the individual member level.

H11b: Team aggregated personality traits will predict team performance.

Study 3 Hypotheses

Study 2 investigated individual team member characteristics and their relations with team conflict and performance. Yet when researchers consider team members' multiple traits or identities in conjunction, they find different results. One way to measure how teams differ on multiple team member traits are through faultlines that calculate rifts between team members based on their attributes (Lau & Murnighan, 1998). These rifts reflect subgroups in teams according to members' demographics, personality, or access to information. Rifts of this type can affect teams through social categorization processes (Turner, Brown, & Tajfel, 1979). Whether through visible differences that team members can perceive immediately or through group membership that is revealed to team members over time, members categorize themselves and others into subgroups with similar traits.

This can promote favouritism and support for in-group members while creating distance between subgroups that can reduce information sharing and collaboration (Thatcher & Patel, 2011). It is through this subgroup creation process than team faultlines can hurt group processes and performance.

Team demographic faultlines can be harmful for team cohesion and conflict, especially when team autonomy is high (Molleman, 2005). Demographic and informational faultlines in top management teams can lead to poorer performance if shared objectives are low (Van Knippenberg, Dawson, West, & Homan, 2011). Demographic faultlines, in which team members perceive subgroups based on objective characteristics, increase the likelihood of coalitions and conflict in the group, lowering satisfaction and group performance (Jehn & Bezrukova, 2010). This is stronger when faultlines are "activated". An activated faultline happens when objective faultline scores are in line with subjective perceptions of team-level rifts. Therefore, I expect that demographic faultlines will be related to higher relationship and process conflict at the team level.

Demographic faultlines take time to become activated and to consequently affect team interactions. A team episode, such as a group disagreement or a stressful external situation, may not happen at the early stages of team interaction. For this reason, I expect demographic faultlines will relate to conflict later in the team lifecycle than the conflict intercept will measure. Thus, demographic faultline scores may relate to conflict slopes, but not conflict intercepts.

H12: Demographic faultlines will positively predict relationship and process conflict slopes at the team level.

H13: Demographic faultlines will negatively predict team performance at the team level.

In addition, some preliminary research suggests that personality faultlines may explain team performance beyond the contribution of each personality factor and demographic information (Byington, 2012). However, other research did not find an effect of personality faultlines on team processes (Molleman, 2005). Little research has been conducted on personality faultlines (Thatcher & Patel, 2012), yet this may be a promising area of future team composition research. However, existing meta-analytic research that analyzes each personality trait in the Big Five model separately provides support for the relation between personality trait variability and team outcomes (Bell, 2007). Other research finds that personality differences among team members is related not only to team performance, but also to team cohesion (Barrick et al., 1998).

One study on faultlines found the frequency of team communication can exacerbate the effects of cultural and personality faultline strength on team conflict (van der Kamp, Tjemkes, & Jehn, 2011). These results find the direct impact of stronger faultlines on team conflict is negative, though team behaviours could amplify or reduce this effect. In related results, researchers found that personality similarity on some traits dampens the effects of relationship conflict on the team (Tekleab & Quigley, 2014). This supports the personality faultline literature as high similarity in personality traits would reflect a weak or nonexistent personality faultline. In addition, other non-demographic characteristics such as educational level and work experience have a negative impact on team learning when there is little common ground and high faultline distance between members (Rupert, Blomme, Dragt, & Jehn, 2016). Existing research on team member traits that are not visible, yet which may influence how members interact (i.e., educational training, work-related characteristics, and personality similarity), support the proposed link between personality rifts and team variables. Thus, I expect personality faultlines will relate to team conflict and performance.

Specifically, I expect personality faultlines will affect conflict later in teams' interactions. As personality is not immediately visible to team members (i.e., it is considered a deep-level trait; Harrison, Price, Gavin, & Florey, 2002), personality faultlines may only affect conflict after extensive team member interaction. This suggests personality faultlines will only influence conflict as it develops, as measured by conflict slopes and not conflict intercepts. All hypotheses are summarized in Table 1.

H14: Personality faultlines will predict conflict slopes at the team level.

H15: Personality faultlines will predict team performance.

Table 1. *Hypotheses for all three studies.*

Hypothesis	Independent Variable	Dependent Variable
Study 1		
H1	TC, PC, and RC	None - correlation
H2a/b	Conflict	None – measurement invariance
H3	Time	Conflict
H4	Ind. Conflict Scores	None – growth mixture modeling
H5	Team Conflict Slopes	None – growth mixture modeling
Study 2		
H6	RC and TC Scores	Performance
H7	RC and TC Slopes	Performance
H8	Demographics	Conflict
H9	Demographics	Performance
H10a/b	Personality	Conflict
H11a/b	Personality	Performance
Study 3		
H12	Demographic Faultlines	Conflict Slopes
H13	Demographic Faultlines	Performance
H14	Personality Faultlines	Conflict Slopes
H15	Personality Faultlines	Performance

Note. TC = Task Conflict, RC = Relationship Conflict, and PC = Process Conflict. Ind. = individual. H1 = Hypothesis 1.

STUDY 1

This study lays the groundwork for the following studies by establishing the measurement properties and the time-based nature of team conflict. By confirming that team conflict scales are consistently measuring the same constructs over time, more researchers can conduct longitudinal analyses of team conflict with the confidence that conflict is represented consistently over time. Cluster-based analyses of team members and entire teams show whether these project-based teams all follow the same conflict trajectories, or if they take different paths. These results, unlike those that that focus on the measurement properties of team conflict, may be more sample-dependent and less likely to generalize to other team contexts. Nevertheless, the first study in this series answers research questions about the consistency and change of team conflict over time.

Methods

Participants and Procedure

I accessed questionnaire responses from an archive of data collected from the members of 273 student project teams enrolled in an 8-month engineering design course at a large Canadian university in the 2014-2015 and 2015-2016 academic years. This engineering design course consisted of multiple design projects completed sequentially that accounted for most of their final grade. Each of the 1,122 team members belongs to one three- to five-member team (M = 4.11) and teams were situated within one of several course classrooms of approximately 50 students. Within each classroom, the TeamWork Lab randomly assigned members to teams.

These archival data were sourced from two academic years: 2014-2015 and 2015-2016. The 2014-2015 academic year contained three surveys: one taken approximately two months into the teams' tenure (i.e., Survey 2), the next, six months into their team tenure (i.e., Survey 3), and the final survey, collected seven months after the team began working together (i.e., Survey 4). The 2015-2016 academic year contained an additional survey during the first work session where team members were initially assigned to teams (i.e., Survey 1). During this session, team members participated in an icebreaker activity designed to simulate the design projects that the team members will complete. Immediately after this activity, team members completed the first survey (i.e., Survey 1).

Of the 1,122 team members, 871 identified as men, 218 identified as women, and 33 did not respond or were missing from, and thus did not respond in, this data collection period. The team members' average age was 18.4 years with a standard deviation of 1.3 years. Of the 1,122 team members, 814 were native English speakers, 273 learned English as a second language, and 35 did not respond or were missing from this data collection period. Team members had seven options to indicate their ethnicity. One hundred and twenty-nine team members selected Arabic or Indian as their primary ethnicity, 196 selected East Asian, 32 selected Black, four selected Native American, 44 selected Southeast Asian, 617 selected White, and 63 selected Other, which may include multiracial team members. Thirty-seven team members did not respond to this ethnicity question or were not present for the first wave of data collection that contained demographic questions.

Context

All team members were enrolled in a mandatory Design and Innovation Studio course. The aims of this course include fostering innovation, increasing problem solving skills, and designing physical products. The instructors communicate the aims of the objectives of this course through the course outline and introductory classes in which they emphasize creativity, teamwork, problem solving, and iteratively designing solutions. Teams complete three projects subsequently over the eight-month course. The first two projects each last for two months and the final project lasts for four months. The third design project contributes the most to team members' final performance grade. To succeed in the third project, teams must create original solutions to practical problems, including reducing barriers to accessibility, developing instructional tools for STEM education, helping older adults live independently, and improving disaster relief and recovery in developing countries. For example, teams' disaster relief solutions include a flooding alert system, a solar-powered water purification device, and a hospital bed that reduces pressure sores for patients. In recent decades, student team projects have become more prevalent in information technology (Brandyberry & Bakke, 2006) and engineering (Borrego, Karlin, McNair, & Beddoes, 2013). This means results from these studies may generalize to other student project teams, especially those completing design-based innovation projects.

These projects have high stakes for team members. Project grades comprise the majority of the final grade for each team member. To progress in the engineering program and receive offers to competitive, paid internships at engineering companies around the world, team members must perform well in this course. At this stage in their degree program, high grades are a key differentiator that assists in receiving internship positions, research opportunities, and entry-level jobs after graduation. Team members whose grades are not high enough to pass must retake the course to complete their degree, resulting in serious disruptions to their progression towards graduation. This reflects the high-stakes nature of the design course and its projects from the team members' perspectives.

The teams' structural characteristics reflect project teams, classified by McGrath (1984) as constrained in time and scope. These teams complete project-based work in the absence of a defined leader with a planned dissolution point immediately after their final deadline. This supports classifying these teams, further, as agile design teams (Lindsjørn, Sjøberg, Dingsøyr, Bergersen, & Dybå, 2016; Tripp, Riemenschneider, & Thatcher, 2016) as they are self-managed and responsible for planning, coordinating, and creating their solutions to problems that require creativity. Team members may choose to assign one leader from the group, yet the similarity across team members in age, experience, and skills suggests teams may share leadership tasks, similar to many modern teams engaged in knowledge work. Although some elements of the teams' design differ from workplace teams, the samples in all three studies share team design elements with workplaces that employ engineers to design software and hardware products in a project-based manner. However, workplace teams that are not deadline-driven or working on high-stress projects may find different results. This may happen because their workplace context and external constraints are different to the project teams analyzed below.

Measures

Team members responded to existing measures of relationship, task, (Jehn, 1995) and process conflict (Behfar, Mannix, Peterson, & Trochim, 2011) in Surveys 1, 2, 3, and 4. Individual team members responded to a four-item scale of relationship conflict on a Likert-type scale from 1 to 5, with 1 = A Very Small Amount to 5 = A Lot. This same response scale is used for all conflict measures. This relationship conflict scale measures team members’ perceptions of the level of character-related disagreements in the group. An example question from this scale reads, “How much are personality conflicts evident in your team?” Task conflict is measured with three items capturing the extent to which team members feel they disagree about their work. An example item from this scale is, “How often do your team members discuss evidence for alternative viewpoints?”

Process conflict measures two sub-types of conflict: contribution conflict, which reflects disagreements about team members’ contributions, and logistical conflict, which measures challenges the team faces regarding scheduling and coordination. An example item measuring contribution conflict reads, “How often is there tension in your team caused by member(s) not performing as well as expected?” An example from the three items measuring logistical conflict reads, “How frequently do your team members disagree about the optimal amount of time to spend on different parts of teamwork?” For all conflict measures, items referred to the team and its members, not the individual responding to the measure. This referent shift to the team level is important to establish a shared team construct (Chan, 1998b).

Statistical Analyses

To test the measurement-related hypotheses in Study 1, I evaluated the conflict measures for reliability at each time point. I then calculated the interitem correlations of each measure using Cronbach’s alpha at the individual level. Subsequently, I conducted individual-level measurement invariance analyses in MPlus (Muthén & Muthén, 2019). This step is important to compare mean changes in conflict over time. If the survey waves show invariance across time, this suggests the same construct is measured consistently (Chan, 1998a).

The first step of measurement invariance includes conducting confirmatory factor analyses to determine the structure of team conflict at each survey wave. This establishes the factor structure that will be used for the three remaining steps of measurement invariance. After this, testing for invariance involves restricting the parameters across survey waves so they are equivalent. The second step involves making factor loadings equivalent, whereas the third step involves holding all intercepts to the same values across survey waves. The final measurement invariance step involves constraining all error residuals to the same at each time. However, some models do not reach these higher steps of equivalence, as the data do not fit the lower-level requirements. Alternatively, some data may partially fit the requirements of a given invariance level, yet only when certain questionnaire items are exempt from this requirement. This can result in partial measurement invariance at a certain step, instead of full measurement invariance at that level.

To determine whether these data pass a particular invariance level, I used the following cutoffs: ΔCFI of ≤ 0.01 (Chen, 2007) and ΔRMSEA of ≤ 0.005. Whereas there are many guidelines for the cutoff of change in RMSEA (e.g., Chen, 2007; Meade, Johnson, & Braddy, 2008), researchers indicate that RMSEA changes should be lower when comparing fewer groups. To reflect the influence of sample size on RMSEA values, I used a conservative cutoff for ΔRMSEA that was lower than the guidelines provided by existing research (i.e., 0.007-0.3).

After replicating the first step of establishing the same factor structure across time at the team level, I then conducted the remaining team-level measurement invariance analyses to ensure the measures are consistent across time for both levels (Jak, 2018; Jak

& Jorgensen, 2017). A common concern with Likert-type scales, such as those used in this study, is that continuous analyses do not reflect their ordinal or ordered categorical nature (e.g., Svensson, 1998). At the individual level, all conflict data for the above factor analyses were collected from measures that use 5-point, Likert-type scales. Response scales with 5-7 response options have similar results in categorical and continuous measurement models (Bovaird & Koziol, 2012; Rhemtulla, Brosseau-Laird, & Savalei, 2012). Thus, I used continuous response options to reduce computational complexity while keeping a close approximation to these data. I allowed all latent factors to correlate with each other, as these three conflict types are not perfectly orthogonal. I explored the CFA output for adequate model fit, problematic cross-loadings, and correlations between latent constructs.

Some teams and team members may start at different levels or follow different conflict trajectories. Due to this heterogeneity, my sample may reflect multiple distinct distributions of teams with unique intercepts and slopes that characterize these distributions (Wang & Bodner, 2007). As with many studies, my analysis may benefit from more personalized methods to model change rather than a single longitudinal growth model. Here, linear trajectories were used for growth modeling due to the limited number of times available for analysis. Although it would be ideal to compare curvilinear and linear growth models, the structure of these data only allowed for linear modeling. Personalized linear growth trajectories may provide more nuanced results than one linear growth trajectory for all teams and team members, despite the limitations of a linear trajectory.

Growth mixture modeling (GMM) identifies multiple intercept and slope distributions within these data to test for unobserved heterogeneity (Muthén, 2001). GMM is relevant to longitudinal studies where growth trajectories differ across individuals and/or teams and where these trajectories belong to discrete groups that are not already classified. This approach is similar to cluster analysis, in that both approaches can identify subgroups of the total sample. Yet, there are some distinctions between these methods. In GMM, one uses the prior probability of class membership to assign individuals or teams to that class.

By using this approach in conjunction with LGM, I can model change over time that better reflects a sample with more than one distribution. This analytic approach follows three steps. First, the approach identifies the number of classes reflecting distinct distributions. Second, the analytic program computes their properties including the mean and variance for the intercept and slope of each construct. Finally, the program specifies which individuals or teams belong in each class (Muthén, Brown, Khoo, Yang, & Jo, 1998; Muthén & Muthén, 1998; Muthén & Shedden, 1999). When modeling multiple, previously unspecified groups, it is necessary to first identify the correct number of classes that best fit the data. To do this, I must compare the fit of each GMM model from one class to many classes.

To evaluate the classification accuracy for placing individuals and teams into each class, researchers recommend using four main fit statistics: the Bayesian Information Criterion (BIC), the proportion of individuals or teams in each class, the average conditional probabilities of class membership (Nagin, 1999), and the entropy measure (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). The Bayesian Information

Criterion (BIC) reflects the model fit. In an iterative analytic approach such as GMM, the BIC, sample-size adjusted BIC, and AIC (Akaike Information Criterion) are lowest at the best-fitting model.

Further, the proportion of individuals or teams in each class should be over 5%. This guideline ensures that each class represents a true subgroup, to prevent over-fitting to the dataset (Rousseau & Mengersen, 2011). The average conditional probabilities of class membership reflect how clearly distinguishable each class is from another. When classifying individuals or teams into their most likely class, one can use a table of the estimated posterior probabilities for each unit of analysis (Nylund, Asparouhov, & Muthén, 2007). If the diagonal probabilities in the aforementioned table are near one and the off-diagonal probabilities are near zero, the model has good classification. Finally, the entropy measure summarizes the clarity of classification. Entropy values closer to one reflect better, more distinct classification, whereas entropy values closer to zero reflect less clarity in classification (e.g., Hix-Small, Duncan, Duncan, & Okut, 2004). After completing these analyses, I determined the number of classes of individuals and teams to analyze in Studies 2 and 3.

Results

Descriptive Statistics

To avoid issues of multicollinearity, I computed the correlations between task, relationship, and process conflict, the latter of which contains two subtypes: logistical and contribution conflict (Table 2). At the within-team level, relationship conflict was strongly correlated with logistical and contribution conflict. Surprisingly, logistical and contribution conflict were also intercorrelated at the same magnitude. This suggests that, at least within this study, relationship conflict and both process conflict subtypes are closely related. At the between-team level, relationship, logistical, and contribution conflict were intercorrelated at even higher magnitudes. This suggests high construct overlap between process conflict types and relationship conflict (Carlson & Herdman, 2012).

Task conflict did not show strong intercorrelations with other conflict types at the individual level. Specifically, relationship and task conflict were not significantly correlated, and neither were task and contribution conflict. Task and logistical conflict had a small positive relation. A similar pattern emerged at the team level; task conflict was significantly related to logistical conflict, not significantly related to contribution conflict, and had a small and slightly significant correlation with relationship conflict. This supports Hypothesis 1, which states that process and relationship conflict will be more highly correlated with each other than these conflict types would be correlated with task conflict. To reduce overlap and avoid multicollinearity in future analyses, I will only analyze task and relationship conflict.

Then, I computed the individual-level means and standard deviations for task and relationship conflict items at each survey administration wave (Table 3). One may note that relationship conflict levels remain much lower than task conflict scores over the four survey waves. These results are in line with team conflict results in other project team settings; for example, a profile analysis of team conflict levels found consistently high task conflict across four subsets of teams, with varying levels of relationship conflict (O’Neill et al., 2018). Mathematically, this suggests average task conflict across team profiles was higher than the average relationship conflict levels. Conceptually, this may relate to the type of work being completed in these project teams; team members must create design ideas and debate them with other members. Compared to teams with other task requirements such as executing routine work systems, these teams' tasks create an environment conducive to high task conflict. Across surveys, standard deviations for each item are relatively consistent. This pattern is similar for the team-level means and standard deviations (Table 4), yet variability at the team level is systematically lower than at the individual level. This suggests some variability was reduced when aggregating team members' scores to the team level.

Table 2. *Correlations between conflict types at the within- and between-team levels.*

	TC	RC	PC-LC	PC-CC
TC		.07*	.22***	.05
RC	.02		.71***	.80***
PC-LC	.05	.68***		.70***
PC-CC	.21***	.58***	.68***	

Note. RC = Relationship Conflict, TC = Task Conflict, PC = Process Conflict, LC = Logistical Conflict, CC = Contribution Conflict. * = $p < .05$, ** = $p < .01$, *** = $p < .001$. Values on the lower left are within-team correlations and values on the upper right are between-team correlations.

Table 3. *Individual means and standard deviations for task and relationship conflict items.*

Items	RC1	RC2	RC3	RC4	*TC1*	*TC2*	*TC3*
Survey 1	1.37 (0.74)	1.35 (0.68)	1.35 (0.75)	1.13 (0.47)	*2.90* *(1.09)*	*3.40* *(0.96)*	*3.23* *(1.02)*
Survey 2	1.32 (0.66)	1.35 (0.67)	1.274 (0.64)	1.15 (0.47)	*2.75* *(1.10)*	*3.37* *(0.98)*	*3.14* *(1.02)*
Survey 3	1.27 (0.60)	1.41 (1.41)	1.27 (0.61)	1.15 (0.54)	*3.19* *(1.08)*	*3.62* *(0.90)*	*3.39* *(1.02)*
Survey 4	1.28 (0.45)	1.43 (0.68)	1.26 (0.49)	1.11 (0.38)	*3.09* *(1.09)*	*3.39* *(0.98)*	*3.18* *(1.06)*

Note. RC = Relationship Conflict, TC = Task Conflict. Standard deviations are in parentheses.

Table 4. *Team means and standard deviations for task and relationship conflict items.*

Items	RC1	RC2	RC3	RC4	*TC1*	*TC2*	*TC3*
Survey 1	1.53 (0.49)	1.49 (0.46)	1.48 (0.48)	1.22 (0.32)	*3.01* *(0.63)*	*3.49* *(0.53)*	*3.34* *(0.62)*
Survey 2	1.48 (0.43)	1.54 (0.5)	1.44 (0.43)	1.26 (0.35)	*2.85* *(0.62)*	*3.34* *(0.56)*	*3.18* *(0.6)*
Survey 3	1.77 (0.74)	1.86 (0.77)	1.68 (0.71)	1.45 (0.61)	*3.25* *(0.58)*	*3.56* *(0.48)*	*3.37* *(0.57)*
Survey 4	2.02 (0.77)	2.04 (0.76)	1.87 (0.72)	1.55 (0.62)	*3.21* *(0.59)*	*3.37* *(0.54)*	*3.27* *(0.63)*

Note. RC = Relationship Conflict, TC = Task Conflict. Standard deviations are in parentheses.

To assess how much variability was accounted for by each level, I calculated the intraclass correlations (ICCs) for each task and relationship conflict item using two distinct clustering methods: within- and between-individuals and within- and between-teams (Table 5). Across each conflict item, ICCs were higher in the within- and between-individual clustering method than in the within- and between-team clustering method. More variation attributed to individual differences than repeated measures within team members reflects higher consistency within individuals and more variability across individuals in the sample. I found lower ICCs using the within- and between-team clustering method. Task conflict items had particularly low ICC values, suggesting that members of the same team had high variability compared to the variability across teams.

Other research finds task and relationship conflict have similar ICC values at the 0.13-0.14 range (Somech, Desivilya, & Lidogoster, 2009) or the 0.22-0.33 range (Greer, Caruso, & Jehn, 2011). One study, that only reported the intraclass correlation for task conflict, had a slightly higher value at 0.10 (Costa, Passos, & Bakker, 2015). Computing ICCs separately for each time point allows for more fine-grained analyses of when within- and between-team variability is highest (Table 6). Whereas the ICCs for task conflict items did not differ significantly across survey administration waves, the ICCs for relationship conflict items were markedly higher in surveys 3 and 4 compared to surveys 1 and 2. This suggests that team members had lower variability in relationship conflict scores in surveys 3 and 4 compared to the variability across teams at these same times.

Table 5. *Intraclass correlations across all times using two clustering methods.*

Items	RC1	RC2	RC3	RC4	*TC1*	*TC2*	*TC3*
Within- & Between-Individual	0.26	0.31	0.28	0.25	*0.27*	*0.22*	*0.25*
Within- & Between-Team	0.16	0.16	0.15	0.12	*0.08*	*0.08*	*0.09*

Note. RC = Relationship Conflict, TC = Task Conflict.

Table 6. *Intraclass correlations at each time using within- & between-team clustering.*

Items	RC1	RC2	RC3	RC4	*TC1*	*TC2*	*TC3*
Survey 1	0.13	0.11	0.13	0.11	*0.07*	*0.07*	*0.13*
Survey 2	0.07	0.12	0.07	0.06	*0.11*	*0.11*	*0.13*
Survey 3	0.35	0.32	0.32	0.27	*0.08*	*0.04*	*0.06*
Survey 4	0.32	0.21	0.25	0.17	*0.10*	*0.06*	*0.12*

Note. RC = Relationship Conflict, TC = Task Conflict.

Measurement Analyses

Confirmatory Factor Analyses: Individual Level. I conducted four individual-level confirmatory factor analyses (CFAs) with relationship and task conflict to examine the factor structure of team conflict at each of four survey administration waves. All four CFAs showed good to excellent fit statistics across categories, except for the root mean squared error of approximation (RMSEA) values for Survey 3 and Survey 4 (Table 7). These two values were higher than 0.05, a commonly accepted rule of thumb for good fit (Awang, 2012; Hair, Black, Babin, & Anderson, 2010). Only the RMSEA for Survey 4 was above the cutoff for acceptable fit, at 0.08 (Awang, 2012). The modification indices in the Survey 4 CFA suggested that some items in the same factor may have correlated residuals. However, the modification indices did not recommend these changes for every survey wave. Thus, I could improve the fit slightly for the Survey 4 CFA model, yet these changes would not improve the fit of all four time points.

All CFI and TLI values were over 0.95, suggesting that fit could not be substantially improved on these metrics (Lance, Butts, & Michels, 2006; Forza & Filippini, 1998). Finally, the standardized root mean squared residual (SRMR) indices were well below cutoffs suggested by Hu and Bentler (1999) of 0.08 and Ringle (2016) of 0.10. This suggests the model fit well for all four waves of survey administration, with slightly poorer fit towards the later time points. Factor structures at the individual level suggest that factor loadings were similar across survey administration waves (Table 8). Whereas all factor loadings were over 0.40 (Cabrera-Nguyen, 2010), suggesting they were not weak, some item loadings became stronger as the team worked together whereas other item loadings slightly decreased. The latent variables calculated for relationship and task conflict were not significantly correlated in Surveys 1-3 (Figure 1). At Survey 4, there was a small yet significant positive correlation between task and relationship conflict. Taken together, these results suggest the proposed measurement model fits these data reasonably well.

Table 7. *Model fit for individual-level CFAs at four surveys.*

Wave	N	Teams	Chi square	RMSEA	CFI	TLI	SRMR
# 1	606	157	$\chi^2(13) = 29.86$***	0.046 [90% CI: 0.024, 0.068]	0.99	0.98	0.033
# 2	1,052	272	$\chi^2(13) = 41.99$***	0.046 [90% CI: 0.031, 0.062]	0.99	0.98	0.039
# 3	1,048	272	$\chi^2(13) = 63.86$***	0.061 [90% CI: 0.047, 0.076]	0.99	0.98	0.037
# 4	1,033	271	$\chi^2(13) = 119.90$***	0.089 [90% CI: 0.075, 0.104]	0.98	0.96	0.048

Note. Wave = Survey administration wave. N = Team member sample size. RMSEA = Root mean squared error of approximation, CFI = Confirmatory Fit Index, TLI = Tucker-Lewis Index, SRMR = Squared root mean residual. All chi square values are significant at *** = $p < .001$.

Table 8. *Factor loadings for individual-level CFAs at four surveys.*

Items	Survey 1	Survey 2	Survey 3	Survey 4
RC1	.76	.80	.88	.90
RC2	.65	.76	.86	.81
RC3	.78	.85	.91	.93
RC4	.70	.70	.83	.78
TC1	*.44*	*.55*	*.65*	*.76*
TC2	*.68*	*.74*	*.83*	*.83*
TC3	*.90*	*.82*	*.77*	*.82*

Note. RC = Relationship Conflict, TC = Task Conflict. All factor loadings are significant at $p < .001$.

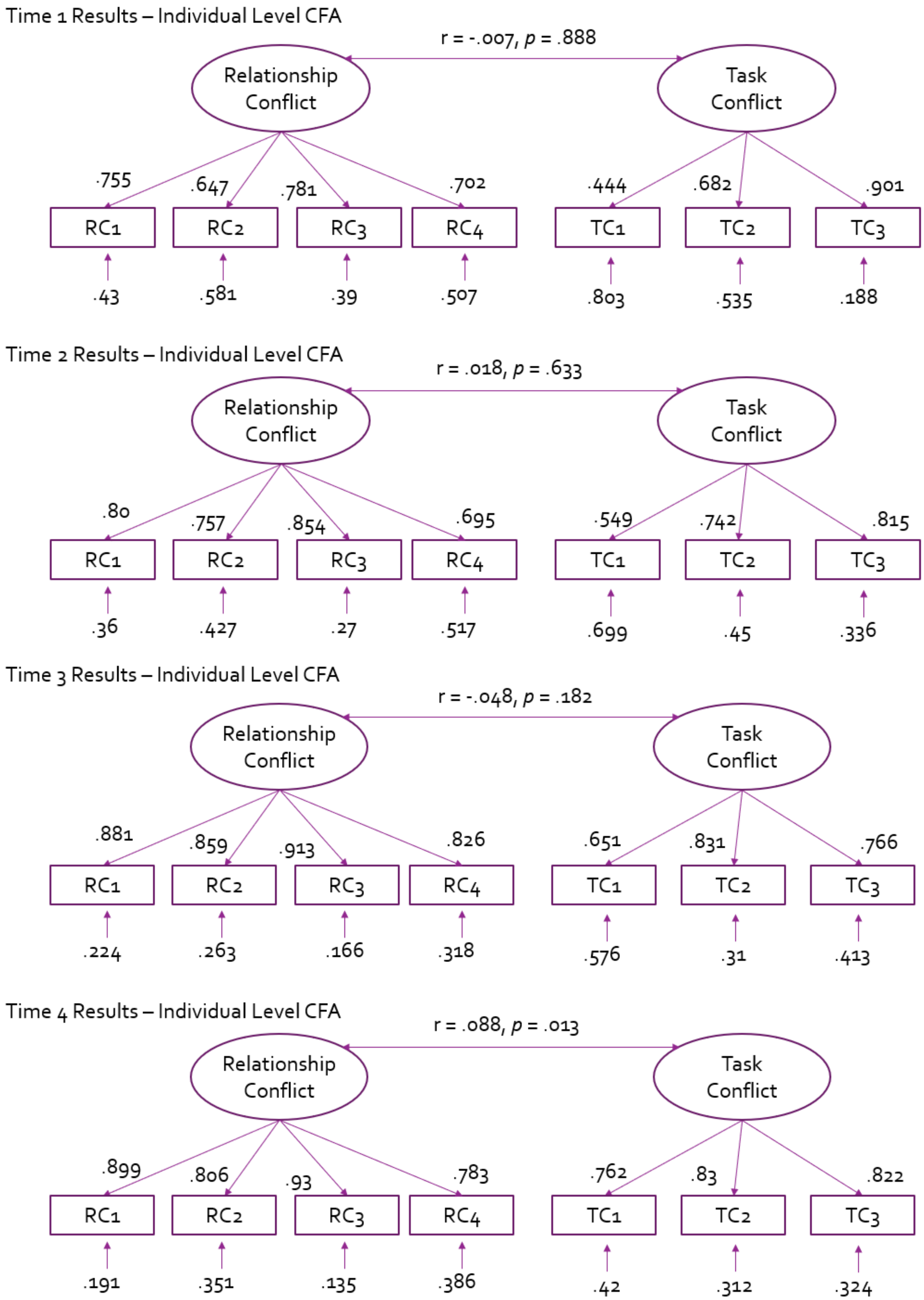

Figure 1. Confirmatory factor analysis results at the individual level.

Confirmatory Factor Analyses: Team-Level. The ideal confirmatory factor analysis method for these data is a multilevel (i.e., two-level) CFA. Unfortunately, this model did not converge. Thus, I conducted four team-level confirmatory factor analyses (CFAs) of relationship and task conflict items to analyze the factor structure of team conflict at each of four survey administration waves. All four CFAs showed acceptable but not good fit statistics across categories (Table 9). The worst fitting time point was Survey 4, in which the root mean squared error of approximation (RMSEA) value was over the acceptable value of 0.08 (Awang, 2012). All but one CFI and TLI value were over 0.95, suggesting that fit could not be improved on these metrics (Hu & Bentler, 1999). The lower TLI value was associated with the team-level model in Survey 4, where the fit statistic was in the “good” range between 0.9 and 0.95, yet below the excellent range of 0.95 and above (Awang, 2012; Forza & Filippini, 1998). Finally, standardized mean squared residual (SRMR) indices were below cutoffs suggested by Hu and Bentler (1999) of 0.08 and Ringle (2016) of 0.10. This suggests the model fit well for all four waves of survey administration, with poorer fit towards the later time points.

The modification indices for each model found no improvements to the Survey 1 CFA and small improvements for Surveys 2-4. Thus, any changes to the factor structure would not benefit all survey waves. Factor structures at the team level suggest that factor loadings were similar across survey administration waves (Table 10). All factor loadings were above 0.40 with some items showing stronger loadings as the team worked together for longer. The relationship and task conflict latent variables were not significantly correlated for the first three surveys (Figure 2), yet there was a significant positive correlation between the two latent variables at the last survey. These results suggest the proposed measurement model fit these data somewhat well at the team level of analysis.

Table 9. *Model fit for team-level CFAs at four surveys.*

Survey	Teams	Chi square	RMSEA	CFI	TLI	SRMR
# 1	157	$\chi^2(13) = 27.46$*	0.084 [90% CI: 0.039, 0.128]	0.97	0.952	0.046
# 2	272	$\chi^2(13) = 24.85$*	0.058 [90% CI: 0.02, 0.092]	0.986	0.978	0.052
# 3	272	$\chi^2(13) = 43.82$***	0.093 [90% CI: 0.064, 0.125]	0.98	0.968	0.057
# 4	271	$\chi^2(13) = 85.96$***	0.144 [90% CI: 0.116, 0.174]	0.955	0.927	0.073

Note. RMSEA = Root mean squared error of approximation, CFI = Confirmatory Fit Index, TLI = Tucker-Lewis Index, SRMR = Squared root mean residual. * = $p < .05$, *** = $p < .001$.

Table 10. *Factor loadings for team-level CFAs at four surveys.*

Items	Survey 1	Survey 2	Survey 3	Survey 4
RC1	.80	.81	.92	.95
RC2	.76	.79	.92	.85
RC3	.84	.88	.96	.96
RC4	.84	.68	.90	.84
TC1	*.52*	*.71*	*.70*	*.85*
TC2	*.75*	*.77*	*.80*	*.81*
TC3	*.93*	*.88*	*.80*	*.86*

Note. RC = Relationship Conflict, TC = Task Conflict. All factor loadings are significant at $p < .001$.

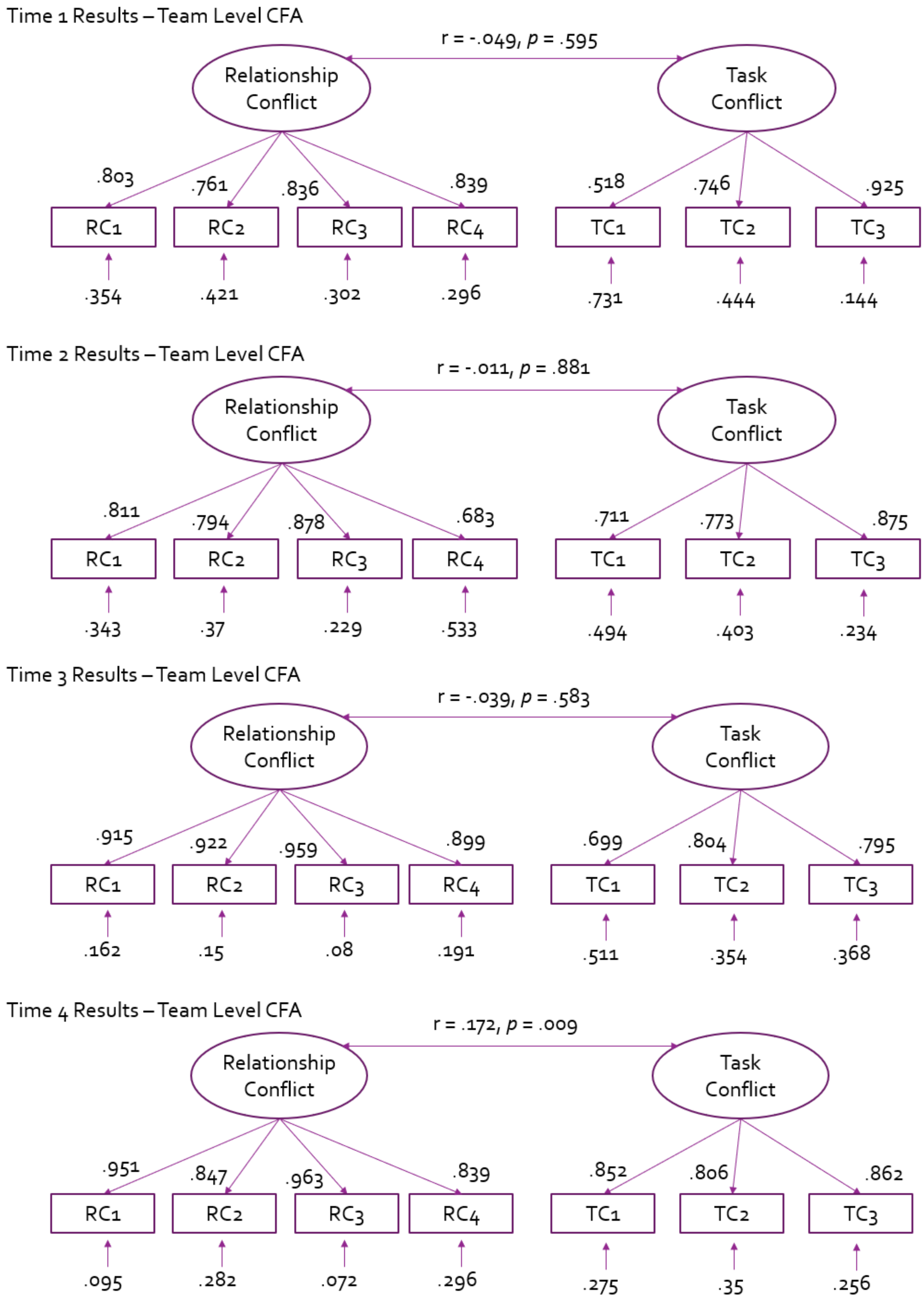

Figure 2. Confirmatory factor analysis results at the team level.

Measurement Invariance: Individual-Level. I conducted measurement invariance analyses across survey administrations by following the four-step method: first testing configural invariance, then weak (loading) invariance, then strong (scalar/intercept) invariance, and finally strict (residual) invariance if the model passed all other levels. Due to missing data at Survey 1, the sample size for this time was approximately half of the other surveys. All analyses used maximum likelihood estimation to handle missing data across surveys. The individual-level model passed the configural, weak (loading), and strong (scalar/intercept) invariance stages, yet it failed to demonstrate similar model fit across time at the strict (residual) invariance stage without modifications (Table 11). Specifically, the model showed significant decreases in fit across all three indices: chi square, CFI, and RMSEA. Thus, I freed three parameters one at a time, starting with parameters associated with the highest modification indices. The model showed minimal decreases in fit across CFI and RMSEA values after freeing the residual error of the fourth relationship conflict item at the fourth survey administration (i.e., Survey 4), after freeing the residual error of the third relationship conflict item at the first survey administration (i.e., Survey 1), and after freeing the residual error the first task conflict item at the fourth survey administration.

By freeing the residual errors of the three items above, the strict measurement invariance model showed minimal changes in fit across the CFI and RMSEA indices. Whereas the chi square value was significantly higher than the strong invariance model, significant chi-square values are common with large sample sizes. Thus, I focused on comparative fit indices such as CFI and RMSEA instead of significance testing, as recommended by Kline (2015). The fit indices’ absolute values were also in the excellent range, as RMSEA and SRMR values were below 0.05 and CFI and TLI values were over 0.95 in the final model. This means that I established partial strict measurement invariance with some modifications (i.e., freeing the residual error values of three items: one at Survey 1 and two at Survey 4) for the individual-level conflict model, supporting Hypothesis 2a. In the studies below, individual-level analyses will show reliable differences across intercepts. The error terms of these questionnaire items are equal for all but three items across time.

Table 11. *Model fit for individual-level measurement invariance analyses.*

Invariance Level	Chi-Square	RMSEA	CFI	TLI	SRMR	AIC / BIC	Δ Chi Square	ΔCFI	ΔRMSEA
Configural	$\chi^2(280)$ = 536.85***	0.029 [90% CI: 0.025, 0.032]	0.982	0.975	.036	57526.08 / 58299.60			
Weak (Loading)	$\chi^2(295)$ = 600.55***	0.030 [90% CI: 0.027, 0.034]	0.978	0.972	.04	57559.78 / 58257.96	$\chi^2(15)$ = 63.7*** fail	.004 pass	.001 pass
Strong (Scalar / Intercept)	$\chi^2(310)$ = 746.35***	0.035 [90% CI: 0.032, 0.039]	0.969	0.962	.04	57675.59 / 58298.42	$\chi^2(15)$ = 145.80*** fail	.009 pass	.005 pass
Strict (Residual)	$\chi^2(331)$ = 1057.14***	0.044 [90% CI: 0.041, 0.047]	0.948	0.941	.058	57944.38 / 58461.73	$\chi^2(21)$ = 310.79*** fail	.021 fail	.009 fail
Strict – freed T4_RC4 residual	$\chi^2(330)$ = 995.005***	0.042 [90% CI: 0.039, 0.045]	0.952	0.946	.053	57884.24 / 58406.62	$\chi^2(20)$ = 248.65*** fail	.017 fail	.007 fail
Strict – freed T1_RC3 residual	$\chi^2(329)$ = 947.02***	0.041 [90% CI: 0.038, 0.044]	0.956	0.949	.052	57838.26 / 58365.66	$\chi^2(19)$ = 200.67*** fail	.013 fail	.006 fail
Strict – freed T4_TC1 residual	$\chi^2(328)$ = 889.97***	0.039 [90% CI: 0.036, 0.042]	0.96	0.954	.048	57783.20 / 58315.63	$\chi^2(18)$ = 143.61*** fail	.009 pass	.004 pass

Note. T4 = Time 4, T1 = Time 1. RC = Relationship Conflict, TC = Task Conflict. RMSEA = Root mean squared error of approximation, CFI = Confirmatory Fit Index, TLI = Tucker-Lewis Index, SRMR = Squared root mean residual, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. The sample size for T1 was 606 team members and the sample size for T2-T4 was 1,122 team members. *** = $p < .001$.

Measurement Invariance: Team-Level. I then conducted measurement invariance analyses across survey administrations at the team level. First, I tested configural invariance, then weak (loading) invariance, then strong (scalar/intercept) invariance, and finally strict (residual) invariance if the model passed all other levels. The team-level model passed the first two stages, suggesting the two conflict types had configural and loading invariance across time. Yet, the team-level model failed to demonstrate similar fit at the strong (scalar/intercept) stage (Table 12). Specifically, the model showed significant decreases in fit across all three indices: chi square, CFI, and RMSEA. Thus, I freed item intercepts one at a time, starting with the items with the highest modification indices. After freeing the intercepts of the first task conflict item at Surveys 3 and 4, the model showed minimal decreases in CFI and RMSEA as measures of model fit. The change in chi-square values was still significant, which suggests the fit is not substantially better on this metric, yet large and significant chi-square values are common for models with a large sample size. I therefore relied on changes in CFI and RMSEA to determine the measurement invariance level.

The fit indices' absolute values suggested good to excellent fit for some measures, as RMSEA values were below 0.05 and CFI and TLI values were over 0.95 in the partial strong measurement invariance model. However, the SRMR values were acceptable but not good, as they were over 0.05 but less than 0.08. Thus, I established partial strong measurement invariance at the team level with two modifications (i.e., freeing the intercepts for the first task conflict item at Survey 3 and 4). Subsequent team-level analyses will show reliable differences across all item intercepts except one. However, some survey administration waves will contain more error than others as this model did not reach the strict, residual invariance stage. This supports Hypothesis 2b, where I posited that conflict would display measurement invariance at the team level.

Table 12. *Model fit for team-level measurement invariance analyses.*

Invariance Level	Chi-Square	RMSEA	CFI	TLI	SRMR	AIC / BIC	Δ Chi Square	ΔCFI	ΔRMSEA
Configural	$\chi^2(280)$ = 416.95***	0.042 [90% CI: 0.034, 0.051]	0.972	0.962	.061	7074.11 / 7629.97			
Weak (Loading)	$\chi^2(295)$ = 452.62***	0.044 [90% CI: 0.036, 0.052]	0.968	0.959	.07	7079.78 / 7581.50	$\chi^2(15)$ = 35.67** fail	.004 pass	.002 pass
Strong (Scalar / Intercept)	$\chi^2(310)$ = 578.17***	0.056 [90% CI: 0.049, 0.063]	0.946	0.934	.07	7175.33 / 7622.90	$\chi^2(15)$ = 125.55*** fail	.022 fail	.012 fail
Strong – freed T4_TC1 intercept	$\chi^2(309)$ = 532.72***	0.051 [90% CI: 0.044, 0.059]	0.955	0.944	.071	7131.88 / 7583.07	$\chi^2(14)$ = 8.1*** fail	.013 fail	.007 fail
Strong – freed T3_TC1 intercept	$\chi^2(308)$ = 505.0***	0.048 [90% CI: 0.041, 0.056]	0.96	0.951	.071	7106.16 / 7560.95	$\chi^2(13)$ = 52.38*** fail	.008 pass	.004 pass

Note. T4 = Time 4, T3 = Time 3. TC = Task Conflict. RMSEA = Root mean squared error of approximation, CFI = Confirmatory Fit Index, TLI = Tucker-Lewis Index, SRMR = Squared root mean residual, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. The sample size for Time 1 was 158 teams and the sample size for Times 2-4 was 273 teams. *** = $p < .001$.

Time-Based Analyses

Time as a Predictor of Conflict Scores. After establishing measurement invariance across time, I analyzed the role of time in predicting team conflict scores. This step is a helpful precursor to the future analyses using growth modeling. As many hypotheses rely on the expectation that conflict scores change over time, it is important for to first establish whether this is true for all conflict types. Specifically, I conducted a multilevel structural equation model (SEM) analysis with time as a predictor of task and relationship conflict within- (i.e., members' scores from each time) and between-individuals (i.e., members' scores irrespective of time). In this model, time was a within-level variable. This multilevel approach separates the variance accounted for at the within-person level, where this hypothesis will be tested, from the variance attributed to the between-person level. The model fit was very good ($\chi^2(31) = 294.69$, $p < .001$, RMSEA = .048, CFI = .97, TLI = .96, SRMR Within = .05, SRMR Between = .06). Within individual members, time predicted relationship conflict ($\beta = 0.27$, SE = 0.02, $p < .001$), yet time did not predict task conflict ($\beta = 0.04$, SE = 0.02, $p = .05$).

Then, I conducted another multilevel SEM analysis using within-team (i.e., teams' aggregated scores from each time) and between-team (i.e., teams' aggregated scores irrespective of time) levels. As above, time was a within-level predictor. However, the multilevel nature of this analysis remains helpful for ensuring higher accuracy by reflecting the clustered nature of this data, even if the hypotheses about time as a predictor cannot be tested at the higher level. This model did not fit the data as well as the previous model ($\chi^2(31) = 260.18$, $p < .001$, RMSEA = .087, CFI = .95, TLI = .92, SRMR Within = .06, SRMR Between = .12). This supports the descriptive ICC results above, which showed that variance accounted for at the between-team level was lower than variance accounted for at the between-person level. However, results on the predictive power of time were similar. Time predicted relationship conflict scores within groups ($\beta = 0.38$, SE = 0.03, $p < .001$), but time did not predict task conflict scores within groups ($\beta = 0.08$, SE = 0.05, $p = .08$). This suggests that relationship conflict for team members and teams differs across time, yet time is not a significant predictor of scores for task conflict. These results largely support the third hypothesis, that time will explain differences in team conflict scores.

Identifying Clusters in Longitudinal Data

Individual-Level Growth Mixture Modeling (GMM). To identify the number of classes that reflect team members' conflict trajectories, I used the scale scores for relationship and task conflict in Surveys 2, 3, and 4. These and future analyses include only three surveys; this is because one full year of data (approximately 50% of the sample) was missing at Survey 1. This can lead to less accurate intercept estimates that reflect a much lower sample size. To avoid this, I used data from Surveys 2, 3, and 4; these scores were collected after teams began working on their assigned design projects, whereas data from Survey 1 was collected upon meeting one's team members. This means conflict begins to represent the team's interactions about their work at Survey 2, not at Survey 1.

Growth mixture modeling involves combining two analytic approaches: latent growth modeling and mixture modeling. LGM uses longitudinal data to estimate a growth curve that best fits the sample. Although these growth curves can take many shapes, a common trajectory is a linear one as it requires fewer time points to compute. As only three surveys reflected the team's interactions about their work in this sample, only linear growth models were possible for this analysis. Thus, for all growth trajectories in these three studies, I will be using linear models. This is for consistency and straightforward interpretation; although one year of team data has four survey waves, the other year only has three. If I were to model one quadratic growth model using four time points in Study 2 and another linear growth model using three time points in Study 3, I could not directly compare the results from Study 2 with those from Study 3. In addition, I could not conduct combined analyses in the present study (i.e., Study 1) if I were to use two types of growth curves in future studies.

This longitudinal growth modeling approach calculates an average intercept and slope for the entire sample, plus the associated variability around these values. Mixture modeling uses a different approach. There are multiple forms of mixture modeling, including factor and growth mixture modeling. Factor mixture modeling is similar to latent class analysis. It identifies subgroups of units (e.g., participants, teams, or animals) within a full sample that reflect a separate normal distribution. When taken together, all the subgroups and their associated normal distributions may better reflect the full set of data than one normal distribution can. GMM uses a similar approach, though it uses growth curves that reflect multiple subgroups instead of raw data or latent variables.

I conducted GMM analyses from one class to four classes using a maximum likelihood robust estimator with 80 and 16 random starts. The GMM models with one to three classes were able to replicate their best loglikelihood value, whereas the model with four classes did not replicate the best solution even when doubling the number of random starts. The model with four classes also had one class containing less than 5% of cases, suggesting the model may be overfitting the data. Despite these issues with the four-class solution, I compared the fit statistics for classes one to four to identify the optimal solution. In addition, I plotted the three fit statistics for all solutions on an elbow plot (Figure 3). This visual inspection can determine the inflection point where additional classes do not provide a substantial improvement to the fit, despite ever-decreasing values.

The BIC, sample size adjusted BIC, and the AIC all indicated that a model with three classes was the best solution as their values were lowest in this condition (Table 13), supporting the fourth hypothesis that multiple classes of team member conflict trajectories exist. However, the elbow plot indicated a slight inflection point at two classes. There were substantial improvements in fit beyond two classes, but these changes were not visually as drastic as the change from one class to two. As we could not test additional classes, the elbow plot may not have enough information to show a reliable “leveling off” point. The RC intercept had negative variance in this three-class solution, so I fixed the variance of this parameter to zero. Only the slopes of task and relationship conflict were correlated at r = 0.02, $p < .05$. Of these three classes, Class 1 is characterized by a similar task conflict intercept, no significant task or relationship conflict slope, and a relationship conflict intercept in the middle of the three classes (Figure 4). Class 2 is characterized by a similar task conflict intercept, a small yet positive task conflict slope, a low relationship conflict intercept, and a large, positive relationship conflict slope. Class 3 is characterized by a similar task conflict intercept, a non-significant task conflict slope, the highest relationship conflict intercept, and a negative relationship conflict slope (Table 14).

Table 13. *Model fit for four GMM model analyses at the individual level.*

# of Classes	BIC	Sample Size Adjusted BIC	AIC	Members in Each Class
One (LGM)	14,933.72	14,870.2	14,833.32	C1 (1,119)
Two	14,543.28	14,463.87	14,417.77	C1 (982) C2 (137)
Three	14,327.79	14,245.21	14,197.27	C1 (194) C2 (846) C3 (79)
Four	14,173.06	14,061.89	13,997.35	C1 (83) C2 (847) C3 (18) C4 (171)

Note. BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion. LGM = Latent Growth Modeling. C1-C4 = Class 1 – Class 4. The sample size for all analyses was constant at 1,119 team members. The Class 4 solution had one sample representing less than 5% of cases; as well, the best solution was not replicated, which means this four-class solution may not be trustworthy.

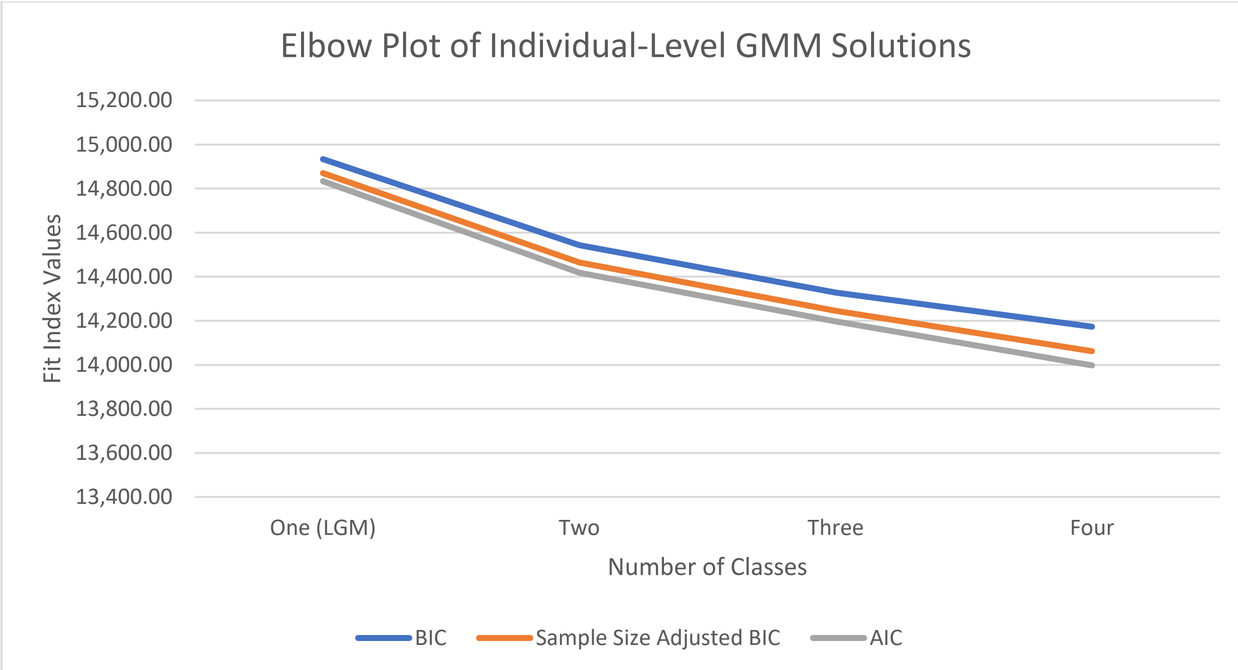

Figure 3. Elbow plot of individual-level growth mixture modeling solutions.

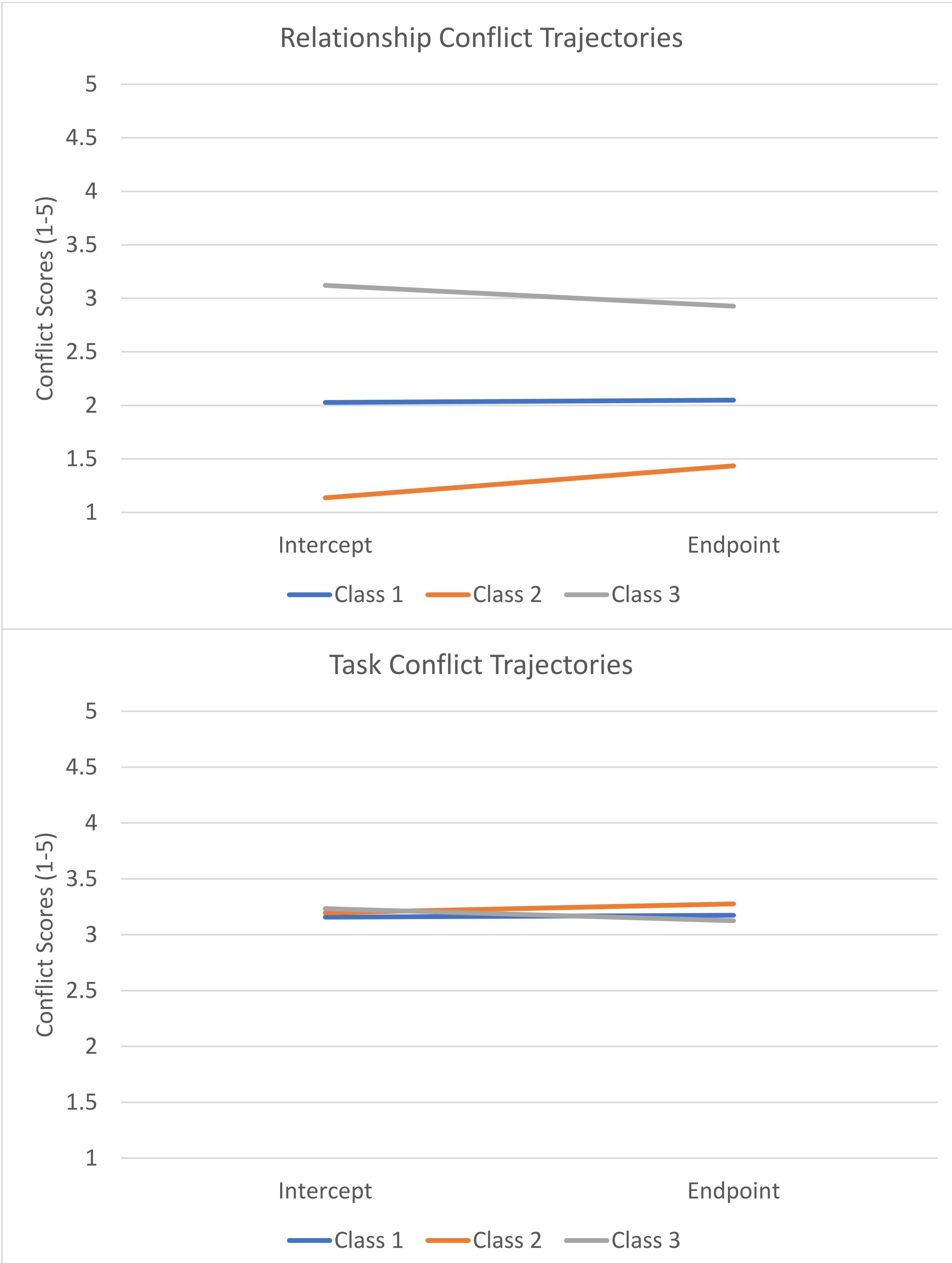

Figure 4. Individual-level classes of conflict trajectories.

Table 14. *Mean and variance of individual conflict slopes and intercepts for a 3-class solution.*

Descriptive Statistics	Relationship Conflict	Task Conflict
Class 1	Slope: 0.022 (0.16***) Intercept: 2.03*** (0)	Slope: 0.017 (0.07**) Intercept: 3.16*** (0.2***)
Class 2	Slope: 0.3*** (0.16***) Intercept: 1.14*** (0)	Slope: 0.08*** (0.07**) Intercept: 3.2*** (0.2***)
Class 3	Slope: -0.19** (0.16***) Intercept: 3.12*** (0)	Slope: 0.11 (0.07**) Intercept: 3.23*** (0.2***)

Note: Variance values are in parentheses; the relationship conflict intercept had negative variance in the model, so its variance is set to zero. ** = $p < .01$, *** = $p < .001$.

Team-Level Growth Mixture Modeling. To classify the team trajectories into subgroups, I then conducted GMM analyses from one to four classes using the same parameters as the individual-level models. The one-class solution is identical to the latent growth model I will use in future studies; these trajectories closely replicate the observed means for task and relationship conflict at each survey (Figure 5). The variance in this analysis suggests there may be multiple unobserved classes in the sample. Specifically, the task conflict intercept has significant variance across teams (variance = 0.077, $p < .001$), as does the relationship conflict intercept (variance = 0.11, $p < .001$). Whereas the task conflict slope's variance is set to zero due to issues with the residual variance of one wave of survey data, the variance of the relationship conflict slope is significant (variance = 0.11, $p < .001$). The GMM models with one to three classes replicated their best loglikelihood value, a score that reflects the best solution for these data. This means the best solution was stable for analyses up to three classes. However, the statistical model with four classes did not replicate the best solution, even when I doubled the number of iterations for the program. The model may not have a stable solution because one class contained less than 5% of the total sample size. This suggests the model may be overfitting the data. Thus, I discarded the four-class solution as it was unstable and compared the fit statistics for up to three classes to find the optimal solution.

The sample size adjusted BIC and the AIC, yet not the standard BIC values, both indicated that a model with two classes was the best solution for the team-level data as their values were lowest for this model (Table 15). This provides support for Hypothesis 5, that multiple conflict trajectory classes exist at the team level. However, the task conflict slope had negative variance in this solution, so I fixed the variance of this parameter to zero to find a stable model of both classes. In the two-class analysis, no slopes or intercepts were significantly correlated. Of these two classes, Class 1 has a slightly lower task conflict intercept, a moderate, positive task conflict slope, a slightly lower relationship conflict intercept, and a large, positive relationship conflict slope (Figure 6). Class 2 has a slightly higher task conflict intercept, a small, positive task conflict slope, a slightly higher relationship conflict intercept, and a moderate, positive relationship conflict slope. (Table 16). Team-level analyses in Studies 2 and 3 will use this two-class solution to group teams' conflict trajectories for increased accuracy and predictive power. I use the number of classes found here in future studies. It would be ideal to recalculate the optimal number of classes in each subsequent study, as the distributions may be different in the scores each year. Unfortunately, the smaller sample sizes in Studies 2 and 3 would increase the risk of overfitting the sample data and reduce the accuracy of the intercept and slope estimates. Considering the circumstances and the large sample size requirements of growth mixture modeling, I will carry the two-class, team-level solution forward to future studies.

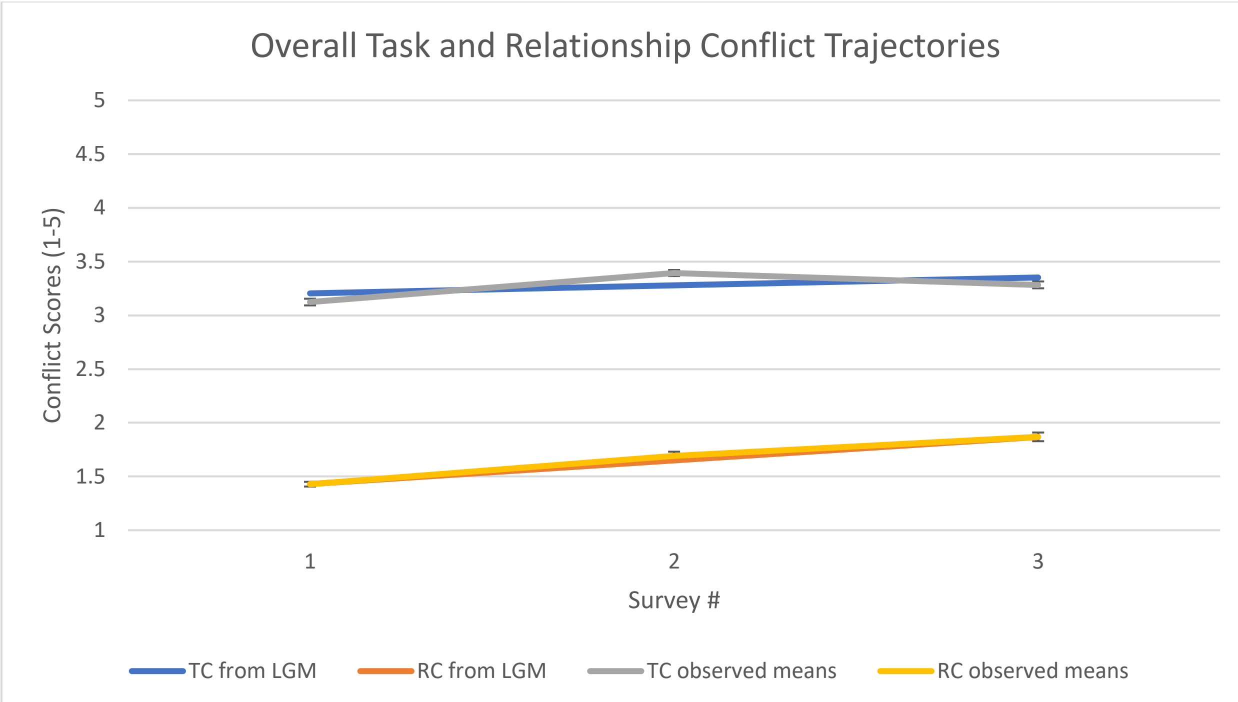

Figure 5. Team-level longitudinal growth model results for task and relationship conflict.

Note. TC = Task Conflict. RC = Relationship Conflict. LGM = Longitudinal Growth Modeling. The values labeled "TC from LGM" and "RC from LGM" represent the intercept and slope from the longitudinal growth model. The values labeled "TC observed means" and "RC observed means" represent the average values for each survey. These values are not perfectly linear, unlike the LGM results. Observed means have standard error bars denoting variation around the average.

Table 15. *Model fit for three GMM analyses at the team level.*

# of Classes	BIC	Sample Size Adjusted BIC	AIC	Teams in Each Class
One (LGM)	2,391.94	2,344.38	2,337.80	C1 (273)
Two	2,369.39	2,305.98	2,297.20	C1 (240) C2 (33)
Three	2,314.58	2,384.33	2,304.93	C1 (187) C2 (60) C3 (26)

Note. BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion. LGM = Latent Growth Modeling. C1-C3 = Class 1 – Class 3. The sample size for all analyses was constant at 273 teams.

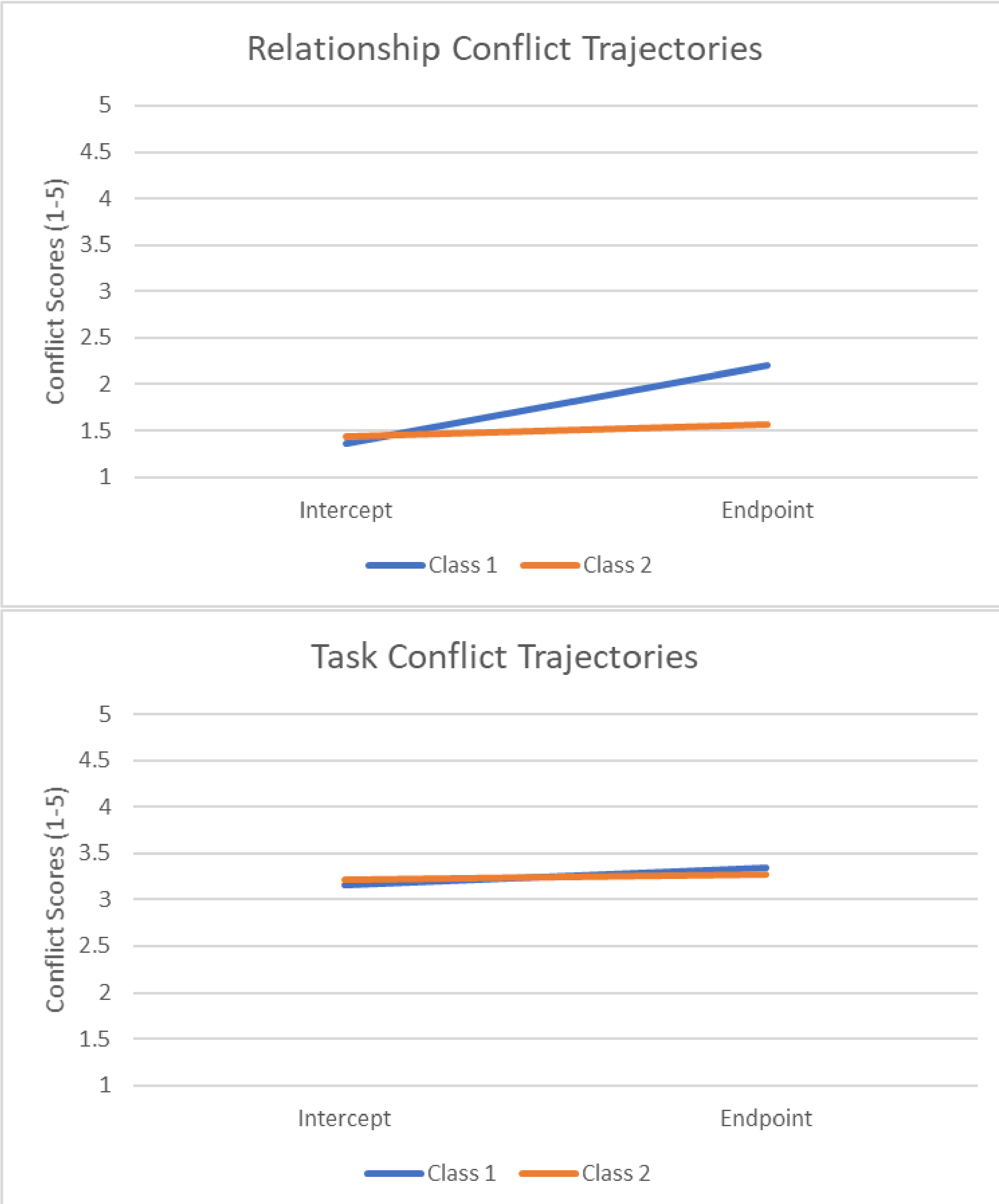

Figure 6. Team-level classes of conflict trajectories.

Table 16. *Mean and variance of team conflict slopes and intercepts for a 2-class solution.*

Descriptive Statistics	Relationship Conflict	Task Conflict
Class 1	Slope: 0.83*** (0.05***) Intercept: 1.37*** (0.11***)	Slope: 0.2** (0) Intercept: 3.15*** (0.077***)
Class 2	Slope: 0.13*** (0.05***) Intercept: 1.44*** (0.11***)	Slope: 0.05** (0) Intercept: 3.21*** (0.077***)

Note: Variance values are in parentheses; the task conflict slope had negative variance in the model, so its variance is set to zero. ** = p < .01, *** = p < .001.

Discussion

Team conflict, specifically of the task and relationship kind, shows a consistent structure across four waves of data collection. These conflict measures have partial strong measurement invariance, suggesting that the factor structure, loadings, and some intercepts are consistent across time. Having established measurement invariance at the strong but not strict level, I continued to evaluate the role of time in predicting task conflict scores at two levels. In a preliminary time-based analysis, time predicted relationship conflict scores but not task conflict scores at both levels. This suggests task conflict levels are consistent for teams across time, whereas levels of relationship conflict differ through the team's lifecycle.

To understand the role of time in teams further, I conducted growth mixture model analyses at the individual and team levels. These analyses identify clusters of teams and team members who display similar patterns of scores over time. These analyses demonstrate that three subgroups of team member trajectories exist at the individual level. One class of team members show flat trajectories and medium intercepts on task and relationship conflict. The second class of team members has positive slopes for task and relationship conflict with low relationship conflict intercepts and moderate task conflict intercepts. The final class of team members show a negative slope of relationship conflict, no significant task conflict slope, and the highest intercepts for task and relationship conflict. This means one class of team members perceive moderate task and relationship conflict consistently across three surveys, another class of team members report increasing conflict of both types that starts with low relationship conflict, and the last set perceives consistent task disagreements with declining relationship conflict.

Two subgroups of team trajectories exist at the team level, with different characteristics than the individual-level classes. One set of teams shows steep, positive relationship and task conflict slopes, whereas the other set of teams shows relationship and task conflict scores that increase more slowly over time. These two classes showed similar intercepts for both conflict types, suggesting that teams begin at similar levels of conflict, yet some teams escalate conflict more rapidly than other teams. To better understand why teams follow different trajectories, Studies 2 and 3 use personality and demographic traits to explain differences within and across classes.

At the start of teams' projects, relationship conflict was particularly low: this may be for a few reasons. First, teams may not have had enough time for disagreements to emerge. As initial surveys were collected as soon as 20 minutes of working together on an icebreaker activity, personal or relationship conflicts likely did not yet come up. Second, team members may feel social pressure to avoid relationship conflict at the start of their interactions. Team members were randomly assigned to groups, suggesting that most members did not have existing relationships with their group colleagues. Social norms of politeness when meeting others (Laver, 2011; Terkourafi, 2005) may put pressure on team members to avoid airing personal disagreements early in teams' lifecycles. These social norms could reduce the amount of relationship conflict that was expressed in teams, even if some team members did in fact personally disagree with other members.

Limitations

There are some limitations in the sample, design, and analysis of this study. This sample comprised engineering trainees in a limited age range who completed a course-based project in a university setting. Thus, the team context does not reflect a workplace situation, thus limiting the generalizability of these findings to work teams. The increased control and standardization of this context, in which all teams complete similar projects under identical evaluation criteria, are beneficial to maintain consistency and make causal conclusions from the data. However, this standardization reduces the external validity and generalizability of this sample.

The design of this study included four waves of survey administration to model team conflict over time. The low number of survey waves reduces the opportunity to conduct fine-grained analyses of changes over time that would be possible with a more frequent data collection method. Other research using daily (e.g., Kurtzberg & Mueller, 2005) or weekly (e.g., Banker, Field, Schroeder, & Sintia, 1996) study designs can identify changes in team processes that may be overlooked in the current design. The study design also includes choices of which measures to use that best reflect each conflict type. In this set of studies, process conflict was measured using Behfar and colleagues' (2011) questionnaire, not the process conflict scale developed by Jehn (2001). Despite reportedly measuring the same construct, Behfar and colleagues' measure mentions time-based conflict more often than does Jehn's measure. Further, value-laden terms such as "tension" appear in Behfar and colleagues' measure, which may increase the overlap between process and relationship conflict in this sample than would appear in a sample using Jehn's (2001) process conflict measure. Other workplace-relevant constructs such as pacing styles for completing work (Gevers, Rutte, & Van Eerde, 2006) and cultural norms of time (Arman & Adair, 2012) may explain differences in team members' process conflict perceptions and team-level conflict levels that are time-related. Future research could compare measures of process conflict and compare this construct to individual-level time perceptions and work styles.

Finally, the analyses conducted were mainly single-level analyses replicated at the between-team level and the between-person level. In part, this was due to difficulties conducting complex multilevel analyses within the software program. Multilevel designs that analyze both levels at once could provide increased rigor and accuracy to conclusions drawn from these analyses, especially at the between-person (i.e., within-team) level in which observations violate the assumption of independence. Studies 2 and 3 aim to address this limitation in analysis by conducting multilevel analyses where possible.

STUDY 2

Study 1 tested the measurement-related and time-based hypotheses about team conflict using both sets of team data. This study extended the findings from Study 1 to address associations and causal relationships between team conflict and other measures. The practical and theoretical benefits of this study are extensive. Research results from these analyses show whether the changing conflict levels of many deadline-driven project teams are due to members' and teams' demographic characteristics and personality traits. These results also contribute to the body of research on whether team conflict types influence team performance. In this study, I explored two sources of team performance measures: results using these team- and other-rated performance scores will inform researchers on the influence of team conflict to members' perceptions of how effective the team is and to external supervisors' perceptions of the team's final product.

Methods

Participants and Procedure

I accessed questionnaire and project grade data from an archive of data collected from the members of 117 student project teams. These team members were enrolled in an 8-month engineering design course at a large Canadian university in the 2014-2015 academic year. This engineering design course consisted of multiple design projects completed sequentially that contributed to the teams' final grades. Each of the 492 team members belonged to one three- to six-member team ($M = 4.37$, $SD = 0.57$). Of the 492 team members, 382 were men, 87 were women, and 33 did not respond or were missing from Survey 1. I created a gender representation score for each team by calculating the percentage of women on each team. On average, this was 18.4% with a median of 20% and a standard deviation of 19.7%. Fifty-one (43.6%) of the 117 teams had no women on them; no teams had only women.

The team members' average age was 18.7 years with a standard deviation of 1.4. Most team members (i.e., 361) were native English speakers, with 106 learning English as a second language and 25 missing responses. At the team level, I created a native language representation score to reflect the percentage of team members for whom English is their second language. On average, 38% of team members had English as a second language; the median was 40% and the standard deviation was 23.1%. Only one team had all English as a second language speakers, whereas 16 teams had no English as a second language speakers.

Team members had seven options to indicate their ethnicity. Across teams, 49 members selected Arabic or Indian as their primary ethnicity, 75 selected East Asian, 8 selected Black, two selected Native American, 17 selected Southeast Asian, 289 selected White, and 27 selected Other, which includes multiracial individuals. Twenty-five team members either did not respond to this ethnicity question or were not present for the first wave of data collection that contained demographic questions. Although it would be more detailed and accurate to calculate team representation scores for all ethnicity options above, the categorical nature of these variables along with the sample size required a simpler approach. To do this, I categorized team members into white and non-white ethnicity groups, collapsing the seven ethnicity options into two. The median team had one-quarter (25%) non-white members and three-quarters (75%) white members; the average for teams was slightly lower at 22.6% (standard deviation = 19.5%). Whereas there were no teams with all non-white members, there were 38 teams with all white members.

Measures

All three surveys included the same conflict measures as Study 1. In Survey 1, team members responded to 60 items measuring their personality with the HEXACO personality inventory (Ashton & Lee, 2009; Appendix A). The response scales for all personality items were

Likert-type scales from 1 = Strongly Disagree to 5 = Strongly Agree. Team members also responded to questions about their native English-speaking ability, their gender, age, and race or ethnicity in Survey 1. Team effectiveness was rated by each team member on five items with a Likert-type scale from 1 = Strongly Disagree to 7 = Strongly Agree. This measure, developed by the TeamWork Lab, reflects team members' perceptions of their team's performance. An example item reads, "Compared to other teams in [course name], how would you rate your team's… overall performance?" Projects were evaluated by the course instructors and teaching assistants; teams received a grade for each project and a final course grade (Appendix B). Individual grades were adjusted from the team grade based on team members' reports of how much effort each team member contributed to the project.

Statistical Analyses

Having established measurement invariance and identified subgroups of team conflict trajectories, I then built a series of path models that tested personality and demographic variables as inputs, conflict intercepts and slopes as process variables, and measures of team performance as outcomes. These models began with a multilevel, multiple regression analysis using task, process, and relationship conflict as predictors and team performance as the dependent variable (Figure 7). This analysis disregarded time as a factor in team conflict scores and simply tested the relation between conflict scores at all time points and performance at the third time point. Next, I created a multilevel model with task and relationship conflict intercepts at the individual team member level and random slopes for these two conflict measures at the team level. This added conflict trajectories to the team level, which were not present in the first model.

For this and future models, the conflict intercepts are not available at the team level, as the conflict slopes were calculated by creating a random slope for each team through plotting conflict scores on the time variable (see Appendix C for the full model MPlus syntax). This multilevel, time-dependent approach is similar to longitudinal growth modeling, with two main differences: 1) the residual error values of each variable are constrained to equality in this model whereas they are not in the traditional LGM approach and 2) the intercept is not available at the team level in this model whereas it is calculated in the LGM approach. As the intercept is not available in this multilevel approach, the first (i.e., intercept-only) model is required to test hypotheses related to conflict intercepts. The second model used these conflict trajectories to test whether the slope of task or relationship conflict is related to team performance. These trajectories use data from three surveys; the team-rated performance measure was collected at the same time as the final team conflict measures.

Whereas this model represented the key research question of this dissertation, it did not test all Study 2 hypotheses. The third model added the first set of inputs into the input-process-outcome model. In the third model, I tested whether team members' demographic characteristics predicted conflict ratings at the individual level and whether team-level demographic representation variables predict conflict slopes at the team level. These demographic characteristics were measured in a survey to team members conducted immediately after researchers formed the teams and conducted a team training session. This provides some temporal distance between the team member inputs and all later team interactions. The demographic and personality characteristics, however, should remain stable through the entire lifecycle of the project teams and beyond. I chose to add the demographic characteristics first as input variables to test the unique contributions of each demographic (i.e., generally surface-level) team diversity measure (Phillips & Loyd, 2006) before adding the personality-based (i.e., deeper-level) measures of team member characteristics.

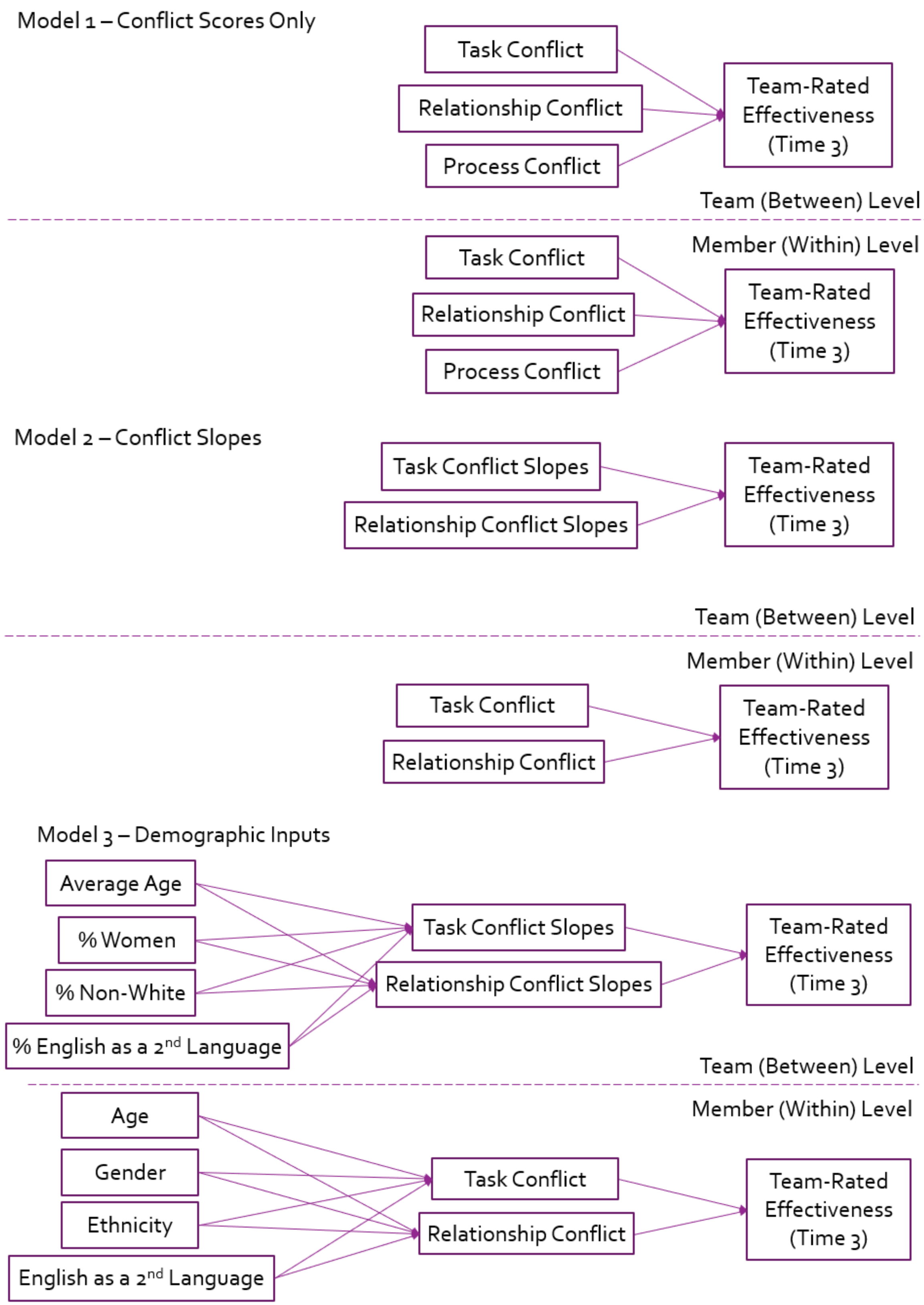

Figure 7. Path models representing Models 1-3 with team-rated effectiveness.

Whereas Model 3 reflected the input-process-output framework using demographic characteristics, it did not test direct links from inputs to outputs. Model 4 tested whether demographic variables directly affect team performance (Figure 8). The fifth model added the remaining input variables to the analysis: this tested the incremental predictive power of HEXACO personality traits on conflict at the individual and team levels. This means Model 5 tested hypotheses about how individual- and team-level personality traits affect conflict, whereas Model 6 tested the direct link between personality and performance, without any intervening process variables (Figure 9). I first analyzed this set of six models using team-rated effectiveness as the outcome measure. Next, I analyzed the same set of models in the same order, using other-rated project grades as the outcome measure for all six path analyses. Whereas the results are numerically and conceptually identical for the input-process relationships, this other performance measure may have different relations with inputs and processes.

After testing this set of path models for two measures of team performance (i.e., team-rated effectiveness and other-rated project grades), I continued to the growth mixture modeling analysis. This analysis separated the full sample of teams into groups based on conflict trajectories. Using this approach, I calculated the average team inputs and outputs for each class. I used the two-class mixture modeling results from Study 1 to compare classes on demographic characteristics, personality traits, and team outcomes. I confirmed that member-rated team outcomes (i.e., team effectiveness scores) had shared team-level variance with descriptive ICC values. Using previously identified classes, I tested whether predictor and outcome means differed across classes of teams. Where possible, Study 2 tested hypotheses at both individual and team levels. This examines whether input-process-output relations are isomorphic (i.e., the same across levels) or if they differ from the team member level to the whole team level.

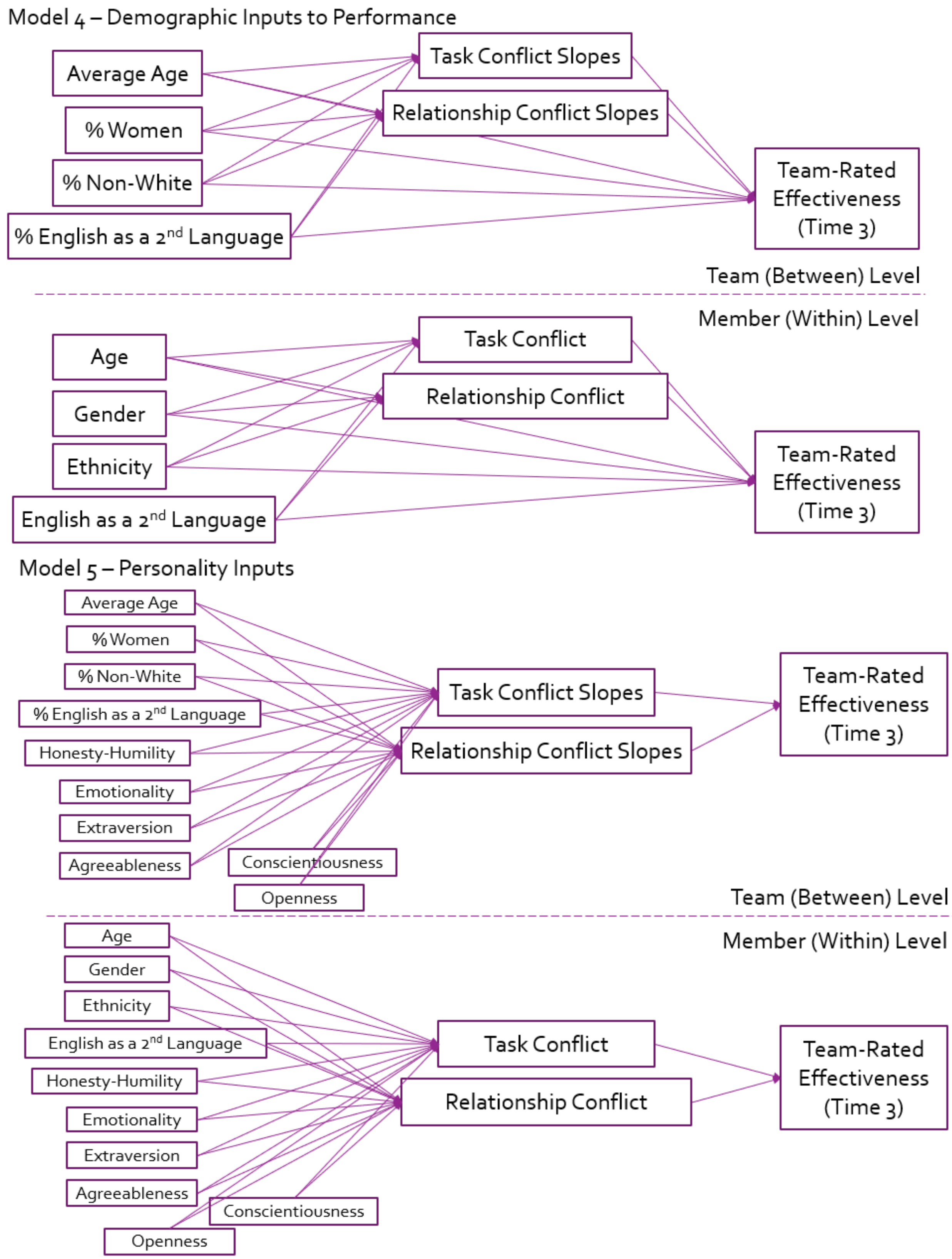

Figure 8. Path models representing Models 4 and 5 with team-rated effectiveness.

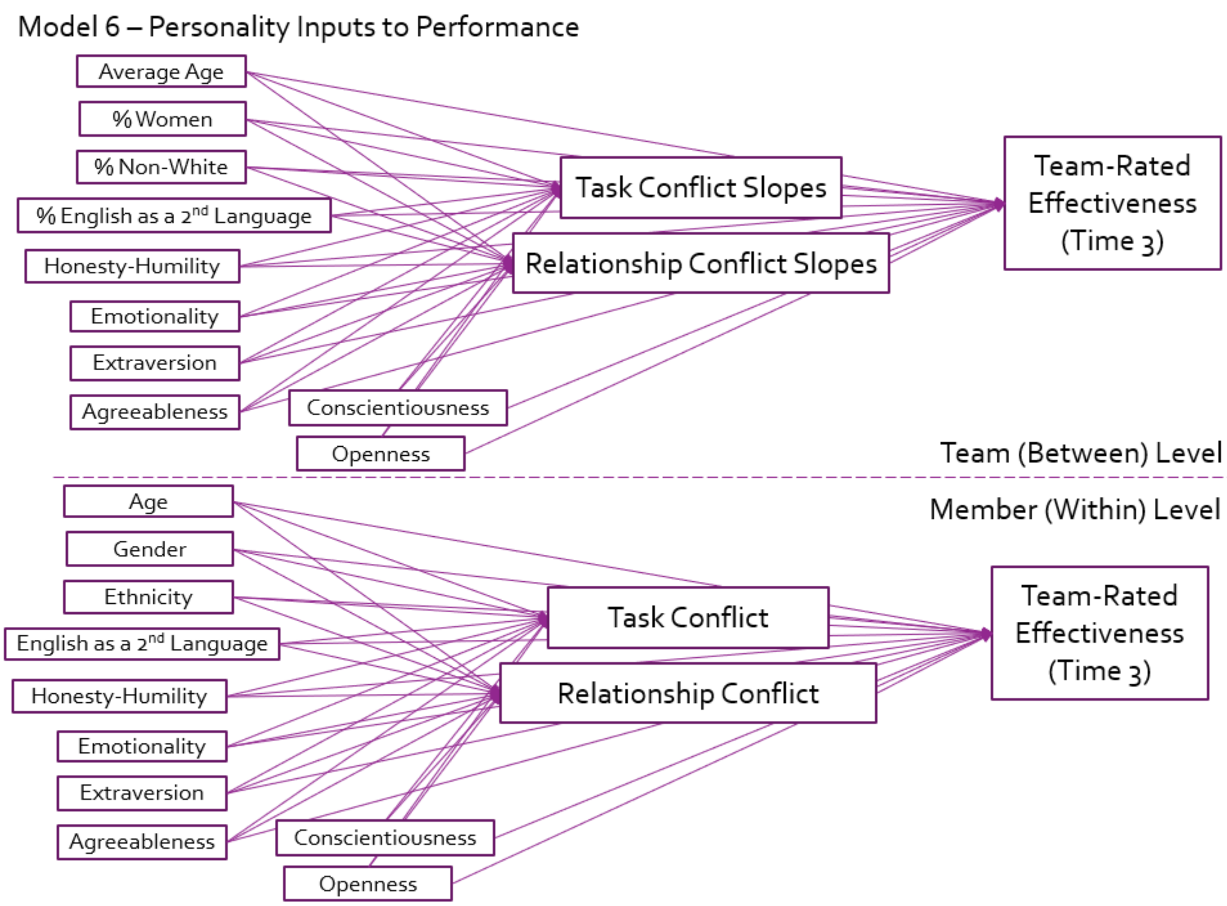

Figure 9. A path model representing Model 6 with team-rated effectiveness.

Results

Measurement Analyses

To establish acceptable psychometric properties for all survey measures, I computed their reliability coefficients (Cronbach, 1951). All variables had acceptable levels of internal consistency; team effectiveness and all survey measures of relationship conflict had the highest Cronbach's alpha values (Table 17). For all team-rated constructs, I also calculated the intraclass correlations at each survey. Intraclass correlations reflect the percentage of variance accounted for at the between-team level. For relationship and logistical conflict, these values were quite low in the first survey and were highest at the second survey. Task conflict had relatively consistent intraclass correlations over time, as the percentage of variance at the team level was higher at the start yet it did not reach the extreme values that relationship or logistical conflict did. Team effectiveness had a high intraclass correlation value in the only survey where it was measured. Correlation matrices at the individual level (Table 18) and the team level (Table 19) are reported below.

Table 17. *Interitem reliability coefficients and intraclass correlations for Study 2.*

Construct	Items	Time 1	Time 2	Time 3
Honesty-Humility	10	.74		
Emotionality	10	.78		
Extraversion	10	.80		
Agreeableness	10	.75		
Conscientiousness	10	.78		
Openness	10	.76		
Relationship Conflict	4	.85 (.045)	.94 (.38)	.89 (.27)
Task Conflict	3	.80 (.19)	.79 (.10)	.85 (.12)
Logistical Conflict	3	.86 (.047)	.80 (.23)	.79 (.10)
Team Effectiveness	5			.95 (.22)

Note. Intraclass correlations are in parentheses.

Table 18. Correlation matrix for the individual-level variables in Study 2.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1. Age																	
2. Gender	.10																
3. Ethnicity	-.10	-.05															
4. English	-.11	-.02	.57														
5. O	.11	.08	-.02	-.02													
6. C	-.07	-.07	.10	.11	.13												
7. A	.06	.15	-.06	-.06	.01	.10											
8. E	-.04	.08	.02	.08	.07	.11	.08										
9. X	-.13	-.38	-.09	-.12	-.12	.05	-.08	-.16									
10. H	.00	-.11	-.09	-.09	.02	.22	.33	-.10	.14								
11. T1_RC	.07	.02	-.05	-.14	-.02	-.19	-.04	-.05	-.02	-.06							
12. T1_TC	.04	.16	.12	.01	.14	.00	.02	.10	-.15	.00	.10						
13. T2_RC	.11	-.01	-.06	-.07	-.01	-.03	-.08	-.01	.03	-.15	.31	-.01					
14. T2_TC	-.04	.09	.02	.05	.03	.03	.02	.05	-.07	.04	-.04	.33	.17				
15. T3_RC	.03	-.06	.01	-.01	.01	-.04	-.10	-.03	.07	-.15	.34	-.01	.47	.01			
16. T3_TC	-.08	.16	.02	.01	.09	.03	.03	.06	-.04	.02	.00	.28	.06	.43	.13		
17. TE	-.07	.09	-.03	-.01	-.01	.09	.05	.15	-.08	-.01	-.20	.04	-.22	.11	-.44	.14	
18. Grades	.00	-.03	.04	.06	-.03	-.01	-.08	.06	-.03	-.05	.06	.09	.02	.01	-.01	.02	.11

Note. H = Honesty-Humility. E = Emotionality. X = Extraversion. A = Agreeableness. C = Conscientiousness. O = Openness. T1 = Time 1. T2 = Time 2. T3 = Time 3. RC = Relationship Conflict. TC = Task Conflict. TE = Team Effectiveness. Gender is coded as 0 = women and 1 = men. Ethnicity is coded as 0 = non-white and 1 = white. English as a second language is coded as 0 = English is not the member's native language and 1 = English is the member's native language. All correlations at or above +/- 0.09 are significant at $p < .05$. All correlations above +/-0.11 are significant at $p < .01$. All correlations above +/-0.14 are significant at $p < .001$.

Table 19. Correlation matrix for the team-level variables in Study 2.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1. Age																	
2. Gender	-.11																
3. Ethnicity	.10	.05															
4. English	.07	.11	.63														
5. H	-.07	-.07	.12	.07													
6. E	-.11	.18	.09	.12	.56												
7. X	.07	-.11	.09	.06	.37	.38											
8. A	.11	-.07	-.02	-.03	.60	.54	.51										
9. C	-.01	-.03	-.09	-.17	.47	.37	.32	.42									
10. O	.17	-.22	.00	.04	.44	.41	.56	.50	.51								
11. T1_RC	.13	.15	.06	.17	-.18	-.03	-.07	-.09	-.15	.08							
12. T1_TC	.06	-.16	-.18	-.06	.04	-.07	.07	.08	.08	.27	.09						
13. T2_RC	.10	.04	.01	.03	-.06	.13	.18	.06	.09	.10	.30	.01					
14. T2_TC	-.03	-.17	-.01	-.08	.06	-.06	.12	-.01	.01	.16	.03	.27	.26				
15. T3_RC	-.05	.15	.07	.06	-.02	.16	.08	.07	.12	.12	.26	-.09	.54	-.04			
16. T3_TC	-.09	-.17	.05	.01	.00	-.11	.02	-.09	.03	.19	.14	.35	.14	.53	.16		
17. TE	-.10	-.03	-.16	-.10	.00	-.12	.06	-.05	.06	.00	-.13	.09	-.35	.20	-.57	.05	
18. Grades	.02	.09	-.10	-.13	-.32	-.23	-.14	-.38	-.21	-.28	.12	.13	.04	.04	-.04	.04	.16

Note. Age = Average team member age. Gender = % of women on the team. Ethnicity = % of non-white team members. English = % of English as a second language members on the team. H = Honesty-Humility. E = Emotionality. X = Extraversion. A = Agreeableness. C = Conscientiousness. O = Openness. T1 = Time 1. T2 = Time 2. T3 = Time 3. RC = Relationship Conflict. TC = Task Conflict. LC = Logistical Conflict. TE = Team Effectiveness. All correlations above +/- 0.19 are significant at $p < .05$. All correlations above +/-0.24 are significant at $p < .01$. All correlations above +/-0.30 are significant at $p < .001$.

Multilevel Modeling

Predicting team-rated performance. Study 1 tested the measurement-related hypotheses in the present body of research. In this section of Study 2, I tested the substantive hypotheses that connect team inputs, conflict, and performance. To test Hypotheses 6 to 11, I evaluated a series of six multilevel models. The first and simplest model sought to answer the question, do conflict scores predict team effectiveness? The final and most elaborate model sought to answer the question, do team inputs, specifically demographic characteristics and personality traits, predict conflict scores, conflict slopes and team effectiveness? Below, I describe the purpose, hypotheses, and results of each model.

In the first model, I examined the unique predictive power of three team conflict type scores on team effectiveness (Table 20). At the within-team or individual level, scores on all three types of team members' conflict were related to their ratings of team effectiveness. Members' relationship conflict was negatively related to their ratings of team effectiveness (b = -0.14, SE = 0.05, $p = .008$), as was logistical conflict (b = -0.16, SE = 0.04, $p < .001$), whereas task conflict was positively related to team effectiveness (b = 0.12, SE = 0.04, $p = .006$). This means all three conflict types, measured through team members' lifecycles, predicted team member-rated performance at the teams' endpoint. Whereas relationship and logistical conflict were negatively related to performance at the individual level, task conflict was beneficial for team members' ratings of their effectiveness. The magnitude of each effect was small and similar across conflict types.

These results may not replicate at the team-aggregate level. Only logistical conflict and task conflict scores were related to team effectiveness at the between-team level. Logistical conflict was negatively related to team effectiveness (b = -1.27, SE = 0.40, $p = .001$), whereas task conflict had a positive connection with effectiveness (b = 0.51, SE = 0.22, $p = .02$). Here, the impact of logistical conflict on effectiveness was much stronger than the impact of task conflict on team-level performance. Relationship conflict was not significantly related to effectiveness across teams. By and large, Hypothesis 6 was supported by these data. Logistical and relationship conflict had high intercorrelations in Study 1, especially at the team level. The null results for relationship conflict at the team level in this multiple regression may come from the high overlap between predictors (i.e., multicollinearity) at the between-team level. Thus, I removed logistical conflict from the following analyses to reduce the multicollinearity of conflict predictors.

The first model answered research questions about the predictive power of static conflict measures. This is because all time points were combined to predict team effectiveness. However, the major contribution of this research is to study the dynamic trajectory of conflict. In the second model, I added random slopes for both relationship and task conflict at the team level to test the relation between conflict trajectories and team effectiveness. At the within-team level, task and relationship conflict predicted team outcomes as shown in the first model. The longitudinal portion of this model, specifically the slopes of task and relationship conflict, were reported at the between-team level. As the multilevel random-slope approach that I used did not calculate team-level intercepts, I could not test this hypothesis for starting conflict values. I was only able to test relations between team-level conflict slope and team effectiveness. Relationship conflict had an average positive slope of 0.17 ($p < .001$) and a significant variance of 0.049 ($p = .001$). Task conflict did not have a significant slope that differed from zero, at 0.034 ($p = .28$); there was no significant variance around this slope level for teams, as the variance was 0.009 ($p = .35$). At the between-team level, the slope of teams' relationship conflict scores over time was negatively related to their team-rated effectiveness (b = -2.43, SE = 0.74, $p = .001$). However, the relation between conflict slopes and intercepts could not be calculated using the multilevel random slope approach in these models. Task conflict slopes were not related to team outcomes (b = 2.77, SE = 3.03, p = .36). These two results support Hypothesis 7 that slopes of relationship conflict, but not task conflict, relate to team outcomes. This replicates the time-based analyses in Study 1 that showed task conflict did not change significantly over time, yet relationship conflict did. These two models show that task and relationship conflict scores are related to team member ratings of performance, yet only relationship conflict's slope is predictive of team effectiveness across groups.

My second model incorporated the teams' slopes of conflict in addition to the individual-level scores. Then I sought to answer questions about how member and team inputs related to conflict. Models 3 and 4 include demographic characteristics of team members and aggregated demographics at the team level. In Model 3, I tested the impact of demographics on conflict. At the within-team level, none of the four demographic characteristics I measured (i.e., age, gender, ethnicity, or English as a first language) were related to relationship conflict. Only gender was uniquely related to team members' task conflict scores (b = 0.23, SE = 0.10, p = .03), such that men reported more task conflict than women did. Age, ethnicity, and English as a native language were characteristics of members not related to team members' reports of task conflict.

At the between-team level, only teams' ethnicity composition, measured here as the proportion of team members with a non-white ethnic background, was related to relationship conflict. The negative relation between ethnicity composition and the slope of relationship conflict suggests that teams with more non-white members experienced higher relationship conflict over time than did teams who were homogeneously Caucasian (b = -0.44, SE = 0.22, p = .04). However, the multiple comparisons in this model can increase the type I error rate. For a minimally significant result such as this, future research replicating this effect is important to establish whether it is an error or a true result. As for task conflict, only teams' average age was related to the slope of task conflict; teams with a younger average age were more likely to experience increased task conflict over time (b = -0.07, SE = 0.03, $p = .04$). This result is also unlikely to pass a more stringent multiple comparison test, suggesting it may not replicate in future studies. This means Hypothesis 8 is largely unsupported, with some exceptions in the minimally significant results above. These relations should be interpreted with caution, as the sample size is relatively small at the team level. The weak significance level suggests these results may not be robust with multiple comparison adjustments. Finally, the composition of this sample skews very young with a majority of white or Caucasian team members. Teams and organizations with a different overall demographic composition may not find the same results for gender, ethnicity, and age.

In Model 4, I tested the relation between demographic characteristics and team effectiveness. At the within-team level, only team member ethnicity was related to ratings of team performance. Specifically, white team members rated their team's performance more poorly than did non-white team members (b = -0.27, SE = 0.11, $p = .02$). At the between-team level, no demographic variables were directly related to team effectiveness, failing to provide substantial support for Hypothesis 9.

Models 5 and 6 involved personality measures and demographic variables as potential predictors of conflict and performance. In Model 5, I tested the relations between HEXACO personality traits and conflict types. At the within-team level, honesty-humility was negatively related to relationship conflict scores (b = -0.12, SE = 0.05, $p = .017$), as was agreeableness (b = -0.09, SE = 0.04, $p = .032$). These results indicate that team members with higher honesty-humility or higher agreeableness reported lower relationship conflict on the team than other members. No other personality traits or demographic characteristics were related to team member relationship conflict scores.

Of the demographic and personality variables examined at the within-team level, gender was uniquely related to task conflict (b = 0.25, SE = 0.097, p = .01), as was honesty-humility (b = 0.099, SE = 0.048, p = .041). However, the latter result is quite close to the significance cutoff used in this study. Due to the high number of hypotheses tested at once in this model, this result may be a type I error and may not replicate in future research. These results show that men reported more task conflict in the team than women team members did and that team members with higher honesty-humility scores had slightly higher task conflict. At the between-team level, aggregated extraversion (b = 0.21, SE = 0.084, p = .015) was related to task conflict slopes. These results mean that teams with higher average extraversion had more positive task conflict slopes. However, previous models find that task conflict has neither a significant slope nor does that slope have significant variance across teams. Any significant results that relate to team-level task conflict slopes may reflect type I errors or spurious correlations due to the high number of inputs in this model. These results show minimal support for Hypothesis 10 that aims to link personality traits to conflict.

Distinct from Model 5, Model 6 involves adding paths between demographics and personality as input variables and team effectiveness as the output. Within teams, only team members' ethnicity was a significant and unique predictor of team effectiveness ratings (b = -0.29, SE = 0.11, p = .011). This means non-white team members rated their team's effectiveness more positively than white team members did. At the team level, average extraversion levels were uniquely related to team effectiveness (b = 1.02, SE = 0.41, p = .011), as were average conscientiousness levels (b = 0.93, SE = 0.47, p = .049). These results, similar to the previous ethnicity – team effectiveness results in the study above, may also be an artefact of the many comparisons made in this model. If replicated, this extraversion and team effectiveness relation or the conscientiousness – performance result may not hold..

There were no other relationships between team inputs and conflict or effectiveness. This is a departure from the earlier models, in which the average team age and the percentage of non-white team members were related to team-level conflict. In Model 5, extraversion was also related to task conflict slopes, yet it is not a significant predictor in Model 6. The changing results from one model to another support the multiple comparisons issue, in which previously significant results may reflect high type I error rates. Alternatively, the addition of personality traits as predictors may have reduced the unique relationship between the demographic characteristics and conflict scores. This does not provide strong support for Hypothesis 11 at the individual level, though it supports the limited indirect approach: team inputs, if they affect team performance, appear to operate through team processes such as conflict. At the team level, Hypothesis 11 is supported as two of the six personality traits were significant, unique predictors of team effectiveness. A summary of statistically significant results from the final model is provided below (Figure 10).

Table 20. *Regressions including team-rated performance with multilevel modeling for Study 2.*

Variable	Level	TC	RC	Team Effectiveness
Model 1				
TC Scores	W			b = 0.12** [0.034, 0.20]
RC Scores	W			b = -0.14** [-0.24, -0.035]
LC Scores	W			b = -0.12*** [-0.2, -0.033]
TC Scores	B			b = 0.51* [0.083, 0.94]
RC Scores	B			b = -0.36 [-0.91, 0.19]
LC Scores	B			b = -1.27** [-2.05, -0.49]
Model 2				
TC Scores	W			b = 0.10* [0.018, 0.18]
RC Scores	W			b = -0.21*** [-0.31, -0.12]
TC Slope	B			b = 2.77 [-3.17, 8.70]
RC Slope	B			b = -2.43** [-3.89, -0.97]
Model 3				
Gender	W	b = 0.23* [-0.43, -0.027]	b = -0.01 [-0.18, 0.16]	
Age	W	b = -0.019 [-0.058, 0.02]	b = 0.043 [-0.051, 0.14]	
Ethnicity	W	b = 0.13 [-0.028, 0.29]	b = 0.094 [-0.045, 0.23]	
English	W	b = -0.075 [-0.27, 0.12]	b = -0.081 [-0.11, 0.27]	
Gender	B	b = 0.24 [-0.93, 0.46]	b = 0.078 [-0.49, 0.34]	
Age	B	b = -0.065* [-0.13, -0.002]	b = -0.028 [-0.16, 0.11]	
Ethnicity	B	b = -0.23 [-0.82, 0.37]	b = -0.44* [-0.87, -0.012]	
English	B	b = 0.13 [-0.35, 0.6]	b = 0.13 [-0.43, 0.68]	
Model 4				
Gender	W			b = -0.25 [-0.52, 0.02]
Age	W			b = -0.057 [-0.74, 0.63]
Ethnicity	W			b = -0.27* [-0.49, -0.051]
English	W			b = -0.014 [-0.28, 0.26]
Gender	B			b = 1.46 [-0.21, 3.14]
Age	B			b = -0.015 [-0.32, 0.29]
Ethnicity	B			b = 1.49 [-0.84, 3.81]
English	B			b = -0.41 [-2.53, 1.71]

Table 20. *(continued).*

Model 5				
H	W	b = 0.099* [0.005, 0.19]	b = -0.12* [-0.22, -0.021]	
E	W	b = -0.014 [-0.13, 0.11]	b = 0.026 [-0.082, 0.13]	
X	W	b = 0.053 [-0.051, 0.16]	b = -0.038 [-0.13, 0.055]	
A	W	b = -0.019 [-0.13, 0.091]	b = -0.085* [-0.16, -0.009]	
C	W	b = 0.023 [-0.10, 0.15]	b = -0.094 [-0.23, 0.041]	
O	W	b = 0.019 [-0.093, 0.13]	b = -0.042 [-0.12, 0.031]	
H	B	b = -0.14 [-0.36, 0.077]	b = -0.28 [-0.76, 0.19]	
E	B	b = 0.062 [-0.054, 0.18]	b = 0.067 [-0.21, 0.34]	
X	B	b = 0.21* [0.04, 0.37]	b = 0.078 [-0.31, 0.46]	
A	B	b = -0.037 [-0.20, 0.12]	b = -0.12 [-0.44, 0.20]	
C	B	b = 0.18 [-0.001, 0.35]	b = 0.092 [-0.32, 0.5]	
O	B	b = 0.033 [-0.089, 0.16]	b = 0.098 [-0.20, 0.40]	
Model 6				
H	W			b = -0.055 [-0.23, 0.12]
E	W			b = -0.099 [-0.28, 0.077]
X	W			b = 0.16 [-0.02, 0.34]
A	W			b = -0.011 [-0.21, 0.18]
C	W			b = 0.14 [-0.03, 0.30]
O	W			b = -0.096 [-0.25, 0.057]
H	B			b = -0.55 [-1.51, 0.41]
E	B			b = 0.32 [-0.17, 0.81]
X	B			b = 1.02* [0.23, 1.82]
A	B			b = 0.19 [-0.73, 1.11]
C	B			b = 0.93* [0.006, 1.85]
O	B			b = -0.11 [-0.76, 0.54]

Note. b = unstandardized regression coefficient. TC = Task Conflict, measured as a team member score at the within-team level and as a slope at the between-team level. RC = Relationship Conflict, measured as a team member score at the within-team level and as a slope at the between-team level. W = Within-Team (i.e., Individual) level. B = Between-Team (i.e., Team) level. Square brackets contain values representing 95% confidence intervals; intervals that contain zero are considered non-significant. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

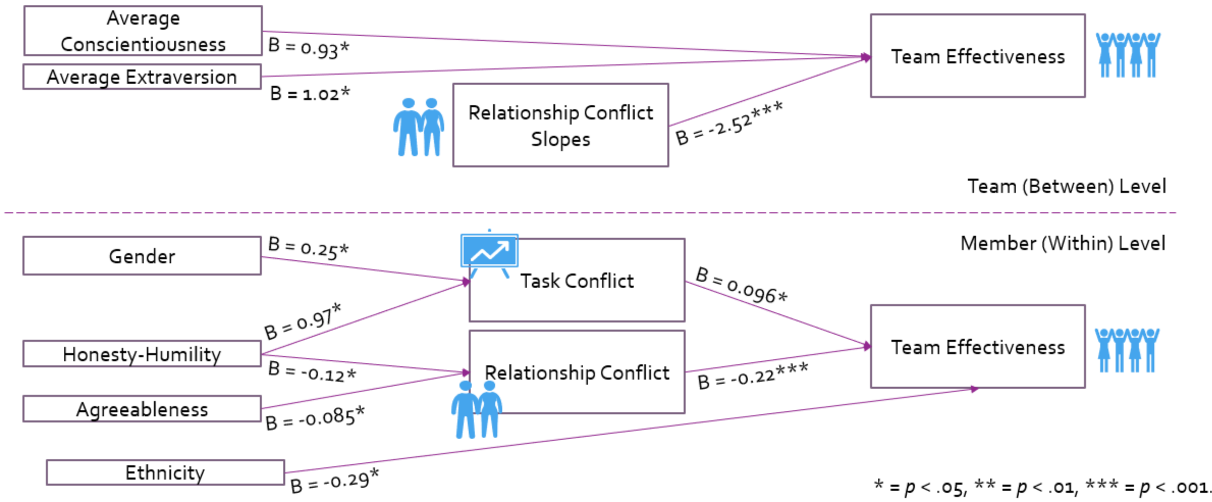

Figure 10. Summary of statistically significant results from Model 6.

Predicting other-rated performance (team project grades). Whereas the results of the analyses above have interesting implications, all the variables in the six models were measured using the same method: team member reports. In the following analyses, team performance was rated, using grades out of 100, by observers outside of the team. To begin, the first model analyzes the effect of conflict scores on grades: at the within-team level, no conflict type was related to project grades (Table 21). This pattern held at the between-team level, in which no conflict type was significantly related to team grades either, which does not support Hypothesis 6. In the remaining models, I kept relationship and task conflict and removed logistical conflict, as the overlap between relationship and logistical conflict was high (see Study 1 results).

In the second model, I computed random slopes for all teams and analyzed the impact of team slopes on team performance. The growth model values for task and relationship conflict were consistent with the team-rated performance analyses above. Relationship conflict had a significant and positive slope on average, at 0.17 ($p < .001$) whereas task conflict had no significant slope in a positive or negative direction, at 0.035 ($p = .27$). Whereas relationship conflict slopes differed across teams, as seen in the significant variance values (variance = 0.047, $p = .001$), task conflict did not have significant variance (variance = 0.009, $p = .27$). As in the first model, relationship and task conflict were not related to individual team member grades. Further, the relationship and task conflict slopes were not related to grades at the team level, failing to support Hypothesis 7. This suggests that other-rated performance measures are not closely related to team processes.

Model 3 includes demographic variables and tests the relations between these input characteristics and conflict types. At the individual level, gender was positively related to task conflict (b = 0.22, SE = 0.10, $p = .03$). This suggests men on the team reported more relationship conflict than women did and replicates the results from models reported above. Yet this gender-linked result is not highly significant. This makes the result susceptible to lower replication rates due to type I error. Age, ethnicity, and English as a second language status did not relate to individual-level task conflict scores. Relationship conflict scores at the individual level were not related to any demographic characteristics.

At the team level, only team gender composition was related to task conflict slopes (b = 0.37, SE = 0.18, $p = .037$), such that teams with more men had steeper task conflict trajectories. Yet this result is not highly significant and task conflict had limited variability with no significant slope; these two caveats reduce the likelihood of this team gender composition result remaining significant in later models. No demographic characteristics of teams were related to relationship conflict slopes, including age, gender, ethnicity, and English language status. Thus, gender was the only demographic characteristic that explained differences in team conflict within and between teams, somewhat supporting Hypothesis 8. Model 4 tests the links between demographic variables and grades. Within teams, no demographic traits were related to grades. Similarly, at the between-team level, no demographic variables were related to other-rated performance, thus finding no support for Hypothesis 9.

Model 5 examines the relation between personality characteristics and conflict scores. Beyond the impact of demographic variables, only agreeableness and honesty-humility had a significant correlation with individual-level relationship conflict scores. Team members who were more agreeable reported lower relationship conflict (b = -0.085, SE = 0.04, $p = .036$), whereas team members with higher honesty-humility also reported lower relationship conflict in the team (b = -0.12, SE = 0.05, $p = .017$). However, the high number of inputs in this model increases the risk of type I errors from significant results. This means the unique personality and demographic predictors may not be significant in a replication of this study or with multiple comparison adjustments. As for task conflict scores at the individual level, only the honesty-humility personality trait was related to task conflict (b = 0.097, SE = 0.049 p = .047). This means members with higher honesty-humility perceived slightly more task conflict, though as above they perceived slightly lower relationship conflict. This barely significant result may not hold when replicated, as it could reflect high type I error. This partially supports Hypothesis 10 which posited that personality traits predict conflict. At the between-team level, no personality traits were related to task or relationship conflict slopes.

Finally, Model 6 adds personality as direct predictors of grades. No personality traits were directly related to grades at the individual or team levels, showing no support for Hypothesis 11. The differences in these results, compared to those described above, reflect the minimal overlap between what the team-rated, and the other-rated, performance constructs are measuring.

Table 21. *Regressions including other-rated performance (grades) with multilevel modeling for Study 2.*

Variable	Level	TC	RC	Team Outcomes (Grades)
Model 1				
TC Scores	W			b = -0.023 [-0.064, 0.018]
RC Scores	W			b = -0.019 [-0.088, 0.05]
LC Scores	W			b = 0.006 [-0.067, 0.079]
TC Scores	B			b = 2.68 [-1.69, 7.04]
RC Scores	B			b = 1.94 [-3.76, 7.64]
LC Scores	B			b = -2.51 [-10.86, 5.84]
Model 2				
TC Slope	B			b = 2.90 [-29.81, 35.61]
RC Slope	B			b = -2.39 [-11.49, 6.72]
Model 3				
Gender	W	b = 0.22* [0.024, 0.42]	b = -0.005 [-0.18, 0.17]	
Age	W	b = -0.019 [-0.058, 0.02]	b = 0.044 [-0.05, 0.14]	
Ethnicity	W	b = 0.13 [-0.028, 0.29]	b = 0.089 [-0.05, 0.23]	
English	W	b = -0.073 [-0.27, 0.11]	b = -0.08 [-0.23, 0.11]	
Gender	B	b = 0.37* [0.023, 0.71]	b = -0.088 [-0.51, 0.33]	
Age	B	b = -0.05 [-0.12, 0.017]	b = -0.053 [-0.16, 0.053]	
Ethnicity	B	b = -0.30 [-0.80, 0.12]	b = -0.26 [-0.71, 0.20]	
English	B	b = 0.081 [-0.63, 0.79]	b = 0.046 [-0.57, 0.66]	
Model 4				
Gender	W			b = 0.12 [-0.076, 0.023]
Age	W			b = -0.023 [-0.45, 0.21]
Ethnicity	W			b = 0.041 [-0.21, 0.29]
English	W			b = 0.19 [-0.18, 0.57]
Gender	B			b = 87.18 [-172.35, 346.71]
Age	B			b = 16.53 [-141.69, 174.39]
Ethnicity	B			b = -11.11 [-824.03, 801.80]
English	B			b = 156.03 [-544.31, 856.36]

Table 21. *(continued).*

Model 5				
H	W	b = 0.097* [0.001, 0.19]	b = -0.12* [-0.22, -0.022]	
E	W	b = -0.015 [-0.13, 0.10]	b = 0.027 [-0.081, 0.14]	
X	W	b = 0.055 [-0.051, 0.16]	b = -0.035 [-0.13, 0.057]	
A	W	b = -0.018 [-0.13, 0.092]	b = -0.084* [-0.16, -0.006]	
C	W	b = 0.026 [-0.099, 0.15]	b = -0.094 [-0.23, 0.041]	
O	W	b = 0.018 [-0.094, 0.13]	b = -0.042 [-0.12, 0.031]	
H	B	b = -0.073 [-1.10, 0.96]	b = -0.27 [-0.75, 0.21]	
E	B	b = 0.015 [-0.28, 0.31]	b = 0.072 [-0.19, 0.33]	
X	B	b = 0.035 [-0.60, 0.67]	b = 0.098 [-0.26, 0.46]	
A	B	b = -0.11 [-1.22, 1.01]	b = -0.06 [-0.41, 0.29]	
C	B	b = 0.012 [-0.34, 0.37]	b = 0.13 [-0.25, 0.51]	
O	B	b = 0.13 [-0.16, 0.42]	b = 0.089 [-0.21, 0.38]	
Model 6				
H	W			b = -0.008 [-0.22, 0.20]
E	W			b = 0.003 [-0.13, 0.14]
X	W			b = -0.052 [-0.16, 0.056]
A	W			b = -0.078 [-0.35, 0.19]
C	W			b = 0.029 [-0.16, 0.22]
O	W			b = -0.038 [-0.27, 0.20]
H	B			b = -7.77 [-23.33, 7.78]
E	B			b = -0.70 [-7.54, 6.13]
X	B			b = 4.84 [-4.78, 14.46]
A	B			b = -7.71 [-23.19, 7.78]
C	B			b = -0.20 [-9.06, 8.66]
O	B			b = -0.16 [-17.89, 17.57]

Note. b = unstandardized regression coefficient. TC = Task Conflict, measured as a team member score at the within-team level and as a slope at the between-team level. RC = Relationship Conflict, measured as a team member score at the within-team level and as a slope at the between-team level. W = Within-Team (i.e., Individual) level. B = Between-Team (i.e., Team) level. Square brackets contain values representing 95% confidence intervals; intervals that contain zero are considered non-significant. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Growth Mixture Modeling at the Team Level

To understand the unique conflict trajectories that teams may follow, I conducted a growth mixture modeling analysis. Following results from Study 1, I used a two-class mixture model at the team level with longitudinal growth models for two classes of team conflict variables (Table 22). In addition to establishing these classes, I compared the team input and outcome variables across these subgroups of the overall sample. Class 1, with 66 teams, had a high task conflict intercept (M = 3.34, $p < .001$) with no significant task conflict slope (M = -0.024, $p = .61$). This class had a low relationship conflict intercept (M = 1.49, $p < .001$) with a significantly positive slope (M = 0.14, $p = .001$). The intercept of relationship conflict did not have significant variance (variance = 0.09, $p = .082$), yet the relationship conflict slope had significant variance (variance = 0.089, $p < .001$). Task conflict's intercept had significant variance across classes (variance = 0.12, $p < .001$), though the task intercept slope variance had to be set to zero to complete this analysis. These variance levels apply for both classes.

Class 2, with 48 teams, had similar intercepts to Class 1; task conflict was also high on average (M = 3.1, $p < .001$) and relationship conflict started at a relatively low level (M = 1.57, $p < .001$). As in Class 1, task conflict did not show a significant slope in either direction (M = 0.07, $p = .17$). Similar to Class 1, relationship conflict in Class 2's teams had a significantly positive slope (M = 0.21, $p < .001$), yet it was steeper (Figure 11). Thus, both classes began with similar intercepts, but the relationship conflict levels in Class 2 rose more quickly than relationship conflict scores did in Class 1. Task conflict stayed at a consistently high level in both classes, suggesting there was considerable project-related debate among most teams throughout their time working together.

To understand how these classes may differ with respect to their team inputs and outcomes, I compared the demographic, personality, and team performance means for both classes. On average, Class 1 teams were composed of 16.8% women, 16.8% non-white members, 9.8% English as a second language speakers, and an average age of 18.58. Some of these results were similar to the Class 2 demographics: 20.4% of team members were women and teams had an average age of 18.75. Yet the ethnicity and language results are markedly different: in Class 2, 56.1% of team members are non-white and 38% of team members on average had English as their second language. The two classes of teams seem to reflect similar levels of HEXACO personality traits. Class 1 has similar scores for all six personality traits: honesty-humility (Class 1 M = 3.18, Class 2 M = 3.24), emotionality (Class 1 M = 2.87, Class 2 M = 3.00), extraversion, (Class 1 M = 3.45, Class 2 M = 3.42), agreeableness (Class 1 M = 3.29, Class 2 M = 3.34), conscientiousness (Class 1 M = 3.70, Class 2 M = 3.58), and openness to experience (Class 1 M = 3.23, Class 2 M = 3.19). One should interpret any differences in these descriptive results with caution.

Regarding team outcomes, Class 1 had a similar mean for instructor-rated performance (i.e., team project grades) that was only higher by approximately 1 percentage point (Class 1 M = 87.14, compared to Class 2 M = 83.71). The variance in grades was very high, at 54.25% ($p < .001$), suggesting no differences between classes on other-rated performance. However, the team effectiveness mean was considerably lower for Class 1 (M = 5.41) than for Class 2 (M = 5.07). There was significant variance in team-rated effectiveness scores overall, at 0.49 points ($p < .001$), which is larger than the magnitude of the difference between the team effectiveness scores in each class. This means there may be no significant differences between how team members perceive their performance across classes and how outside observers rate the outputs of each class of teams’ work.

Table 22. *Growth mixture modeling with two classes for Study 2.*

Class	Conflict	Demographics	HEXACO	Team Performance
# 1 (n = 66)	TC Intercept: 3.34*** (0.12***) TC Slope: -0.024 (0) RC Intercept: 1.49*** (0.09) RC Slope: 0.14*** (0.089***)	Age: 18.58 Gender: 16.8% women Race: 23.1% non-white English: 9.8% ESL	H: 3.18 E: 2.87 X: 3.45 A: 3.29 C: 3.70 O: 3.23	Team Effectiveness: 5.41 (0.49***) Grades: 87.14% (54.24***)
# 2 (n = 48)	TC Intercept: 3.01*** (0.12***) TC Slope: 0.07 (0) RC Intercept: 1.57*** (0.09) RC Slope: 0.21*** (0.089***)	Age: 18.75 Gender: 20.4% women Race: 56.1% non-white English: 38% ESL	H: 3.24 E: 3.00 X: 3.42 A: 3.34 C: 3.58 O: 3.19	Team Effectiveness: 5.07 (0.49***) Grades: 83.71% (54.24***)

Note. Variances are in parentheses. ESL = English as a second language. H = Honesty-Humility, E = Emotionality, X = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness. *** = $p < .001$.

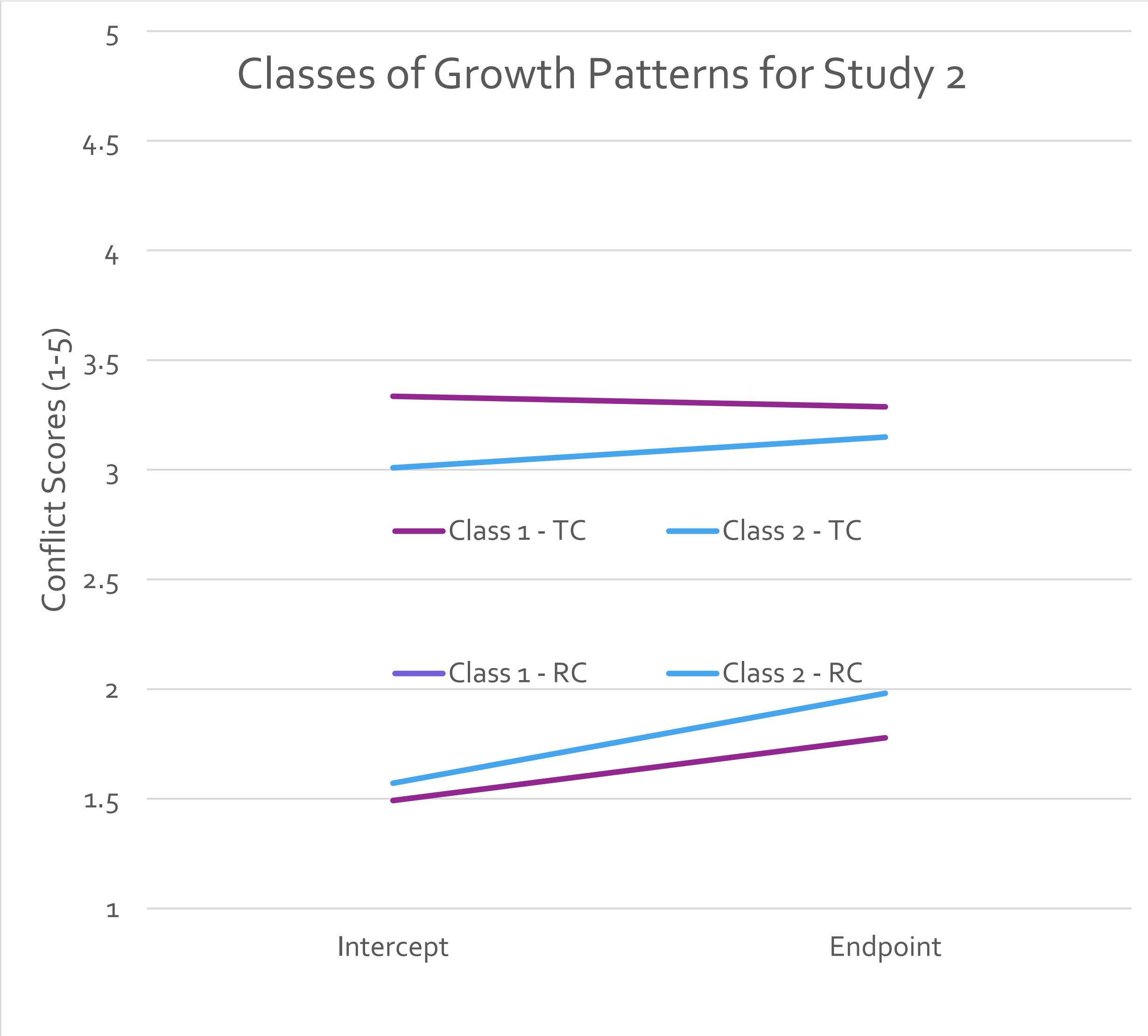

Figure 11. Growth mixture modeling results for Study 2.

Note. Relationship conflict trajectories for both classes are below the legend and task conflict trajectories for both classes are above the legend.

Discussion

The results of Study 2 demonstrate the importance of capturing the change in conflict scores over time. The importance of this research design is shown in how conflict connects team inputs and outcomes. In particular, relationship conflict levels in teams changed more over time than did task conflict levels, resulting in stronger connections between relationship conflict slopes and team-rated effectiveness. The negative relation between relationship conflict trajectories and team performance, at least when performance was rated by team members, replicates and extends meta-analytic findings suggesting that relationship conflict may be detrimental to team performance (De Wit et al., 2012; O'Neill et al., 2013). This detrimental effect of relationship conflict may be due to lower team learning behaviours, as previous research indicates that disrupted team learning explains the link between higher relationship conflict and poorer performance (van Woerkom & Van Engen, 2009).

The null relation between conflict and other-rated performance suggests that our field needs more research on the differences between these measures and significant predictors of supervisor-rated performance. Previous meta-analyses found that conflict types had different effects for team-, expert-, and supervisor-rated performance (O'Neill et al., 2013). Although other-rated performance may seem more valid than self-rated performance due to lower common method variance, there is ample evidence that team member ratings are related to many other important outcomes from the team's perspective. This includes team potency (Pearce, Gallagher, & Ensley, 2010; Sivasubramaniam, Murry, Avolio, & Jung, 2002), the team's shared belief that their group can achieve results, and team efficacy (Gully, Incalcaterra, Joshi, & Beaubien, 2002), the team's perceptions they are capable of specific tasks.

New methods used in this study can help researchers to identify classes of teams with excessively high relationship conflict. These new analytic methods include growth mixture modeling at the team level, as used in this study. The two classes of teams found here suggest that team conflict trajectories, not their initial levels of conflict, differentiate teams at risk for underperformance due to harmful conflict. Yet these results are tentative; they are based on relatively small team-level sample sizes and may not replicate across team tasks and other contexts. Future work is needed in workplaces and other settings with much larger sets of comparable teams.

Building on the mixed effects of demographic diversity on team experiences (Pelled, Eisenhardt, & Xin, 1999), the results in this study showed mixed patterns between individual-level, and team-level, diversity-team conflict relations. In this study, team members' gender was linked to their task conflict perceptions. This held across multiple analyses, suggesting that the following result is consistent in this sample: men on the team reported higher task conflict than did women on the team. These results for gender are surprising, as men reported higher task conflict than women did on teams. In one analysis, teams with a younger average age had steeper task conflict slopes; however, this result did not hold when personality traits were added to the analysis. This inconsistent result for teams' age may be due to the restriction of range on this variable. The average team member age was quite young, and its distribution was skewed towards the lower end of the range, with few team members in their mid- to late-twenties and older. Of all the demographic characteristics measured in this study, team members' age is least representative of the general population and of work teams to which this research aims to generalize.

Other research, often conducted at the team level, shows some relationships between team gender composition and team processes such as conflict, including the importance of gender identity salience (Randel, 2002). One cross-cultural study found that gender heterogeneity interacted with national culture to predict cognitive, but not affective, conflict (Watson, Cooper, Torres, & Boyd, 2008). In a study of status conflict, highly gender diverse teams had a weaker relation between status conflict and team psychological safety than less gender diverse teams (Lee, Choi, & Kim, 2018). Finally, an investigation of gender faultlines found that emotional conflict mediated the relation between activated gender faultlines and team creativity (Pearsall, Ellis, & Evans, 2008). However, none of these studies directly compare task and relationship conflict scores at the individual level.

Insights from the romantic relationships literature may help to understand how men and women express and perceive conflict. In one meta-analysis of couples, women were more likely to express hostility and distress during a relationship conflict, whereas men were more likely to express withdrawal (Woodin, 2011). Interestingly, women in romantic relationships tend to be more affected by relationship negativity due to more interdependent views of themselves (Wanic & Kulik, 2011). This would suggest the opposite relation between gender and conflict perceptions, if this effect held in the project team context. However, romantic relationship conflict may be conceptually more similar to team relationship conflict than to team task conflict, due to the personal nature of relationship conflicts. Gender differences in conflict management strategies are an unexplored area of research that may interest future team researchers. In an experimental study, women were more likely than men to choose communal conflict management strategies with friends, and agentic conflict management strategies with romantic partners (Keener, Strough, & DiDonato, 2012). Studies of conflict management among team members may find differences in how team members resolve conflicts, based on their gender.

More research is needed to understand how team members with unique demographic characteristics perceive shared group processes. Age may affect conflict resolution skills (Gbadamosi, Baghestan, & Al-Mabrouk, 2014; Owens, Daly, & Slee, 2005); this suggests teams with younger members may have more difficulty addressing and diffusing conflict, leading disagreements to escalate over time. No demographic traits showed connections to individual-level relationship conflict, yet mixed-ethnicity teams showed steeper increases in relationship conflict than teams with a larger proportion of white members. Interestingly, team members with a non-white background had higher self-rated team effectiveness scores than did other team members. As mixed-ethnicity teams had higher relationship conflict slopes, at least in an intermediate model of the analyses, this team effectiveness result may reflect the personal disagreements happening in these teams toward the end of the project teams' time together. Recent disagreements of a personal nature could create strong negative memories (Bravo-Rivera & Stores-Bayon, 2020; Small, Kenny, & Bryant, 2011) that appear when white team members are rating their performance.

Trust, a key mechanism of teamwork, may explain the diversity-conflict relations seen here. In a longitudinal study, higher team trust at the *initial stages of a team's time together* predicted lower relationship conflict later on (Curşeu, & Schruijer, 2010). Previous research has shown that team diversity is related to lower trust (Garrison, Wakefield, Xu, & Kim, 2010) and can increase conflict and reduce social integration (Stahl, Maznevski, Voigt, & Jonsen, 2010). In addition, culturally diverse teams may be more individually focused when starting to work together (Watson, Johnson, Kumar, & Critelli, 1998). Thus the generally visible (i.e., surface-level) diversity traits, such as ethnicity differences shown here, may lower initial trust levels or familiarity across ethnicity groups and increase conflict as the team progresses.

Some team member personality traits played a role in team conflict. Within teams, members reported lower relationship conflict when they were more agreeable or had higher honesty-humility. This may mean that honest, humble, and/or agreeable team members are less likely to engage in dyadic relationship conflict. Alternatively, these team members may see relationship conflict as less severe when it does occur. This may come from different conflict expression types (Weingart, Behfar, Bendersky, Todorova, & Jehn, 2015) or different styles of conflict resolution (Behfar, Peterson, Mannix, & Trochim, 2008). Between teams, the average extraversion level on teams was related to team members' ratings of team effectiveness. This may occur if teams in which members are more outspoken believe they are more effective. Descriptive results from the growth mixture modeling analysis showed substantial differences on all six HEXACO personality traits across classes. This brings up many questions, including which differences are related to conflict, whether any class differences in personality affect performance, and how stable these differences will be for analyses with larger sample sizes.

Whereas these team-level results come from one aggregation approach, specifically averaging team member personality traits, other compositional approaches may show different effects. Team member skewness is another compositional approach used in previous team conflict research (Sinha et al., 2016), that may be appropriate for these input and process scores. Nevertheless, the results of Study 2 suggest that some team inputs, including demographic and personality traits, explain differences in conflict levels and trajectories, which in turn impact team-rated performance. These findings provide fruitful avenues for future research to classify and reduce detrimental conflict trajectories for work teams.

Limitations

There are unique limitations of this study that did not characterize Study 1. Namely, the study involves weak causal inference and low variance on team members' demographic characteristics. This study was non-experimental; thus, it has weak causal inference for the relations between team inputs, processes, and outcome variables. Whereas this study design contained random assignment into groups, multiple time points for causal ordering, and multiple sources of ratings for team performance, experimental studies would strengthen these causal inferences. For example, controlled studies that manipulate team composition variables could support the relation between team inputs and dependent variables such as conflict and performance. In addition, this study's design did not allow for cross-lagged analyses or reverse causation models that may strengthen the results of this research. Future research could measure team performance more often to investigate whether the conflict-performance relation is reciprocal.

Next, the demographic and personality composition of this sample is limited in some areas that restricts its generalizability. Specifically, the age range of this sample is narrow and members are younger, which may limit the applicability of these results to older age groups. In addition, the personality profile of these young engineering trainees may not reflect the broader population. As personality may change within one's lifetime (Anusic & Schimmack, 2016) and, in some cases, across birth cohorts (Twenge, 2001), the personality-based results in this study may not hold across populations. Another way this research design limits the interpretation of team member characteristics is the unbalanced ethnicity diversity on teams in this sample. The average proportion of non-white team members was 38%; 16 all-white teams appeared in this sample, compared to one all non-white team. The imbalanced nature of this field sample limits the diversity-related insights this study can provide. Thus, Study 2 has unique limitations that may be addressed in future research with stronger causal study designs and varying participant samples.

STUDY 3

Study 2 explored the relations between demographic characteristics, personality traits, conflict types, and two measures of team outcomes. Study 3 replicated these relations at the within-team level and extended the exploration of team inputs at the between-team level. This study contributes to multiple fields: team conflict, team performance, group diversity, and personality research. The methods used here advance our understanding of dynamic conflict types unfolding over time, how researchers and practitioners can investigate multiple team member inputs in one measure, and how unique subgroups of teams may have different conflict experiences than the average group.

Methods

Participants and Procedure

I collected questionnaire and project grade data from the members of 157 student project teams enrolled in an 8-month engineering design course at a large Canadian university in the 2015-2016 academic year. This engineering design course consisted of multiple design projects completed sequentially that contributed to most of their final grade. Each of the 632 students belonged to one three- to six-member team ($M = 3.96$, $SD = 0.54$). The TeamWork Lab randomly assigned students to these teams with one restriction: students were randomly assigned within each classroom. I collected data from three surveys: one taken on the first day teams were created (i.e., Survey 1), the second approximately two months into the teams' tenure (i.e., Survey 2), and the third, collected approximately seven months after the team began working together (i.e., Survey 3). Of the 632 students, 493 identified as men and 131 identified as women; 7 did not respond to the survey containing demographic information. The average team had 20.6% women (median = 25%, SD = 21.2%), with no teams having over 80% women and 64 teams having no women members. Three-hundred and twenty-eight individuals reported their ethnicity as white or Caucasian, whereas 290 selected other ethnicity options including multi-racial: 12 individuals did not respond or were missing from this survey. Teams had 47% non-white members on average (median = 50%, SD = 27.3%), with 14 all-white teams and 13 all non-Caucasian teams. Across members, 453 individuals had English as their first language, 167 learned English as a second language, and 10 did not respond or were missing from this survey. The average team had 26.9% members with English as their second language (median = 25%, SD = 22.7%), with 45 teams having no members whose native language was not English, and no teams with all non-native English speakers. The students' average age was 18.2 years with a standard deviation of 1.3 years.

Measures

All measures were identical to those used in Study 2, with demographic information, HEXACO personality variables (Ashton & Lee, 2009), and task and relationship conflict collected at Survey 1. The same conflict variables were collected at Survey 2 and the same conflict variables, along with team effectiveness measures, were collected at Survey 3.

Statistical Analyses

In previous studies, I established measurement invariance, identified classes of intercepts and slopes, and tested the relations among conflict, personality and demographic predictors, and outcome variables. In this study, I replicated and extended the analytic approaches in Study 2 using faultline measures as team-level inputs. This answered research questions about how multiple team member traits are organized within a group, and how these traits influence conflict and performance.

To calculate team faultlines from the personality scores and demographic characteristics provided in Survey 1, I used the asw.faultlines R package (Meyer & Glenz, 2013). To my knowledge, this is the only program available for calculating team faultlines. Specifically, I used the Gibson function that takes continuous and categorical data and computes the overlap between team members' attributes. This results in a single, team-level faultline measure of the strength of rifts within a team (Gibson & Vermeulen, 2003). To find this team-level measure, the faultline program starts with the individual-level data for all four demographic characteristics or all six personality traits, plus the team membership information. After labeling each variable as a nominal (categorical) trait or a numeric (ordinal, interval, or ratio) trait, the program runs through many iterations of the faultline calculation process to find the stable value representing the multiple correlation between demographic or personality traits. When the program is finished, it returns one value between 0 and 1 for each team on its demographic faultline and its personality faultline.

This program has multiple versions of team faultlines available, based on eight distinct published papers in the faultline literature. Some measures identify subgroup membership at the individual level, whereas others identify multiple subgroups within a team. Other faultline conceptualizations include methods that calculate one faultline perception score for each group member or methods that assume each team has two homogeneous subgroups present. Yet another method finds the distance and strength of faultlines between group members. However, these other methods do not provide a single, team-level score that represents the strength of rifts on the team for variables that may be intercorrelated as personality and demographics are. I found these faultline values from HEXACO personality scores (Ashton & Lee, 2009) and demographic information (i.e., gender, age, ethnicity, and English as a first language) to calculate two scores for each team: the former reflects a 'personality faultline score' and the latter reflects a ‘demographic faultline score’. Using growth mixture modeling, I tested whether these scores predict conflict slopes and intercepts at the team level.

To build on this model, I added team performance as an outcome variable predicted by conflict slopes and intercepts at the team level. This replicated the analysis conducted in Study 2. By testing the direct relation between team faultline scores and team performance, I extended our understanding of how team traits and their configurations can influence performance. At the individual level, I replicated the analysis conducted in Study 2 by analyzing personality traits and demographic variables independently in each team member.

Results

Measurement and Descriptive Analyses

To establish acceptable psychometric properties for all survey measures, I computed their reliability coefficients (Cronbach, 1951) and intraclass correlations (Table 23). All six personality traits had moderate internal reliability, whereas relationship conflict and team effectiveness had consistently high internal reliability. Task conflict scores had low internal consistency (i.e., Cronbach's alpha scores) in earlier survey administrations; Cronbach's alpha scores were higher as teams progressed in their projects. As task conflict was measured with only three items, these inconsistent Cronbach's alpha values may reflect the small number of scale items. Intraclass correlations were higher for relationship conflict rather than task conflict, except for the first survey in which the percentage of variance at the team level was the same for both types of conflict. Team effectiveness had an acceptable intraclass correlation value, suggesting that team members largely agreed on their team's performance. In addition, I determined the intercorrelations between all variables at the individual (Table 24) and team levels (Table 25).

Table 23. *Interitem reliability scores and intraclass correlations for Study 3.*

Construct	Items	Time 1	Time 2	Time 3
Honesty-Humility	10	.76		
Emotionality	10	.77		
Extraversion	10	.79		
Agreeableness	10	.76		
Conscientiousness	10	.77		
Openness	10	.73		
Relationship Conflict	4	.86 (.099)	.91 (.31)	.93 (.27)
Task Conflict	3	.67 (.11)	.79 (.033)	.84 (.068)
Team Effectiveness	5			.91 (.20)

Note. Intraclass correlations are in parentheses for team constructs.

Table 24. *Individual-level intercorrelations for Study 3.*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Gender																
2. Age	-08															
3. Ethnicity	-01	.09														
4. English	-02	.24	.51													
5. H	-.13	.02	.11	.05												
6. E	-.39	.01	.11	.08	.11											
7. X	.01	-.10	-.10	-.11	-.03	-.14										
8. A	.04	-.01	.06	.03	.35	.01	.17									
9. C	-.19	.02	-.14	-.10	.15	.09	.30	.09								
10. O	-.03	.06	.03	.04	.13	-.01	.07	.00	.04							
11. T1_RC	.01	.09	.12	.21	-.08	.10	-.05	-.07	-.03	.04						
12. T1_TC	.11	.02	-.02	-.01	-.09	.01	.07	-.05	.01	.02	-.01					
13. T2_RC	-.01	.05	.07	.12	-.15	.16	-.11	-.18	-.01	-.04	.35	.09				
14. T2_TC	.11	-.04	.00	-.02	.08	-.08	.20	.13	.13	.11	-.10	.25	-.18			
15. T3_RC	-.07	-.01	-.01	.00	-.14	.10	-.02	-.11	-.03	.03	.30	.09	.54	-.09		
16. T3_TC	.06	-.01	-.07	-.04	-.01	-.02	.12	-.05	.09	.10	.02	.22	.00	.38	.07	
17. TE	.06	.09	.09	.11	.02	.01	.12	.01	-.01	.06	-.06	.07	-.20	.17	-.29	.16

Note. H = Honesty-Humility. E = Emotionality. X = Extraversion. A = Agreeableness. C = Conscientiousness. O = Openness. T1 = Time 1. T2 = Time 2. T3 = Time 3. RC = Relationship Conflict. TC = Task Conflict. TE = Team Effectiveness. For the gender measure, 0 = women and 1 = men. For ethnicity, this variable is coded as 0 = white and 1 = non-white. For English as a second language, 1 is coded as English is the member's native language and 2 is coded as English is not the member's native language. Correlations at or above r = +/-.08 are significant at p < .05. Correlations at or above r = +/-.11 are significant at p < .01. Correlations above r = +/-.13 are significant at p < .001.

Table 25. *Team-level intercorrelations for Study 3.*

	1.	2.	3.	4.	5.	6.	7.	8.
1. Demographic Faultlines								
2. Personality Faultlines	-.04							
3. T1_RC	.01	.12						
4. T1_TC	-.15	-.03	-.04					
5. T2_RC	.02	.16	.48	.07				
6. T2_TC	.04	-.04	-.18	.23	-.26			
7. T3_RC	-.05	.24	.33	.10	.64	-.15		
8. T3_TC	-.02	.00	-.04	.27	.04	.38	.13	
9. TE	.22	-.24	-.13	.04	-.19	.21	-.31	.16

Note. T1 = Time 1. T2 = Time 2. T3 = Time 3. RC = Relationship Conflict. TC = Task Conflict. TE = Team Effectiveness. Correlations at or above r = +/-.16 are significant at $p < .05$. Correlations at or above r = +/-.21 are significant at $p < .01$. Correlations at or above r = +/-.26 are significant at $p < .001$.

Multilevel Modeling

To test Hypotheses 6 to 15, I conducted a series of multilevel models similar to the analytic method used in Study 2. The first model started by testing whether relationship and task conflict scores, measured across three surveys, predicted team effectiveness, measured at the final of three surveys, at the within-team and between-team levels (Table 26). This step was necessary to test hypotheses about static (i.e., not time-related) effects of conflict on team performance at both levels. Within teams, member ratings of relationship conflict, regardless of the time when they were measured, negatively related to team effectiveness at the final survey (b = -0.15, SE = 0.04, $p < .001$) whereas member ratings of task conflict positively related to team effectiveness (b = 0.11, SE = 0.03, $p = .005$). Between teams, the pattern of results was identical; higher relationship conflict scores negatively related to team effectiveness (b = -0.67, SE = 0.17, $p < .001$) whereas aggregated task conflict positively related to team effectiveness (b = 0.86, SE = 0.31, $p = .006$). This means more relationship conflict was associated with poorer team-rated performance and more task conflict was associated with better team-rated performance. These results provide support for Hypothesis 6 and replicate results from Study 2.

For Model 2, I computed random slopes for each team, on both conflict types, to investigate whether a steeper conflict slope was related to differences in team effectiveness. Following the multilevel approached used in Study 2, I calculated random slopes for each time by regressing conflict scores on the within-team time variable. This approach does not provide team-level intercept scores as a traditional longitudinal growth model would. Because of this, only conflict slopes were available at the team level and only conflict scores were available at the individual member level. The slopes of task conflict (slope = 0.12, $p < .001$) and relationship conflict (slope = 0.26, $p < .001$) were significantly higher than zero. However, task conflict slopes did not have significant variance at the team level (variance = 0.005, $p = .37$) whereas relationship conflict slopes did have significant team-level variance (variance = 0.064, $p < .001$). This means task conflict slopes did not change significantly across team members, whereas relationship conflict slopes were considerably different from one team to another. This replicates the results from Studies 1 and 2, in which relationship conflict differed across classes of teams and through significant variance, leading to relationships between relationship conflict slopes and team performance.

At the within-team level, the results for Model 2 were identical to the first model: members' relationship conflict levels were negatively related to team outcomes and members' task conflict was positively related to team outcomes. Yet the pattern of relations for team conflict slope was not the same. Specifically, relationship conflict slopes negatively predicted team effectiveness (b = -1.09, SE = 0.32, $p < .001$), whereas task conflict slopes were unrelated to team effectiveness (b = 8.48, SE = 8.85, $p = .34$). Study 1 may explain these results; whereas relationship conflict scores changed over time, leading to a significant slope for relationship conflict, this did not hold for task conflict. I found that time did not explain task conflict scores in Study 1 and the largest class of team conflict trajectories showed no significant slope for task conflict. Thus, there may not be enough variability in task conflict slopes to predict team outcomes. This supports Hypothesis 7, that task conflict over time had no effect on team performance.

For Model 3 in the multilevel analysis, I added demographic characteristics at the within-team level and demographic faultline scores at the between-team level. In adding these input variables, I aimed to compare the predictive power of each demographic characteristic on the team conflict process. None of the four demographic variables (i.e., age, gender, ethnicity, or English as a first language) explained differences in relationship conflict slopes between team members. Only gender explained unique differences in within-team task conflict (b = 0.18, SE = 0.061, p = .01); age, ethnicity, and English as a first language had no unique relation to task conflict. These results show that men report higher *task conflict* than did women, yet there is no gender difference in reports of *relationship conflict*. Although this result is consistent with the previous study's findings, type I error rates may still be a concern with multiple predictors and comparisons in one model. Therefore, this result may not remain significant when personality predictors are included in the analysis or in a replication of this study. This only partially supported Hypothesis 8: although one demographic characteristic (i.e., gender) was related to one conflict type, demographic traits overall did not relate to conflict.

At the between-team level, demographic faultline scores positively predicted the slope of relationship conflict (b = 0.28, SE = 0.10, p = .01), whereas demographic faultlines were not related to the task conflict slope (b = -0.05, SE = 0.03, p = .11). This supports Hypothesis 12, which stated that relationship conflict will relate to demographic faultlines and that task conflict would not. This pattern of results is also consistent with the Model 2, as task conflict at the team level seems to have no significant slope. The link between demographic faultlines and relationship conflict slopes indicate that teams with deeper demographic rifts and subgroups have increased relationship conflict later in their projects. This may explain how relationship conflict starts and escalates, contributing to its negative impact on team outcomes.

Model 4 tested the relation between demographic characteristics and team effectiveness. At the within-team level, this involved testing the predictive strength of all four demographic characteristics and team effectiveness. At the between-team level, I tested the connection between demographic faultline strength and team outcomes. Only team members' native language (i.e., English or otherwise) explained differences in their ratings of team effectiveness (b = 0.40, SE = 0.14, p = .004). This result shows that team members for whom English was a second language rated their group as more effective than those whose native language was English. Ethnicity, age, and gender were not related to team member performance ratings. Although the effect of native language on team effectiveness is strong and it is highly significant, this result may also suffer from inflated type I error rates that accompany models with a high number of predictors. This partially supported Hypothesis 9, as only one of four demographic traits related to team performance. At the between-team level, demographic faultline strength was not directly related to team effectiveness (b = -0.53, SE = 0.58, p = .36). Thus, Hypothesis 13 was not supported. These results suggest that demographic characteristics of team members and entire teams do not contribute directly to team outcomes a great deal, yet they may act through team processes to influence performance.

Models 5 and 6 concern personality and its faultline strength on team conflict processes and performance outcomes. In the fifth multilevel model, I added HEXACO personality traits to the model to test each trait's relation to team conflict at the individual level and their combined faultline score at the team level. Team member relationship conflict was associated with emotionality (b = 0.15, SE = 0.045, p = .001), honesty-humility (b = -0.16, SE = 0.043, p < .001), and agreeableness (b = -0.084, SE = 0.04, p = .035). This indicates that team members with higher emotionality, lower honesty and humility, or less agreeable tendencies reported higher relationship conflict in the group. Of these results, the relation between agreeableness and relationship conflict had the smallest effect size and weakest significance level: this result may be a borderline finding that does not replicate in future research due to elevated type I error rates. As for team members' task conflict scores, conscientiousness was a positive predictor (b = 0.095, SE = 0.048, p = .046), openness was positively related to task conflict (b = 0.093, SE = 0.043, p

= .029), and extraversion was related to higher task conflict (b = 0.14, SE = 0.052, p = .009). This means that team members with higher prudence and detail orientation, openness to experience, or higher extraversion tendencies reported higher task-related disagreements. The three personality relations here may also be susceptible to inflated type I error rates. The results above may be nonsignificant if one made conservative adjustments for significance according to the multiple comparisons in this model. One can consider these results to partially support Hypothesis 10, as three personality traits relate to relationship conflict and three other personality traits relate to task conflict. At the team level, personality faultline strength was not related to relationship conflict slopes (b = -0.17, SE = 0.20, p = .39), or task conflict slopes (b = 0.15, SE = 0.08, p = .07). Hypothesis 14 was not supported for either relationship or task conflict. This shows no link between team-level personality rifts and team conflict trajectories.

The sixth multilevel model added paths between personality and team effectiveness. No personality traits were directly related to team effectiveness at the within-team level, rejecting Hypothesis 11. As with demographic faultline strength in the previous model, the strength of rifts in the team – along personality lines – was not related to team effectiveness (b = 0.86, SE = 1.26, p = .50). Accordingly, Hypothesis 15 posited that personality faultlines predict team performance: this was not supported. This compositional approach did not relate to team outcomes, unlike other single-variable approaches featured in previous meta-analytic work (e.g., Bell, 2007). The sample and context differences, the temporal gap between personality and performance, and the way performance was measured in this study may explain differences between these results and other team personality research beyond the limitations of aggregating members' personality traits.

Across models, few input and process variables were related to team effectiveness: only task conflict scores, relationship conflict scores, and team members' English as a second language status were related to team member effectiveness ratings. At the team level, only relationship conflict slopes were related to performance. Interesting patterns emerged, however, with the personality and demographic predictors of conflict at the individual level. Of the six personality traits measured in this study, three traits related significantly to task conflict (i.e., conscientiousness, openness, and extraversion), whereas the other three HEXACO traits related to relationship conflict (i.e., emotionality, honesty-humility, and agreeableness). Though team members' gender was connected to task conflict perceptions, no demographic characteristics were linked to relationship conflict scores. Demographic faultline strength across teams was a significant positive predictor of steeper relationship conflict, though it did not relate to task conflict slopes. These results (Figure 12) show the unique predictors and impacts of conflict types, across levels (i.e., within- and between-teams) and across analytic approaches (i.e., scores and trajectories).

Table 26. *Regressions including team-rated performance with multilevel modeling for Study 3.*

Variable	Level	TC	RC	Team Effectiveness
Model 1				
TC Scores	W			b = 0.11** [0.032, 0.18]
RC Scores	W			b = -0.15*** [-0.23, -0.064]
TC Scores	B			b = 0.86** [0.25, 1.47]
RC Scores	B			b = -0.67** [-1.00, -0.33]
Model 2				
TC Slope	B			b = 8.48 [-8.86, 25.82]
RC Slope	B			b = -1.09** [-1.72, -0.46]
Model 3				
Gender	W	b = 0.18** [0.058, 0.30]	b = -0.041 [-0.16, 0.075]	
Age	W	b = -0.012 [-0.053, 0.029]	b = 0.026 [-0.027, 0.079]	
Ethnicity	W	b = -0.021 [-0.14, 0.095]	b = 0.016 [-0.11, 0.15]	
English	W	b = -0.038 [-0.17, 0.095]	b = 0.083 [-0.06, 0.23]	
Demo FL	B	b = -0.05 [-0.11, 0.013]	b = 0.28** [0.082, 0.48]	
Model 4				
Gender	W			b = 0.062 [-0.18, 0.30]
Age	W			b = 0.041 [-0.059, 0.14]
Ethnicity	W			b = 0.20 [-0.045, 0.45]
English	W			b = 0.38** [0.12, 0.65]
Demo FL	B			b = -0.53 [-1.66, 0.60]
Model 5				
H	W	b = -0.012 [-0.088, 0.064]	b = -0.16*** [-0.24, -0.076]	
E	W	b = 0.047 [-0.047, 0.14]	b = 0.15*** [-0.23, -0.066]	
X	W	b = 0.13* [0.022, 0.23]	b = -0.034 [-0.14, 0.068]	
A	W	b = 0.02 [-0.08, 0.12]	b = -0.082* [-0.16, -0.004]	
C	W	b = 0.095* [0.001, 0.19]	b = 0.013 [-0.067, 0.093]	
O	W	b = 0.093* [0.009, 0.18]	b = 0.036 [-0.035, 0.11]	
Hexaco FL	B	b = 0.15 [-0.015, 0.31]	b = -0.17 [-0.56, 0.22]	
Model 6				
H	W			b = 0.066 [-0.095, 0.23]

Table 26. (continued).

E	W	b = 0.046 [-0.12, 0.21]
X	W	b = 0.16 [-0.018, 0.34]
A	W	b = 0.095 [-0.095, 0.29]
C	W	b = -0.15 [-0.32, 0.021]
O	W	b = 0.04 [-0.11, 0.19]
Hexaco FL	B	b = 0.86 [-1.61, 3.34]

Note. b = unstandardized regression coefficient. TC = Task Conflict, measured as a team member score at the within-team level and as a slope at the between-team level. RC = Relationship Conflict, measured as a team member score at the within-team level and as a slope at the between-team level. W = Within-Team (i.e., Individual) level. B = Between-Team (i.e., Team) level. Demo FL = Demographic Faultlines. Hexaco FL = Personality Faultlines. Square brackets contain values representing 95% confidence intervals; intervals that contain zero are considered non-significant. * = p < .05, ** = p < .01, *** = p < .001.

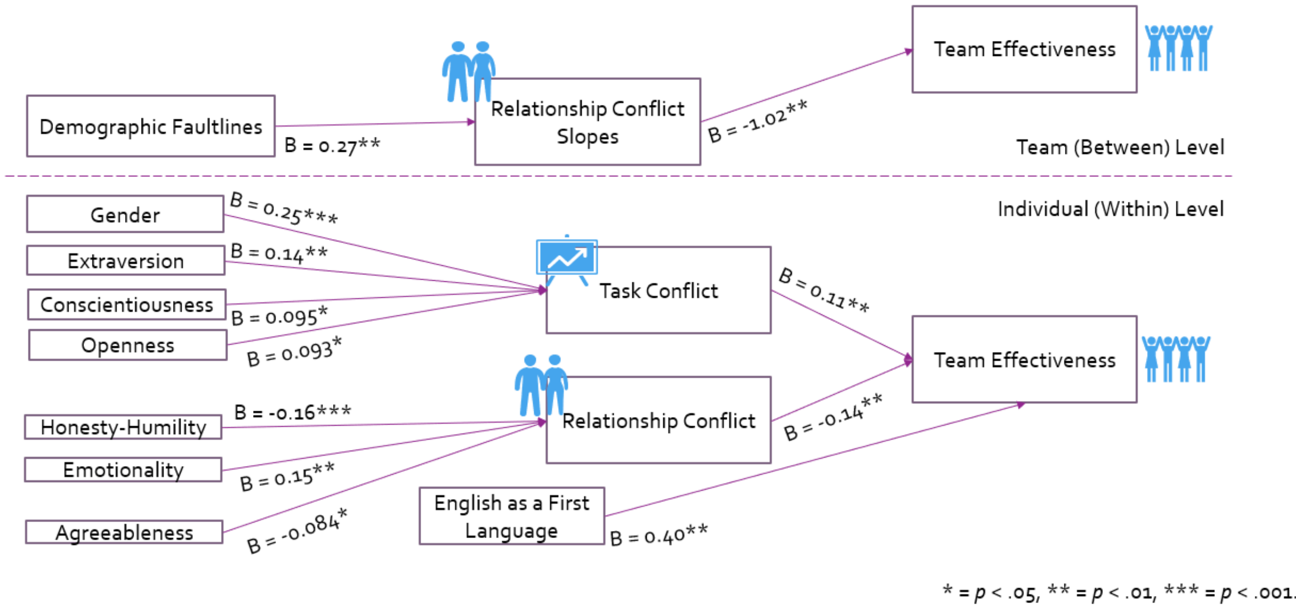

Figure 12. Summary of significant results for Model 6 in Study 3.

Note. Team-level relationship and task conflict intercepts are missing from the figure above, as the multilevel model of longitudinal data used for these analyses did not calculate team intercepts. As well, task conflict slopes are missing from this figure because there were no significant relations between task conflict slopes and team inputs or team effectiveness.

Growth Mixture Modeling at the Team Level

Next, I replicated the growth mixture modeling approach in Studies 1 and 2 with the team-level conflict scores from all three survey administrations. As Study 1 found two classes, I used a two-class mixture model to measure the intercept and slope of task and relationship conflict. In the first of two models, I computed the intercept and slope for both conflict types and measured their intercorrelations while computing the team effectiveness mean for each class. The task conflict intercepts were similar across classes (Table 27), yet Class 1 had a significantly positive task conflict slope whereas task conflict in Class 2 had no significant slope (Figure 13). The relationship conflict intercepts for both classes were also similar, yet Class 1 had a much flatter relationship conflict slope at 0.20 ($p < .001$) than Class 2, where the slope was over five times steeper, at 1.07 ($p < .001$). This suggests one cannot distinguish the two classes of teams at the beginning of their time together, as they have similar intercepts but differing slopes. None of the slopes or intercepts were significantly intercorrelated.

The team effectiveness mean was higher for Class 1 at 5.15 ($p < .001$) than for Class 2 at 4.60 ($p < .001$). This is consistent with the multilevel model results for Studies 2 and 3, which show that relationship conflict has a negative relation to team effectiveness. The average personality faultline score was similar for both classes. Class 1's personality faultline mean was 0.26 ($p < .001$), whereas Class 2's personality faultline mean was 0.28 ($p < .001$). However, the demographic faultline mean was somewhat higher for Class 1 at 0.87 ($p < .001$), than for Class 2 at 0.74 ($p < .001$). As Class 2 is very small, these mean differences may not be robust. Interestingly, this suggests that the class with the higher relationship conflict slope has a slightly lower demographic faultline average; this contrasts with the positive relation between rifts in the team along demographic lines and increasing relationship conflict. In this model, the mean differences suggest that teams with a steeper relationship conflict slope have lower team effectiveness than teams with a shallower slope for relationship conflict. However, Class 2 is much smaller than Class 1; this means any conclusions should be tempered according to the sample sizes of these classes.

Table 27. *Growth mixture modeling with two classes for Study 3.*

Class	Conflict	Faultlines	Team Performance
# 1 (n = 148)	TC Intercept: 3.12*** TC Slope: 0.082** RC Intercept: 1.40*** RC Slope: 0.20***	Demographic: 0.87*** Personality: 0.26***	Team Effectiveness: 5.15***
# 2 (n = 11)	TC Intercept: 3.16*** TC Slope: 0.25 RC Intercept: 1.35*** RC Slope: 1.07***	Demographic: 0.74*** Personality: 0.28***	Team Effectiveness: 4.60***

Note. TC = Task Conflict, RC = Relationship Conflict. ** = $p < .01$, *** = $p < .001$.

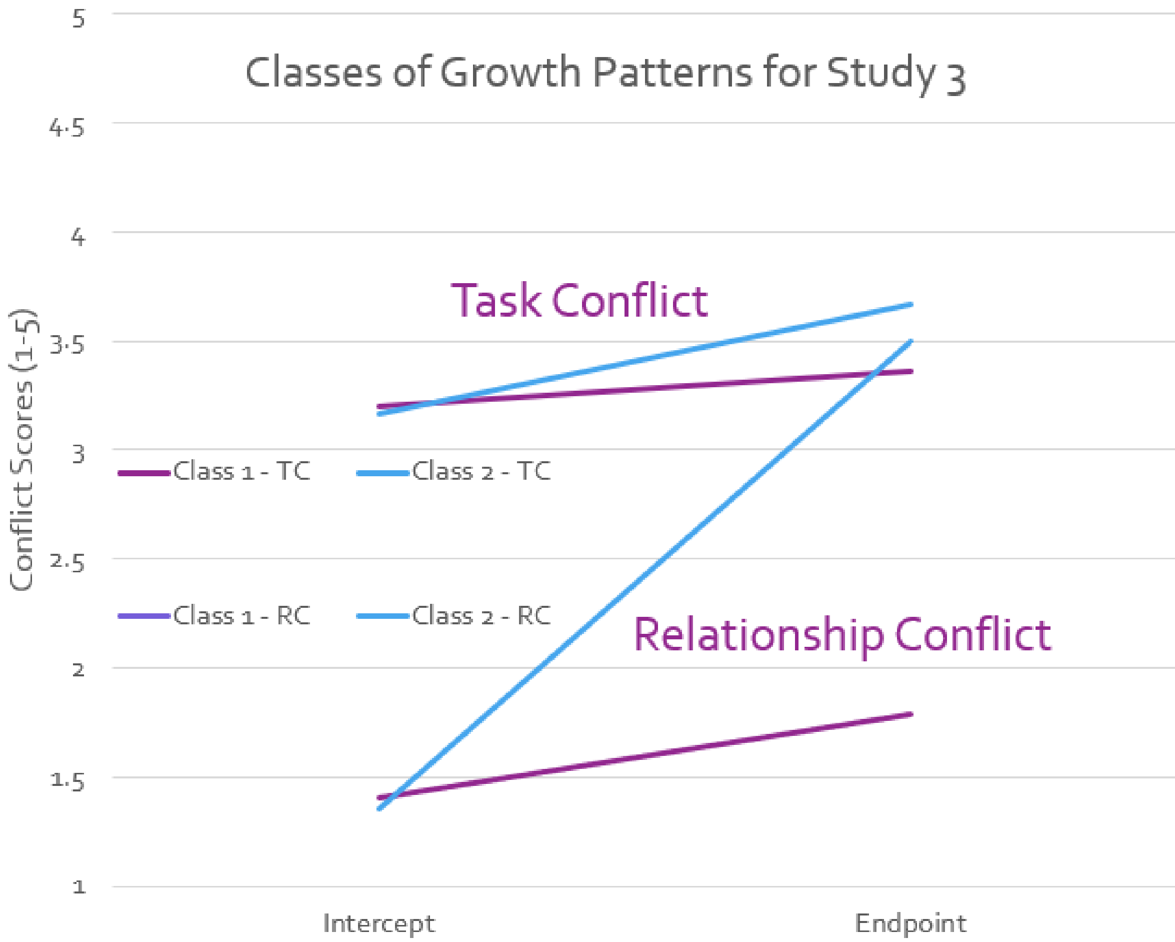

Figure 13. Growth mixture modeling results for Study 3.

Note. Relationship conflict trajectories for both classes begin below the legend and task conflict trajectories for both classes are above the legend.

In the second model, I tested the overall relations among team inputs, conflict, and team effectiveness (Table 28). I analyzed the predictive paths between faultline scores, conflict intercepts and slopes, and team effectiveness across these two classes. There was no connection between faultline scores and conflict slopes or intercepts, yet personality faultline scores were positively related to team effectiveness (b = 1.19, SE = .0.52, *p* = .02) and demographic faultline scores were negatively related to team effectiveness (b = -0.42, SE = 0.21, *p* = .045). Finally, task conflict slopes were positively related to team effectiveness (b = 1.09, SE = 0.45, *p* = .02), whereas relationship conflict slopes were negatively related to team effectiveness (b = -0.77, SE = 0.28, *p* = .006). These results suggest that differences in personality and having more debates about the task over time can help teams perform effectively, whereas demographic differences and escalating personal disagreements can hurt teams' effectiveness.

Table 28. *Regressions with growth mixture modeling for Study 3.*

Variable	Task Conflict Slope	Relationship Conflict Slope	Team Effectiveness
Personality Faultlines	b = -0.037 [-0.36, 0.29]	b = -0.14 [-0.43, 0.16]	b = 1.19* [0.18, 2.20]
Demographic Faultlines	b = -0.036 [-0.20, 0.12]	b = 0.19 [-0.03, 0.41]	b = -0.42* [-0.84, -0.009]
Task Conflict Slope			b = 1.09* [0.21, 1.96]
Relationship Conflict Slope			b = -0.77** [-1.32, -0.22]

Note. N = 138 teams. b = unstandardized regression coefficient. Square brackets contain values representing 95% confidence intervals; intervals that contain zero are considered non-significant. * = p < .05, ** = p < .01.

Discussion

Study 3 used separate traits, at the individual level, to test the relationships between member traits and conflict. This study also used all demographic characteristics and all HEXACO personality traits, at the team level, to create configural faultline measures. Relationship and task conflict scores were related to final-stage performance as rated by team members, whereas relationship (but not task) conflict slopes were related to team-rated outcomes. These null results for task conflict as compared to the significant results for relationship conflict may be explained by a restriction of range in this sample. Task conflict scores did not change substantially from the first to the last survey, as supported by the nonsignificant variance in task conflict slopes. Task conflict slopes showed no relations to team inputs or outputs. Future research could experimentally manipulate task conflict to create a larger task conflict range and provide a stronger test of the task conflict-performance link.

Demographic characteristics showed some significant relations with conflict and performance. Men reported higher levels of task conflict within the team, whereas non-native English speakers rated their team performance higher than native English speakers in the team. This sample was collected in an environment of mostly men (i.e., approximately 80%); there may be many reasons why men reported higher task conflict. Future research conducted in more gender-balanced industries, or comparative research conducted in industries such as engineering along with industries that have a majority of women (i.e., nursing) may be informative for explaining these results. However, all other demographic characteristics were not related to team processes or outcomes. Future research on differential member perceptions of conflict and performance in the team, similar to Sinha and colleagues' (2016) study, may help clarify these demographic differences in ratings.

An interesting pattern of personality results emerged for individual-level task conflict: agreeableness, emotionality, and honesty-humility were related to team member reports of relationship conflict. Team members with higher emotionality may see task conflict as spilling over into relationship conflict more often than team members with lower emotionality do. Research has highlighted the importance of team issues and task conflict emotionality in explaining when task and relationship conflict are highly coupled (Rispens, 2012). This framework for categorizing team issues may help to explore when team members with higher emotionality experience and/or perceive more conflict. Further, employees perceived conflict more negatively if the issue was unresolved (Gayle & Preiss, 1998), suggesting there may be a mutually reinforcing connection between team member emotionality and perceptions of unresolved conflict within the team.

Interestingly, emotionality was not uniquely related to task conflict, though gender was associated with task conflict ratings. As women consistently report higher emotionality (Lee & Ashton, 2020; Lynn & Martin, 1997; Moshagen, Thielmann, Hilbig, & Zettler, 2019), one may expect that emotionality could explain gender differences in task conflict perceptions. However, this was not the case in this study, in which men reported higher task conflict. This result would support the opposite result than would be hypothesized by existing personality research. However, relationship conflict, for which gender was not a unique predictor of individual-level scores, may be more closely related to emotionality than task conflict was. Future research could test whether gender differences in conflict perceptions are mediated by personality traits such as emotionality.

At the individual level, task conflict scores were related to the three remaining HEXACO personality traits: conscientiousness, openness to experience, and extraversion. These traits likely relate to higher task conflict perceptions for distinct reasons. Team members with higher conscientiousness may be more inclined to focus on the task at hand given their higher prudence and they may suggest changes to the team's approach due to their high detail orientation. Team members with more openness to experience may discuss or remember divergent ideas more often than team members who are less open. Finally, extraverted team members may participate in task-related discussions more often, increasing their availability (Tversky & Kahneman, 1973) and resulting in higher task conflict scores. Whereas this study does not directly test the mechanisms through which personality traits affect conflict types, these results begin to explain why some team members may perceive more conflict than others.

Demographic faultline scores, unlike personality faultline scores, were related to conflict in this study. Specifically, stronger demographic faultlines in teams were linked to steeper relationship conflict slopes but were unrelated to task conflict slopes. This is consistent with previous research on team diversity and conflict generally (Goyal, Maruping, & Robert, 2008). Specifically, demographic faultlines are negatively related to team functioning (Molleman, 2005). Team subgroups, reflected in strong demographic faultlines, may create an "us versus them" attitude in team members (Labianca, Brass, & Gray, 1998) and result in higher relationship conflict over time. One fruitful avenue for future research is to design studies that test existing interventions aimed at deactivating demographic faultlines (van der Kamp et al., 2011) to monitor their impact on team conflict.

The null results for personality faultlines, observed in this research, may have many explanations; these include limitations associated with the conceptualization and assessment of personality faultlines, the randomization process used to select team members, and the variables chosen in this study. Neither faultline approach was directly related to performance, which does not support previous meta-analytic findings showing that faultlines were related to group performance and satisfaction (Thatcher & Patel, 2012). However, these researchers only reported the direct relation between faultline strength and one measure that included performance and satisfaction; thus, relations between faultlines and performance may be weaker than reported. Yet the interaction between faultline strength and faultline distance was negatively related to group performance, independent from measures of group satisfaction in the same study. Perhaps team behaviours other than conflict may relate strongly to personality differences, such as creativity and innovation (Somech & Drach-Zahavy, 2011). Different conceptualizations, therefore, may result in significant relations between faultlines and team variables.

The results above result from treating all teams as part of one class, represented by a single sample distribution. However, this may not reflect the data in Study 3 accurately. Using a two-class solution, I found differences in conflict slopes, outcomes, and faultlines across both classes. Class 1 had a steeper relationship conflict slope, higher team performance scores, and higher average demographic faultline strength than Class 2. The higher team performance score for Class 1 is surprising in light of the negative connection between steeper relationship conflict slopes and lower performance scores. Further investigation may clarify differences between these multi-class results and more conventional analytic approaches.

Limitations

In addition to the limitations mentioned in Studies 1 and 2, there are some unique limitations to the analytic approach I used in Study 3. Faultline strength calculations are one method to compute the interaction between multiple traits on a team. However, the obtained faultline value depends heavily on the number and characteristics of the selected traits. Very little previous research has used personality-based faultlines (e.g., Molleman, 2005; van der Kamp et al., 2011), and none of this research has used the HEXACO structure of personality to compute their faultline scores though its structure is highly similar to the Big Five model. For this reason, the appropriateness of this approach is unknown.

However, existing research on demographic and informational faultlines suggests that faultlines on deeper-level attributes may be relevant to these project-based design teams. Under time pressure, with complex tasks, and during periods of intense collaboration, differences between groups become activated and can hurt collaboration and knowledge sharing between team members (Gratton, Voigt, & Erickson, 2007). These group differences can reflect surface-level attributes such as age, gender, and ethnicity, upon which most initial research was conducted, or deep-level attributes such as tenure, job function, and personality (Thatcher & Patel, 2011). The present study involved strong deadlines, complex tasks, and collaboration; due to these characteristics, one would expect faultlines based on multiple types of team member attributes (i.e., both surface- and deep-level) to influence team processes. This is supported by previous research in the team diversity field. In one study, team members paid more attention to diversity in personality traits than surface-level differences between members – even when the salience of those visible differences were manipulated to attract more attention through the researchers’ experimental design (Meyer, Shemla, & Schermuly, 2011).

As well, informational faultlines have been studied more often than personality faultlines. These constructs reflect deep-level attributes related to team members’ experiences, such as their tenure in an organization, their department, and their job function. When teams are highly autonomous and they have stronger faultlines across conscientiousness and educational background, their performance was lower than for teams with weaker deep-level faultlines (Rico, Molleman, Sánchez-Manzanares, & Van der Vegt, 2007). In another study, informational faultlines had a negative impact on the creativity levels of research and development teams when teams had less external knowledge acquisition and did not integrate their knowledge well internally (Qu & Liu, 2017). Finally, personality variation within teams is related to poorer performance in meta-analytic research (e.g., Bell, 2007) and in a study of organizational work teams (Barrick et al., 1998). Further research on many faultline approaches may discover the ideal methods to compute personality faultlines and the correct theoretical foundation for these investigations.

Next, there are many approaches for computing team faultline scores (Meyer & Glenz, 2013). Although the approach I used in this study, developed by Gibson and Vermeulen (2003) has been previously validated, the results in this study could change slightly if I used another method of calculating group faultlines. Though surface-level and demographic faultline research has a longer history than the study of personality-based faultlines, the demographic traits used to compute faultline scores vary across studies. This can lead to large differences in faultline scores that depend on which demographic traits that researchers choose for this computation. This relatively new field of study lacks clear guidelines for when to use each calculation method, which introduces more variability between studies and research groups. These differences in using faultline approaches may result in a field of study where research on the same topic cannot be easily compared. Thus, the results in this study may not hold if faultlines include different or fewer traits, if faultlines are computed differently, or if further research in personality-based faultlines develops clearer norms for these methods.

Finally, results from the multi-class analyses using growth mixture modeling may be overstated. The sample sizes of classes in this study are highly mismatched, which may create results that are difficult to replicate. These sample sizes may produce higher overlap between average scores in Classes 1 and 2 due to larger standard error values in the smaller sub-sample.

Future research with much larger samples of teams can extend these results in a more robust manner to determine if these faultline and performance differences hold across classes.

GENERAL DISCUSSION

In these three studies, I sought to compare team conflict scores across time, uncover any underlying classes that reflect conflict patterns, and test an input-process-output model of project teams. Taken together, I establish that conflict can be reliably measured across time, that teams diverge into two classes after their intercept, and that some paths exist between demographics and personality as team inputs, conflict, and team-rated performance. Further, conflict slopes add a useful layer of explanatory power to differences in team performance. Whereas task and relationship conflict scores predicted team performance at both levels, only relationship conflict slopes were related to team outcomes.

Comparing results across Studies 2 and 3, some demographic results were not consistent. Whereas task conflict scores at the individual level consistently showed gender differences, with men reporting more task conflict in both studies, some results were not so robust within or across studies. Team member ethnicity was a significant predictor of team effectiveness in Study 2, yet not in Study 3. Team members' English as a first language status, however, was related to members' ratings of team effectiveness in Study 3. This result may be a proxy for the ethnicity finding from the previous study, as ethnicity ratings and English language status were highly correlated at the individual level in both studies.

The distinct relations between personality traits and conflict show that, in this set of studies, no personality traits were consistently and simultaneously related to both task and relationship conflict. Though honesty-humility was related to individual-level task and relationship conflict in Study 2, these relations were not highly significant and did not both reappear in the Study 3 results. In addition, the results for demographic characteristics and personality traits were not all consistent across Studies 2 and 3. In Study 2, only two personality traits (i.e., Honesty-Humility and Agreeableness) were related to conflict at the individual level, whereas all six traits were related to some form of individual-level conflict in Study 3. Whereas personality faultlines had no relations with variables at the team level, the average extraversion level in teams had a significant relation to team effectiveness at the team level in Study 2. This suggests that some of these relations may be sample-dependent, or that some individual and team inputs may reflect high type I error and may not replicate for this reason. As teams were composed by randomly assigning members, the composition of teams based on personality and the interactions between members with unique personality profiles may lead to different individual- and team-level relationships. Despite many similarities between the samples and data collection methods in Studies 2 and 3, teams may be composed of different personalities from one study to the next.

Whereas demographic faultline results were in line with previous research, personality faultline scores were not related to either conflict or performance. Many individual personality traits were related to conflict types in these studies. However, other compositional approaches of personality also relate to team performance (Bell, 2007); this means personality faultlines may be a fruitful new avenue of research for team composition (e.g., Molleman, 2005). Overall, these results contribute to the body of literature on team inputs, dynamic processes, and outcomes for project teams.

Implications

Theoretical implications. These three studies have implications for teamwork in addition to other areas of research including team diversity and personality. Here, I advance the study of team conflict over the lifecycle of project teams by showing that teams and their members can follow different conflict trajectories. This adds nuance to theories of project team processes by showing how groups respond to time pressure and stress from project deadlines. I use the IPO framework (McGrath, 1964), specifically expanding our knowledge of team processes or mediators (Ilgen, Hollenbeck, Johnson, & Jundt, 2005), to include longitudinal, multilevel conflict following Humphrey and Aime's (2014) recommendations for theorizing team interactions. Expanding upon team inputs, I advanced the study of newer and relatively unexplored composition approaches, such as personality faultlines (e.g., Molleman, 2005). By analyzing how relationship and task conflict co-occur, I advance the contextual approach to team conflict that considers multiple conflict types at once. This serves to qualify the mixed results observed between task conflict and performance in previous research that uses meta-analytic methods (De Dreu & Winegart, 2003; De Wit et al., 2012).

This research also provides important measurement information for analyzing the dynamics of team conflict. I established measurement invariance, providing confidence for future researchers that these commonly used measures of team conflict are reliable over time. Task and relationship conflict had minimal differences in factor structure, loadings, intercepts, and residuals over time at the individual level and consistent factor structure, loadings, and intercepts at the team level. This suggests that future research on dynamic task and relationship conflict will be consistently measured across the lifecycle of project teams.

Diversity researchers may also benefit from the advances in this research. Though there were small and sometimes inconsistent relations between individual demographic characteristics and team variables, this research adds to the study of multiple demographic traits at once. This research finds that the way in which demographic characteristics are organized within teams, such as the faultlines calculated in Study 3, can be more influential for conflict at the team level than each trait examined one by one. The minimal impact of each demographic variable on conflict and performance supports existing meta-analytic results that show surface-level (i.e., generally demographic) diversity has a small, if significant, relation with team processes and outcomes (Bell et al., 2010; van Dijk et al., 2012). The configural approach used here, by calculating demographic and personality faultlines, showed the distribution of easily visible, surface-level traits on a team affects the trajectory of that team's relationship conflict, yet the distribution of deep-level (i.e., personality) traits did not affect teams' conflict levels or effectiveness. As I aggregated individual team members' personality traits, these results do not contribute much to the literature on diversity of personality or other deep-level traits.

However, Studies 2 and 3 apply to the broader study of personality in teams and individuals. These results show that individual members perceive team conflict differently: those with higher honesty-humility or higher agreeableness were less likely to report experiencing relationship conflict in both studies. Other personality and conflict results at the individual level differed across studies; this suggests individual members' personality traits may interact with team dynamics to determine how each member sees their team processes unfolding. At the team level, the higher average level of conscientiousness or extraversion on the team, the higher the team rated their effectiveness overall. These results, whether due to different team perceptions of the same performance or due to truly different performance in these teams, can advance the current research conducted on team inputs.

Practical implications. It seems reasonable to suggest that this research has relevance for team composition, conflict interventions, and performance improvement. Whenever possible when composing teams, managers and human resources professionals should consider each team members' characteristics including demographic and personality traits. Using a validated personality measure (e.g., Ashton & Lee, 2009) and member demographic characteristics available in many human resources information system databases, practitioners can staff teams with members who are less likely to spark negative conflict types, such as teams with minimal demographic faultlines. Team members' demographics may play a role in team conflict, depending on how they interact with other team members' attributes in forming faultlines. Building on existing demographic faultline research (e.g., Thatcher & Patel, 2012), practitioners can compose teams low in demographic faultlines to encourage more positive conflict expression and a lower relationship conflict trajectory. Whereas no single demographic trait was detrimental for teams, the value of reducing faultlines may come from finding the right "fit" for all team members.

To reduce team conflict, practitioners can begin by measuring conflict early and often. Although teams began with similar conflict intercepts, growth mixture modeling analyses showed that classes of teams separated relatively quickly. By six to eight weeks into a project, practitioners can identify teams with increasing levels of relationship conflict that may threaten the performance and effectiveness of the team. From this point of early identification, practitioners can train teams to resolve task conflicts before they become personal. Using conflict expression theory (Weingart et al., 2015), teams can learn to reduce their oppositional intensity and be more direct when expressing conflict to stimulate healthy debate. To prevent any spillover from task to relationship conflict, teams can practice mindfulness, as team mindfulness relates to lower relationship conflict and weakens the link between relationship and task conflict (Yu & Zellmer-Bruhn, 2018). To increase team performance as rated by the team, practitioners can track relationship conflict and intervene before relationship conflict escalates further, potentially through mediation (Jehn, Rupert, & Nauta, 2006).

Limitations

These three studies have some limitations that constrain the causal inference, assessment of levels of analysis, and generalizability of the results. First, these studies cannot make strong causal inference claims due to the lack of experimental control. A tradeoff exists between a highly controlled experimental sample and a field sample such as that used in this set of studies. Here, experimentation was not possible due to the high-stakes nature of the team's outcomes. Second, the longitudinal nature of the analyses precluded testing reverse causation models to explore alternative explanations for the data. This limits the strength of causal inference further.

Although this research used a multilevel approach, I did not study team conflict from a dyadic perspective. Disagreements often begin as interpersonal interactions between two team members (Humphrey & Aime, 2014); thus, studying conflict at the dyadic level would be appropriate. This analysis would require other measurement approaches, such as peer ratings and network analysis to measure the presence and strength of dyadic conflict. In this set of studies, I conceptualized individual-level conflict as a team member’s perception of the multiple dyadic exchanges they experience when working in the group along with their observations of group conflict as a third-party observer. This means group-level conflict ratings may reflect rough estimates of the dyadic conflicts personally experienced by each team member and viewed by other members as ‘bystanders’ to the conflict (Korsgaard et al., 2008).

Another limitation of the current research stream concerns whether other predictors of team performance (e.g., Salas, Shuffler, Thayer, Bedwell, & Lazzara, 2015) may account for the causal link between conflict and performance. There are many other potential predictors of team performance, including general mental ability, collective intelligence, cohesion, potency, collective efficacy, and information sharing. These predictors may explain why team conflict relates to performance. Future research should explore the relative strength and interactions of the team performance predictors in the literature that were omitted here. Finally, this homogeneous sample contained project teams completing similar tasks with identical instructions and performance evaluation methods. This set of age-constrained engineering trainees limits the generalizability of these studies. Future research should aim to replicate these results in a more heterogeneous sample of project teams in a workplace environment.

Study Strengths

Despite the limitations of these three studies, this sample is helpful for conducting initial research on this topic for many reasons. First, these teams had no formal hierarchy, reflecting many knowledge-based teams in modern organizations. This improves the generalizability to knowledge-based teams, especially those completing engineering design-type projects. Second, although the stakes may seem low to an observer, the groups' tasks and consequences are meaningful for members themselves. Indeed, course grades provide a high-stakes and realistic performance metric with consequences for future academic and career success. This means my results generalize best to other high-stakes environments, such as deadline-driven project teams. Third, these studies benefit from the consistency of a controlled study in a constructed environment (i.e., the classroom) and the realistic consequences and longer lifespan of a field study. This strengthens conclusions about dynamic change over many months, which reflects the length of software projects completed by some engineering teams in high-technology companies (e.g., Colomo-Palacios, Casado-Lumbreras, Soto-Acosta, García-Peñalvo, & Tovar, 2014). These fast-paced, agile design teams (Lindsjørn et al., 2016; Tripp et al., 2016) tend to follow a work style that matches the study design in this research program; thus, practitioners in agile technology companies may particularly benefit from this line of inquiry.

Finally, members were randomly assigned to teams at the beginning of their time together. This ensures that members were largely unfamiliar with each other and that previous social relationships would be unlikely to contribute to the team interactions experienced during the studies. Finally, all teams started and finished each project at the same time. Teams were given identical tasks, team member inputs were assessed, and multiple performance criteria (i.e., team-rated effectiveness and other-rated objective performance) were measured consistently. This avoided alternative explanations, based on differences between group tasks and structure, for team conflict trajectories and performance.

Future Research

Future research can investigate two major areas of study from this work: team inputs (i.e., demographic characteristics and personality traits) and team conflict. To build on this team input research, scholars can: investigate links between multiple demographic and personality traits, for example through interaction analyses; study teams from other environments to test the impact of these inputs at individual and team levels; and use different methods for measuring personality faultlines. Faultlines are one method of testing the interaction between multiple variables. Other methods include profile analyses (Espinoza, Daljeet, & Meyer, 2020) and interactions between traits. Profile analyses take a person-centred approach by considering how multiple traits are represented within each individual. To my knowledge, this approach has not yet been used for team-level inputs. However, existing research finds interesting results for profile analyses on team conflict types (O’Neill et al., 2018). Two- or three-way interactions are another, more traditional method of testing how multiple traits affect each other. However, these approaches tend to have low power (e.g., Aguinis, 1995) and are limited to two or three traits at once for this reason.

As mentioned above, this context is unique in its skewed proportion of men on teams and its quick, deadline-driven nature. The demographic characteristics and personality traits that impact conflict and performance here may not show the same results in women-dominated environments or in more stable work environments with less time pressure. Further investigations are necessary to understand how widespread these results are. Finally, personality faultline research is in its infancy relative to demographic faultline research and other, traditional approaches to team diversity. To ensure consistency across the field and to increase the ability to compare results across studies, researchers can establish guidelines for conducting personality faultline research.

Researchers can build on this work in team conflict in three ways: by testing solutions to conflict-induced performance challenges, by measuring conflict in different ways, and by studying conflict at different levels. To improve team functioning, future research should explore conflict expression, resolution, and management, psychological safety, and team mindfulness to track their effects on conflict over time. Recent research on conflict expression (Weingart et al., 2015) provides a relatively new avenue for team conflict interventions. For example, research found that frequent mild task conflict can instill positive emotions in team members, by motivating members to acquire more information (Todorova, Bear, & Weingart, 2014). One particularly novel study used robots to intervene after a team conflict episode (Jung, Martelaro, & Hinds, 2015). Using this paradigm with longitudinal research designs, researchers can measure the impact of conflict expression training and manipulating team expectations about directness and oppositional intensity, the two elements of conflict expression.

Conflict management and resolution is a similar research area that warrants further investigation; teams that take a more cooperative (rather than competitive) approach to conflict management can improve cohesion within the team (Tekleab, Quigley, & Tesluk, 2009). Further, teams that manage conflict directly can build more constructive team environments that enhance performance (Cameron, 2000; Montoya-Weiss, Massey, & Song, 2001). Conflict resolution may involve interventions that change the nature of team members' conflict expression, specifically to increase the directness and to reduce the oppositional intensity of disagreements. This and other interventions can be delivered in multiple formats to improve team functioning. For example, behaviour modeling training (Taylor, Russ-Eft, & Chan, 2005) is one training method that uses active practice, reflection, and feedback to improve performance. Teams may benefit from this participatory method of instruction to shift the type of conflict that members experience and to reduce the negative impact of relationship conflict on performance. Studies which track and influence positive conflict management styles can extend this research area through stronger theoretical and practical implications.

There are other team constructs researchers can draw on to alleviate the negative impact of conflict, including psychological safety (O'Neill & McLarnon, 2017) and team mindfulness. In their study of project teams, Bradley and colleagues (2012) found that a psychologically safe team climate can reduce relationship conflict while promoting some task conflict. This team experience, in which team members feel comfortable taking social risks and being open with one another (Edmondson, 1999), can improve team performance. In the future, researchers might also explore how psychological safety affects the type and tone of information sharing within teams to promote productive conflict.

Mindfulness has been used as a team characteristic to describe to what extent team interactions reflect awareness about the present and non-judgmental processing of experiences that team members have (Kabat-Zinn, 2005; Yu & Zellmer-Bruhn, 2018). Recent research on this construct finds that teams who practice shared mindfulness have lower relationship conflict and less spillover from relationship to task conflict. Other research finds that team mindfulness mediates the relation between individual mindfulness and work engagement (Liu, Xin, Shen, He, & Liu, 2020). Thus, team mindfulness and psychological safety may shift teams' conflict trajectories from a poor performance path, potentially marked by higher stress and less constructive group interactions, towards better performance.

Beyond the methods used in this set of studies, researchers can expand the ways team conflict is measured. Conflict can be measured through other types including status conflict (Bendersky & Hays, 2012), through other compositional methods, and through different data collection methods. Status conflict is a newly discovered form of team conflict that is distinct from the three types of conflict discovered by Jehn (1995) and Behfar and colleagues (2011). Incorporating status conflict in these analytic approaches may add nuance to the present research. For example, status conflicts may spark the transfer from dyadic conflict to group-level disagreements or from task-related to relationship conflict. Next, future research can extend this work beyond agreement-based indices of team conflict to use skewness (e.g., Sinha et al., 2016), variance, minimum, and maximum team member scores to reflect team conflict. These approaches can be used in multilevel analyses (e.g., Cole et al., 2011). Finally, methods that allow for more frequent data collection, as well as behavioural- or observation-based measures of conflict, can advance the study of team conflict over time. By increasing the frequency of data collection, researchers can analyze fine-grained changes in individual, dyadic, and team-level perceptions of conflict and discover the optimal time for intervention. Given the limitations of self-report measures (e.g., Donaldson & Grant-Vallone, 2002), other behaviour-based or observational approaches to measuring conflict may avoid the pitfalls of survey-based research.

Humphrey and Aime (2014) called for more multilevel, dynamic investigation into teams. This set of studies aimed to answer this call, yet much more can be done to analyze conflict and other variables at multiple levels. Future research can explore dyadic conflict, multi-team systems, and consensus emergence. The dyadic level is where team disagreements likely start (Humphrey & Aime, 2014); thus, dyadic-level analyses may help to advance research on how conflicts are instigated within teams. New techniques and approaches, including network analysis, may help to test advanced theories about conflict by mapping how every team member sees conflict with every other member of the group (Park, Mathieu, & Grosser, 2020). Above the team level, multi-team systems are a new consideration for teams that are embedded in organizations (West et al., 2015). Future research can build on published studies on conflict in multi-team systems (e.g., Berg, Curseu, & Meeus, 2014) to measure how inter-team conflict can impact intra-team performance. Researchers can use multiple levels to show if and how consensus emerges (Lang, Bliese, & de Voogt, 2018) among team members over time. Longitudinal measures of conflict will be more informative if they are paired with longitudinal measures of team performance from multiple sources. Future research that measures team performance alongside conflict can test reciprocal relationships between conflict and performance and further compare results for team- vs other-rated performance metrics.

Conclusion

This set of studies sought to explore the dynamic, multilevel nature of team conflict, its antecedents, and its outcomes. Whereas nearly all teams experienced conflict, the trajectory of their disagreements differed. When composing teams, single demographic traits, demographic rifts in the team (i.e., faultlines), and personality characteristics determine the level and direction of conflict. For project-based design teams, their focus on innovation may explain why task conflict is consistently high among teams and across time. Demographic faultlines may spark rapidly increasing relationship conflict, which in turn dampens team-rated performance. Future research should measure conflict more often over project teams' lifecycles and test interventions in workplace teams to improve project team success.

References

Aeron, S., & Pathak, S. (2017). Personality, conflict and performance: Exploring predictive relationships. *IUP Journal of Organizational Behavior*, *16*(2), 35-54.

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, *21*(6), 1141-1158.

Amason, A. C., & Mooney, A. C. (1999). The effects of past performance on top management team conflict in strategic decision making. *International Journal of Conflict Management*, *10*(4), 340-359.

Amason, A. C., & Schweiger, D. M. (1994). Resolving the paradox of conflict, strategic decision making, and organizational performance. *International Journal of Conflict Management*, *5*(3), 239-253.

Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, *110*(5), 766-781.

Arman, G., & Adair, C. K. (2012). Cross-cultural differences in perception of time: Implications for multinational teams. *European Journal of Work and Organizational Psychology*, *21*(5), 657-680.

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*(6), 1531-1544.

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340-345.

Awang, Z. (2012). *A Handbook on Structural Equation Modeling Using AMOS.* Universiti Technologi MARA Press, Malaysia.

Banker, R. D., Field, J. M., Schroeder, R. G., & Sintia, K. K. (1996). Impact of work teams on manufacturing performance: A longitudinal field study. *Academy of Management Journal*, *39*(4), 867-890.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*(3), 377-391.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, *92*(3), 595-615.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management*, *37*(3), 709-743.

Behfar, K. J., Mannix, E. A., Peterson, R. S., & Trochim, W. M. (2011). Conflict in small groups: The meaning and consequences of process conflict. *Small Group Research*, *42*(2), 127-176.

Behfar, K. J., Peterson, R. S., Mannix, E. A., & Trochim, W. M. (2008). The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes. *Journal of Applied Psychology*, *93*(1), 170-188.

Bendersky, C., & Hays, N. A. (2012). Status conflict in groups. *Organization Science*, *23*(2), 323-340.

Bentein, K., Vandenberghe, C., Vandenberg, R., & Stinglhamber, F. (2005). The role of change in the relationship between commitment and turnover: A latent growth modeling approach. *Journal of Applied Psychology*, *90*(3), 468-482.

Berg, W. V. D., Curseu, P. L., & Meeus, M. T. (2014). Emotion regulation and conflict transformation in multi-team systems. *International Journal of Conflict Management*, *25*(2), 171-188.

Bikfalvi, A., Jäger, A., & Lay, G. (2014). The incidence and diffusion of teamwork in manufacturing – evidence from a Pan-European survey. *Journal of Organizational Change Management*, *27*(2), 206-231.

Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology*, *69*(5), 890-902.

Bono, J. E., Boles, T. L., Judge, T. A., & Lauver, K. J. (2002). The role of personality in task and relationship conflict. *Journal of Personality*, *70*(3), 311-344.

Bovaird, J. A., & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 495-511). New York, NY, US: The Guilford Press.

Bradley, B. H., Baur, J. E., Banford, C. G., & Postlethwaite, B. E. (2013). Team players and collective performance: How agreeableness affects team performance over time. *Small Group Research*, *44*(6), 680-711.

Bradley, B. H., Klotz, A. C., Postlethwaite, B. E., & Brown, K. G. (2013). Ready to rumble: How team personality composition and task conflict interact to improve performance. *Journal of Applied Psychology, 98*(2), 385-392.

Bradley, B. H., Postlethwaite, B. E., Klotz, A. C., Hamdani, M. R., & Brown, K. G. (2012). Reaping the benefits of task conflict in teams: The critical role of team psychological safety climate. *Journal of Applied Psychology*, *97*(1), 151-158.

Brandyberry, A. A., & Bakke, S. A. (2006). Mitigating negative behaviors in student project teams: An information technology solution. *Journal of Information Systems Education, 17*(2), 195-209.

Bravo-Rivera, C., & Sotres-Bayon, F. (2020). From isolated emotional memories to their competition during conflict. *Frontiers in Behavioral Neuroscience, 14*, 36.

Byington, E. (2012, July). OCEAN Faultlines: An Investigation of Personality-Based Subgroups in Teams. In *Academy of Management Proceedings* (Vol. 2012, No. 1, p. 17033). Briarcliff Manor, NY 10510: Academy of Management.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research*, *1*(2), 99-103.

Cameron, D. (2000). *Good to Talk? Living and Working in a Communication Culture.* Thousand Oaks, CA: Sage.

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17-32.

Carlson, S. R., & Iacono, W. G. (2006). Heritability of P300 amplitude development from adolescence to adulthood. *Psychophysiology*, *43*(5), 470-480.

Chan, D. (1998a). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*(4), 421-483.

Chan, D. (1998b). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*(2), 234-246.

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

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.

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.

Collins, C. G., Gibson, C. B., Quigley, N. R., & Parker, S. K. (2016). Unpacking team dynamics with growth modeling: An approach to test, refine, and integrate theory. *Organizational Psychology Review*, *6*(1), 63-91.

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., García-Peñalvo, F. J., & Tovar, E. (2014). Project managers in global software development teams: A study of the effects on productivity and performance. *Software Quality Journal*, *22*(1), 3-19.

Costa, P. L., Passos, A. M., & Bakker, A. B. (2015). Direct and contextual influence of team conflict on team resources, team work engagement, and team performance. *Negotiation and Conflict Management Research*, *8*(4), 211-227.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Curşeu, P. L., & Schruijer, S. G. (2010). Does conflict shatter trust or does trust obliterate conflict? Revisiting the relationships between team diversity, conflict, and trust. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 66-79.

De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, *88*(4), 741-749.

De Wit, F. R., & Greer, L. L. (2008, August). The black-box deciphered: a meta-analysis of team diversity, conflict, and team performance. In *Academy of Management Proceedings* (Vol. 2008, No. 1, pp. 1-6). Briarcliff Manor, NY 10510: Academy of Management.

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

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245-260.

Driskell, J. E., Hogan, R., & Salas, E. (1987). Personality and group performance. In C. Hendrick (Ed.), *Review of personality and social psychology, Vol. 9. Group processes and intergroup relations* (pp. 91-112). Thousand Oaks, CA, US: Sage Publications, Inc.

Eagly, A. H. (2016). When passionate advocates meet research on diversity, does the honest broker stand a chance? *Journal of Social Issues*, *72*(1), 199-222.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350-383.

Espinoza, J. A., Daljeet, K. N., & Meyer, J. P. (2020). Establishing the structure and replicability of personality profiles using the HEXACO-PI-R. *Nature Human Behaviour*, *4*(7), 713-724.

Esquivel, M. A., & Kleiner, B. H. (1996). The importance of conflict in work team effectiveness. *Team Performance Management: An International Journal*, *2*(3), 42-48.

Forza & Filippini, R. (1998). TQM impact on quality conformance and customer satisfaction: A causal model. *International Journal of Production Economics*, *55*(1), 1–20.

Gallo, M. C. (2017). *The Impact of Need for Affect and Personality on Relationship Conflict in Groups* (Doctoral dissertation, The Chicago School of Professional Psychology).

Garrison, G., Wakefield, R. L., Xu, X., & Kim, S. H. (2010). Globally distributed teams: The effect of diversity on trust, cohesion and individual performance. *ACM SIGMIS Database: The Database for Advances in Information Systems*, *41*(3), 27-48.

Gayle, B. M., & Preiss, R. W. (1998). Assessing emotionality in organizational conflicts. *Management Communication Quarterly, 12*(2), 280-302.

Gbadamosi, O., Baghestan, A. G., & Al-Mabrouk, K. (2014). Gender, age and nationality: Assessing their impact on conflict resolution styles. *Journal of Management Development*, *33*(3), 245-257.

Gersick, C. J. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, *31*(1), 9-41.

Gersick, C. J. (1989). Marking time: Predictable transitions in task groups. *Academy of Management Journal*, *32*(2), 274-309.

Gersick, C. J. (1991). Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review*, *16*(1), 10-36.

Gevers, J. M., Rutte, C. G., & Van Eerde, W. (2006). Meeting deadlines in work groups: Implicit and explicit mechanisms. *Applied Psychology*, *55*, 52-72.

Gibson, C., & Vermeulen, F. (2003). A healthy divide: Subgroups as a stimulus for team learning behavior. *Administrative Science Quarterly*, *48*(2), 202-239.

Gould, S. J., & Eldredge, N. (1986). Punctuated equilibrium at the third stage. *Systematic Zoology, 35*(1), 143-148.

Goyal, S., Maruping, L., & Robert, L. (2008, August). Diversity and conflict in teams: A faultline model perspective. In *Academy of Management Proceedings* (Vol. 2008, No. 1, pp. 1-6). Briarcliff Manor, NY 10510.

Gratton, L., Voigt, A., & Erickson, T. J. (2007). Bridging faultlines in diverse teams. *MIT Sloan Management Review, 48*(4), 22-29.

Graziano, W. G., Jensen-Campbell, L. A., & Hair, E. C. (1996). Perceiving interpersonal conflict and reacting to it: The case for agreeableness. *Journal of Personality and Social Psychology, 70*(4), 820-835.

Greer, L. L., Caruso, H. M., & Jehn, K. A. (2011). The bigger they are, the harder they fall: Linking team power, team conflict, and performance. *Organizational Behavior and Human Decision Processes, 116*(1), 116-128.

Greer, L. L., & Jehn, K. A. (2007, August). Where perception meets reality: The effects of different types of faultline perceptions, asymmetries, and realities on intersubgroup conflict and workgroup outcomes. In *Academy of Management Proceedings* (Vol. 2007, No. 1, pp. 1-6). Briarcliff Manor, NY 10510.

Greer, L. L., Jehn, K. A., & Mannix, E. A. (2008). Conflict transformation: A longitudinal investigation of the relationships between different types of intragroup conflict and the moderating role of conflict resolution. *Small Group Research*, *39*(3), 278-302.

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology, 87*(5), 819-832.

Guenter, H., van Emmerik, H., Schreurs, B., Kuypers, T., van Iterson, A., & Notelaers, G. (2016). When task conflict becomes personal: The impact of perceived team performance. *Small Group Research*, *47*(5), 569-604.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis: A Global Perspective,* 7th Edition. New Jersey: Pearson Educational Inc.

Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). *Human judgment and decision making: Theories, methods, and procedures*. Praeger Publishers, Westport, Connecticut.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning. *Academy of Management Journal, 45*(5), 1029-1045.

Hideg, I., & Adair, W. L. (2010). Is Conflict always Detrimental for Group Performance: The Case of Faultline Groups. Paper presented at the *23rd Annual IACM Conference.*

Hix-Small, H., Duncan, T. E., Duncan, S. C., & Okut, H. (2004). A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology and Behavioral Assessment*, *26*(4), 255-270.

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.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Hüffmeier, J., & Hertel, G. (2011). Many cheers make light the work: How social support triggers process gains in teams. *Journal of Managerial Psychology, 26*(3), 185-204.

Humphrey, S. E., & Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals, 8*(1), 443-503.

Humphrey, S. E., Aime, F., Cushenbery, L., Hill, A. D., & Fairchild, J. (2017). Team conflict dynamics: Implications of a dyadic view of conflict for team performance.

Organizational Behavior and Human Decision Processes, 142, 58-70.

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*, 517-543.

Ilies, R., Johnson, M. D., Judge, T. A., & Keeney, J. (2011). A within-individual study of interpersonal conflict as a work stressor: Dispositional and situational moderators. *Journal of Organizational Behavior*, *32*(1), 44-64.

Jak, S. (2018). Cross-Level Invariance in Multilevel Factor Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 1-16.

Jak, S., & Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Frontiers in Psychology*, *8*, 1640-1649.

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly, 40*(2), 256-282.

Jehn, K. A. (1997). A qualitative analysis of conflict types and dimensions in organizational groups. *Administrative Science Quarterly, 42*(3), 530-557.

Jehn, K. A., & Bezrukova, K. (2010). The faultline activation process and the effects of activated faultlines on coalition formation, conflict, and group outcomes. *Organizational Behavior and Human Decision Processes, 112*(1), 24-42.

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, *44*(2), 238-251.

Jehn, K., Rispens, S., Jonsen, K., & Greer, L. (2013). Conflict contagion: a temporal perspective on the development of conflict within teams. *International Journal of Conflict*

Management, *24*(4), 352-373.

Jehn, K. A., Rupert, J., & Nauta, A. (2006). The effects of conflict asymmetry on mediation outcomes. *International Journal of Conflict Management, 17*(2), 96-109.

Joseph, D. L., Jin, J., Newman, D. A., & O'Boyle, E. H. (2015). Why does self-reported emotional intelligence predict job performance? A meta-analytic investigation of mixed EI. *Journal of Applied Psychology*, *100*(2), 298-342.

Jung, M. F., Martelaro, N., & Hinds, P. J. (2015, March). Using robots to moderate team conflict: The case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 229-236).

Kabat-Zinn, J. (2005). *Wherever You Go There You Are: Mindfulness Meditation in Everyday Life*. New York, NY: Hyperion.

Keener, E., Strough, J., & DiDonato, L. (2012). Gender differences and similarities in strategies for managing conflict with friends and romantic partners. *Sex Roles*, *67*(1-2), 83-97.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling.* Guilford Publications.

Korsgaard, M. A., Jeong, S., Mahony, D. M., & Pitariu, A. H. (2008). A multilevel view of intragroup conflict. *Journal of Management*, *34*(6), 1222-1252.

Kristjansson, S. D., Kircher, J. C., & Webb, A. K. (2007). Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. *Psychophysiology*, *44*(5), 728-736.

Kurtzberg, T. R., & Mueller, J. S. (2005). The influence of daily conflict on perceptions of creativity: A longitudinal study. *International Journal of Conflict Management*, *16*(4), 335-353.

Labianca, G., Brass, D. J., & Gray, B. (1998). Social networks and perceptions of intergroup conflict: The role of negative relationships and third parties. *Academy of Management Journal, 41*(1), 55-67.

Labianca, G., Moon, H., & Watt, I. (2005). When is an hour not 60 minutes? Deadlines, temporal schemata, and individual and task group performance. *Academy of Management Journal*, *48*(4), 677-694.

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

Lang, J. W., Bliese, P. D., & de Voogt, A. (2018). Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*, *71*(2), 255-281.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review*, *23*(2), 325-340.

Laver, J. (2011). Linguistic routines and politeness in greeting and parting. In *Conversational Routine* (pp. 289-304). De Gruyter Mouton.

Lee, K., & Ashton, M. C. (2020). Sex differences in HEXACO personality characteristics across countries and ethnicities. *Journal of Personality*, *88*(6), 1075-1090.

Lee, H. W., Choi, J. N., & Kim, S. (2018). Does gender diversity help teams constructively manage status conflict? An evolutionary perspective of status conflict, team psychological safety, and team creativity. *Organizational Behavior and Human Decision Processes, 144*, 187-199.

Li, J. (2011). *Time in teams: Methodological issues in the study of temporal dynamics*. Dissertation from Maastricht University.

Li, J., & Roe, R. A. (2012). Introducing an intrateam longitudinal approach to the study of team process dynamics. *European Journal of Work and Organizational Psychology*, *21*(5), 718-748.

Lindsjørn, Y., Sjøberg, D. I., Dingsøyr, T., Bergersen, G. R., & Dybå, T. (2016). Teamwork quality and project success in software development: A survey of agile development teams. *Journal of Systems and Software*, *122*, 274-286.

Liu, D., Mitchell, T., Lee, T., Holtom, B., & Hinkin, T. (2012). When employees are out of step with coworkers: How job satisfaction trajectory and dispersion influence individual- and unit-level voluntary turnover. *Academy of Management Journal*, *55*, 1360–1380.

Liu, S., Xin, H., Shen, L., He, J., & Liu, J. (2020). The influence of individual and team mindfulness on work engagement. *Frontiers in Psychology*, *10*:2928.

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *The Journal of Social Psychology, 137*(3), 369-373.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*(5), 951-966.

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A century of work teams in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 452-467.

McGrath, J. E. (1984). Groups: Interaction and performance. Englewood Cliffs, N.J: Prentice-Hall.

Meyer, B., & Glenz, A. (2013). Team faultline measures: A computational comparison and a new approach to multiple subgroups. *Organizational Research Methods*, *16*(3), 393-424.

Meyer, B., Shemla, M., & Schermuly, C. C. (2011). Social category salience moderates the effect of diversity faultlines on information elaboration. *Small Group Research, 42*(3), 257-282.

Michinov, E., & Michinov, N. (2007). Identifying a transition period at the midpoint of an online collaborative activity: A study among adult learners. *Computers in Human Behavior*, *23*(3), 1355-1371.

Miner Jr, F. C. (1984). Group versus individual decision making: An investigation of performance measures, decision strategies, and process losses/gains. *Organizational Behavior and Human Performance*, *33*(1), 112-124.

Misangyi, V. F., LePine, J. A., Algina, J., & Goeddeke Jr, F. (2006). The adequacy of repeated-measures regression for multilevel research: Comparisons with repeated-measures ANOVA, multivariate repeated-measures ANOVA, and multilevel modeling across various multilevel research designs. *Organizational Research Methods*, *9*(1), 5-28.

Mohammed, S., & Angell, L. C. (2004). Surface-and deep-level diversity in workgroups: Examining the moderating effects of team orientation and team process on relationship conflict. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 25*(8), 1015-1039.

Molleman, E. (2005). Diversity in demographic characteristics, abilities and personality traits: Do faultlines affect team functioning? *Group Decision and Negotiation*, *14*(3), 173-193.

Moshagen, M., Thielmann, I., Hilbig, B. E., & Zettler, I. (2019). Meta-analytic investigations of the HEXACO Personality Inventory (-Revised). *Zeitschrift für Psychologie, 227*, 186-194.

Montoya-Weiss, M. M., Massey, A. P., & Song, M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal*, *44*(6), 1251-1262.

Muthén, B. (2001). Latent variable mixture modeling. *New developments and techniques in structural equation modeling*, *2*, 1-33.

Muthén, B., Brown, C. H., Khoo, S., Yang, C., & Jo, B. (1998). General growth mixture modeling of latent trajectory classes: Perspectives and prospects. In *meeting of the Prevention Science and Methodology Group*, Tempe, AZ.

Muthén, L. K., & Muthén, B. (2019). Mplus. *The Comprehensive Modelling Program for Applied Researchers: User's Guide*, *5*.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463-469.

Nagin, D. S. (1999). Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychological Methods*, *4*(2), 139-157.

Neuman, G. A., Wagner, S. H., & Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management*, *24*(1), 28-45.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535-569.

Okhuysen, G. A., & Waller, M. J. (2002). Focusing on midpoint transitions: An analysis of boundary conditions. *Academy of Management Journal*, *45*(5), 1056-1065.

O'Neill, T. A., Allen, N. J., & Hastings, S. E. (2013). Examining the "pros" and "cons" of team conflict: A team-level meta-analysis of task, relationship, and process conflict. *Human*

Performance, *26*(3), 236-260.

O'Neill, T. A., & McLarnon, M. J. (2018). Optimizing team conflict dynamics for high performance teamwork. *Human Resource Management Review*, *28*(4), 378-394.

O’Neill, T. A., McLarnon, M. J., Hoffart, G. C., Woodley, H. J., & Allen, N. J. (2018). The structure and function of team conflict state profiles. *Journal of Management*, *44*(2), 811-836.

Owens, L., Daly, A., & Slee, P. (2005). Sex and age differences in victimisation and conflict resolution among adolescents in a South Australian school. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *31*(1), 1-12.

Park, S., Mathieu, J. E., & Grosser, T. J. (2020). A network conceptualization of team conflict. *Academy of Management Review*, *45*(2), 352-375.

Parker, P. A., & Kulik, J. A. (1995). Burnout, self-and supervisor-rated job performance, and absenteeism among nurses. *Journal of Behavioral Medicine*, *18*(6), 581-599.

Pearce, C. L., Gallagher, C. A., & Ensley, M. D. (2002). Confidence at the group level of analysis: A longitudinal investigation of the relationship between potency and team effectiveness. *Journal of Occupational and Organizational Psychology*, *75*(1), 115-119.

Pearsall, M. J., Ellis, A. P., & Evans, J. M. (2008). Unlocking the effects of gender faultlines on team creativity: Is activation the key? *Journal of Applied Psychology, 93*(1), 225-234.

Peeters, M. A., Van Tuijl, H. F., Rutte, C. G., & Reymen, I. M. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality: Published for the European Association of Personality Psychology, 20*(5), 377-396.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative Science Quarterly*, *44*(1), 1-28.

Peterson, R. S., & Behfar, K. J. (2003). The dynamic relationship between performance feedback, trust, and conflict in groups: A longitudinal study. *Organizational Behavior and Human Decision Processes*, *92*(1-2), 102-112.

Phillips, K. W., & Loyd, D. L. (2006). When surface and deep-level diversity collide: The effects on dissenting group members. *Organizational Behavior and Human Decision Processes, 99*(2), 143-160.

Qu, X., & Liu, X. (2017). Informational faultlines, integrative capability, and team creativity. *Group & Organization Management*, *42*(6), 767-791.

Randel, A. E. (2002). Identity salience: A moderator of the relationship between group gender composition and work group conflict. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *23*(6), 749-766.

Ringle, C. (2016). *SimPLS-The PLS Agent* (Version 2). CoMSES Computational Model Library.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103-124.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373.

Rico, R., Molleman, E., Sánchez-Manzanares, M., & Van der Vegt, G. S. (2007). The effects of diversity faultlines and team task autonomy on decision quality and social integration. *Journal of Management*, *33*(1), 111-132.

Rispens, S. (2012). The influence of conflict issue importance on the co-occurrence of task and relationship conflict in teams. *Applied Psychology*, *61*(3), 349-367.

Rousseau, J., & Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(5), 689-710.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, *50*(3), 540-547.

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, *54*(4), 599-622.

Shah, P. P., Peterson, R. S., Jones, S. L., & Ferguson, A. J. (2020). Things are not always what they seem: The origins and evolution of intragroup conflict. *Administrative Science Quarterly*, https://doi.org/10.1177/0001839220965186.

Sinha, R., Janardhanan, N. S., Greer, L. L., Conlon, D. E., & Edwards, J. R. (2016). Skewed task conflicts in teams: What happens when a few members see more conflict than the rest? *Journal of Applied Psychology*, *101*(7), 1045-1056.

Sivasubramaniam, N., Murry, W. D., Avolio, B. J., & Jung, D. I. (2002). A longitudinal model of the effects of team leadership and group potency on group performance. *Group & Organization Management*, *27*(1), 66-96.

Small, L., Kenny, L., & Bryant, R. A. (2011). The cost in remembering of ruminating on negative memories. *Emotion, 11*(6), 1434–1438.

Somech, A., Desivilya, H. S., & Lidogoster, H. (2009). Team conflict management and team effectiveness: The effects of task interdependence and team identification. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 30*(3), 359-378.

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.

Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, *41*(4), 690-709.

Svensson, E. (1998). Ordinal invariant measures for individual and group changes in ordered categorical data. *Statistics in Medicine*, *17*(24), 2923-2936.

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology*, *90*(4), 692–709.

Thatcher, S., & Patel, P. C. (2011). Demographic faultlines: A meta-analysis of the literature. *Journal of Applied Psychology*, *96*(6), 1119-1139.

Thatcher, S. M., & Patel, P. C. (2012). Group faultlines: A review, integration, and guide to future research. *Journal of Management*, *38*(4), 969-1009.

Tekleab, A. G., & Quigley, N. R. (2014). Team deep-level diversity, relationship conflict, and team members' affective reactions: A cross-level investigation. *Journal of Business Research, 67*(3), 394-402.

Tekleab, A. G., Quigley, N. R., & Tesluk, P. E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group & Organization*

Management, *34*(2), 170-205.

Terkourafi, M. (2005). Beyond the micro-level in politeness research. *Journal of Politeness Research*, 1(2), 237-262.

Todorova, G., Bear, J. B., & Weingart, L. R. (2014). Can conflict be energizing? A study of task conflict, positive emotions, and job satisfaction. *Journal of Applied Psychology*, *99*(3), 451-467.

Tripp, J. F., Riemenschneider, C., & Thatcher, J. B. (2016). Job satisfaction in agile development teams: Agile development as work redesign. *Journal of the Association for Information Systems*, *17*(4), 267-307.

Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European Journal of Social Psychology*, *9*(2), 187-204.

Twenge, J. M. (2001). Birth cohort changes in extraversion: A cross-temporal meta-analysis, 1966–1993. *Personality and Individual Differences*, *30*(5), 735-748.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207-232.

van der Kamp, M., Tjemkes, B. V., & Jehn, K. A. (2011, June). Faultline activation and deactivation and their effect on conflict. In *IACM 24th Annual Conference Paper*.

Van Dijk, H., Van Engen, M. L., & Van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*(1), 38-53.

Van Knippenberg, D., Dawson, J. F., West, M. A., & Homan, A. C. (2011). Diversity faultlines, shared objectives, and top management team performance. *Human Relations*, *64*(3), 307-

336.

van Woerkom, M., & Van Engen, M. L. (2009). Learning from conflicts? The relations between task and relationship conflicts, team learning and team performance. *European Journal of Work and Organizational Psychology*, *18*(4), 381-404.

Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, *10*(4), 635-656.

Wang, G., Jing, R., & Klossek, A. (2007). Antecedents and management of conflict: Resolution styles of Chinese top managers in multiple rounds of cognitive and affective conflict. *International Journal of Conflict Management*, *18*(1), 74-97.

Wanic, R., & Kulik, J. (2011). Toward an understanding of gender differences in the impact of marital conflict on health. *Sex Roles*, *65*(5-6), 297-312.

Watson, W., Cooper, D., Torres, M. J. L. N., & Boyd, N. G. (2008). Team processes, team conflict, team outcomes, and gender: An examination of US and Mexican learning teams. *International Journal of Intercultural Relations*, *32*(6), 524-537.

Watson, W. E., Johnson, L., Kumar, K., & Critelli, J. (1998). Process gain and process loss: Comparing interpersonal processes and performance of culturally diverse and non-diverse teams across time. *International Journal of Intercultural Relations*, *22*(4), 409-430.

Weber, B., & Hertel, G. (2007). Motivation gains of inferior group members: A meta-analytical review. *Journal of Personality and Social Psychology*, *93*(6), 973–993.

Wei, X., Liu, Y., & Chen, S. (2015). A meta-analysis of the relationship between team demographic diversity and team performance. *Acta Psychologica Sinica*, *47*(9), 1172-

1187.

Weingart, L. R., Behfar, K. J., Bendersky, C., Todorova, G., & Jehn, K. A. (2015). The directness and oppositional intensity of conflict expression. *Academy of Management Review*, *40*(2), 235-262.

West, C., Landry, K., Graham, A., Graham, L., Cianciolo, A. T., Kalet, A., ... & Sherman, D. W. (2015). Conceptualizing interprofessional teams as multi-team systems—Implications for assessment and training. *Teaching and Learning in Medicine*, *27*(4), 366-369.

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.

Woodin, E. M. (2011). A two-dimensional approach to relationship conflict: Meta-analytic findings. *Journal of Family Psychology, 25*(3), 325–335.

Woolley, A. W. (1998). Effects of intervention content and timing on group task performance. *The Journal of Applied Behavioral Science*, *34*(1), 30-46.

Yu, L., & Zellmer-Bruhn, M. (2018). Introducing team mindfulness and considering its safeguard role against conflict transformation and social undermining. *Academy of Management Journal*, *61*(1), 324-347.

Appendix A

Table 29. Measures used in Studies 1-3.

Item	Measure
I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.	Honesty-Humility
If I want something from someone, I will laugh at that person's worst jokes.*	Honesty-Humility
I wouldn’t pretend to like someone just to get that person to do favours for me.	Honesty-Humility
If I knew that I could never get caught, I would be willing to steal a million dollars.*	Honesty-Humility
I would never accept a bribe, even if it were very large.	Honesty-Humility
I’d be tempted to use counterfeit money, if I were sure I could get away with it.*	Honesty-Humility
Having a lot of money is not especially important to me.	Honesty-Humility
I would get a lot of pleasure from owning expensive luxury goods.*	Honesty-Humility
I think that I am entitled to more respect than the average person is.*	Honesty-Humility
I want people to know that I am an important person of high status.*	Honesty-Humility
I would feel afraid if I had to travel in bad weather conditions.	Emotionality
I sometimes can’t help worrying about little things.	Emotionality
When I suffer from a painful experience, I need someone to make me feel comfortable.	Emotionality
I feel like crying when I see other people crying.	Emotionality
When it comes to physical danger, I am very fearful.	Emotionality
I worry a lot less than most people do.*	Emotionality
I can handle difficult situations without needing emotional support from anyone else.*	Emotionality
I feel strong emotions when someone close to me is going away for a long time.	Emotionality
Even in an emergency I wouldn’t feel like panicking.*	Emotionality
I remain unemotional even in situations where most people get very sentimental.*	Emotionality
I feel reasonably satisfied with myself overall.	Extraversion
I rarely express my opinions in group meetings.*	Extraversion
I prefer jobs that involve active social interaction to those that involve working alone.	Extraversion
On most days, I feel cheerful and optimistic.	Extraversion
I feel that I am an unpopular person.*	Extraversion
In social situations, I’m usually the one who makes the first move.	Extraversion
The first thing that I always do in a new place is to make friends.	Extraversion
Most people are more upbeat and dynamic than I generally am.*	Extraversion
I sometimes feel that I am a worthless person.*	Extraversion

Table 29. *(continued).*

When I'm in a group of people, I'm often the one who speaks on behalf of the group.*	Extraversion
I rarely hold a grudge, even against people who have badly wronged me.	Agreeableness
People sometimes tell me that I am too critical of others.*	Agreeableness
People sometimes tell me that I'm too stubborn.*	Agreeableness
People think of me as someone who has a quick temper.*	Agreeableness
My attitude toward people who have treated me badly is "forgive and forget."	Agreeableness
I tend to be lenient in judging other people.	Agreeableness
I am usually quite flexible in my opinions when people disagree with me.	Agreeableness
Most people tend to get angry more quickly than I do.	Agreeableness
Even when people make a lot of mistakes, I rarely say anything negative.	Agreeableness
When people tell me that I'm wrong, my first reaction is to argue with them.*	Agreeableness
I plan ahead and organize things, to avoid scrambling at the last minute.	Conscientiousness
I often push myself very hard when trying to achieve a goal.	Conscientiousness
When working on something, I don't pay much attention to small details.*	Conscientiousness
I make decisions based on the feeling of the moment rather than on careful thought.*	Conscientiousness
When working, I sometimes have difficulties due to being disorganized.*	Conscientiousness
I do only the minimum amount of work needed to get by.*	Conscientiousness
I always try to be accurate in my work, even at the expense of time.	Conscientiousness
I make a lot of mistakes because I don't think before I act.*	Conscientiousness
People often call me a perfectionist.	Conscientiousness
I prefer to do whatever comes to mind, rather than stick to a plan.*	Conscientiousness
I would be quite bored by a visit to an art gallery.*	Openness
I'm interested in learning about the history and politics of other countries.	Openness
I would enjoy creating a work of art, such as a novel, a song, or a painting.	Openness
I think that paying attention to radical ideas is a waste of time.*	Openness
If I had the opportunity, I would like to attend a classical music concert.	Openness
I've never really enjoyed looking through an encyclopedia.*	Openness
People have often told me that I have a good imagination.	Openness
I like people who have unconventional views.	Openness
I don't think of myself as the artistic or creative type.*	Openness
I find it boring to discuss philosophy.*	Openness
To what extent does your team argue the pros and cons of different opinions?	Task Conflict
How often do your team members discuss evidence for alternative viewpoints?	Task Conflict
How frequently do members of your team engage in debate about different opinions or ideas?	Task Conflict

Table 29. *(continued).*

How much friction is there among members of your team?	Relationship Conflict
How much are personality conflicts evident in your team?	Relationship Conflict
How much tension is there among team members?	Relationship Conflict
How much emotional conflict is there among team members?	Relationship Conflict
How frequently do your team members disagree about the optimal amount of time to spend on different parts of teamwork?	Logistical (Process) Conflict
How frequently do your team members disagree about the optimal amount of time to spend in meetings?	Logistical (Process) Conflict
How often do your team members disagree about who should do what?	Logistical (Process) Conflict
How often is there tension in your team caused by member(s) not performing as well as expected?	Contribution (Process) Conflict
To what extent is there tension in your team caused by member(s) not completing their assignment(s) on time?	Contribution (Process) Conflict
How much tension is there in your team caused by member(s) arriving late to team meetings?	Contribution (Process) Conflict
Compared to other teams in [course name], how would you rate your team's…efficiency?	Team Effectiveness
…quality of innovation?	Team Effectiveness
…goal attainment?	Team Effectiveness
…adherence to schedules?	Team Effectiveness
…overall performance?	Team Effectiveness

Note. Starred items are reverse-coded. HEXACO items were developed by Ashton and Lee (2009). Task and relationship conflict items were developed by Jehn (1995). Process conflict items were developed by Behfar and colleagues (2011). Team effectiveness items were developed by the TeamWork Lab.

Appendix B

Examples of evaluation criteria for team design projects

ES 1050 – Introductory Design and Innovation Studio   Design Project I – Creativity - 2013

Evaluation Criteria for Design Report:

Rubric category	Maximum points
Creativity/Innovation:	
• Many concepts generated (at least 3)	5
• Quality of concepts and degree of innovation	5
• Effective use of concept generation techniques	5
Problem Formulation:	
• Problem and need clearly defined	5
• Objectives and constraints	5
• Whole problem is considered	5
• Definition may lead to innovative solution	5
Effective Communication:	
• Clarity and proper sequence of information	5
• Good formatting and free of typo-graphical and grammatical errors	5
• Quality of sketches and diagrams	10
Design Process:	
• Using systematic engineering approach (identify need, concept generation, iteration etc.)	10
• Critical evaluation of generated concepts	15
• Proper design iteration	10
• Learning from mistakes and providing proper recommendations for future projects	10
Total	100

Evaluation Criteria for Design Report:

Rubric category	**Maximum points**
Design Description • Problem and need clearly defined • Objectives and constraints stated • Final design clearly and fully documented	5 5 20
Effective Communication: • Clarity and proper sequence of information • Good formatting and free of typo-graphical and grammatical errors • Quality of sketches and diagrams	5 5 10
Iteration: • Testing results clearly presented • Critical evaluation and interpretation of testing results • Proper design iteration – evidence of logical design revisions based on testing results • Learning from mistakes and providing proper recommendations for future projects	10 10 20 10
Total	100

Note: There is no grade distribution for Project II. All members are expected to contribute equally to the project.

Design Project III – Major Design Project Presentations

In-Studio Presentations

The design team will give a formal presentation in studio which fully describes the team's final design. The presentation should include the problem definition, specifications, criteria, constraints, and at least 3 novel approaches to the problem. The focus will be on the particulars of the final design. The presentation should be 7 minutes in duration, and all group members should participate. Powerpoint and/or other visual aids may be used.

Criteria	Points
Oral Presentation – Preparation and Mechanics: • Presenting confidently • Well organized • Graphics and material relate to topic • Using allotted time effectively	10 10 10 10
Oral presentation – Delivery: • Enthusiastic • Proper eye contact and body language • Proper voice tone • Coherence and effectiveness of answers to questions	5 5 5 5
Oral Presentation – Content: • Having knowledge and insightful understanding of the project • Accuracy and relevance of presented material	20 20
Total	100

Final Presentation (Design Showcase)

Each team will design and setup a persuasive tabletop display which will showcase the final design, the prototype and detail its significant features. A team of judges will assess the design effectiveness, innovation, thoroughness and quality of the presentation and prototype. Prizes will be awarded for the best display as determined by a vote by the attendees.

Criteria	Weight
1. Engineering Design Process The problem is clearly defined, project-specific objectives and constraints are identified and a suitable solution has been developed through iterative design.	10%
2. Engineering Science The application of engineering science in the development of the design should be clearly demonstrated. Engineering science may include design calculations, CAD modeling and analysis, and/or physical testing of the prototype or components.	10%
3. Prototype Each display must include a prototype which is appropriate to its intended purpose. It may be a scale model of a large system, a functioning component of the system, or a fully functioning design. Criteria for evaluation include: use of appropriate materials and construction methods, appropriate size, cost-effectiveness, minimal complexity, and effectiveness in achievement of the purpose.	30%
4. Visual Communication The display should effectively convey information about the project. The display should be well organized, well written and effectively formatted. Effective visual communication includes use of diagrams, engineering drawings, CAD renderings, photographs, graphs and plots, etc.	20%
5. Oral Communication The design team must describe the project and the prototype orally. The overall project objective, the purpose of the project and function of the prototype, must be clearly explained.	20%
6. Innovation The prototype and display should effectively highlight, demonstrate and communicate innovative concepts incorporated in the design.	10%
Total (out of 100)	

Appendix C

Full model syntax for Study 2.

```
TITLE: LGM study 2 2014 multi level apr 25
DATA: FILE IS s2apr22.dat; ! text file containing raw data in wide format
data widetolong:
   wide = T1_RC T2_RC T3_RC | T1_TC T2_TC T3_TC;
   long = RC | TC;
   idvariable = ID;
   repetition = time;
VARIABLE: NAMES =
ID studioteam Age Gender Ethnicity English O C A Ex Em HH T1_RC T1_TC T1_LC
T2_RC T2_TC T2_LC T3_RC T3_TC T3_LC teff dp3;
USEVARIABLES = Age Gender Ethnicity English hh Em Ex A C O teff RC TC time;
MISSING = ALL (-99);
CLUSTER IS studioteam; ! Level-2 grouping identifier
WITHIN is time;
ANALYSIS:
ESTIMATOR = MLR;
TYPE = TWOLEVEL RANDOM;
MODEL: ! model specification follows
%WITHIN%
RC TC teff on Age Gender Ethnicity English hh Em Ex A C O;
src | RC on time;
stc | TC on time;
teff on RC TC;
%BETWEEN%
src stc teff on Age Gender Ethnicity English hh Em Ex A C O;
teff on src stc;
OUTPUT: Tech1 Tech4 Tech8;
```

Curriculum Vita

Natasha E. Ouslis

Education

University of Western Ontario
Doctorate of Philosophy, Industrial/Organizational Psychology 2018-2021
Supervisor: Dr. Natalie J. Allen
Dissertation: Finding Teams that Fight Fair: Exploring Trajectories of Team Conflict Over Time

University of Western Ontario
Master of Science, Industrial/Organizational Psychology 2016-2018
Supervisor: Dr. Natalie J. Allen
Thesis: Testing a new model of team interdependence

University of Toronto
Honours Bachelor of Science, Psychology Research Specialist 2012-2016

Work Experience

Associate, McKinsey & Company 2021-Present

Founder, NEOconsulting 2019-2021

Organizational Behaviour Specialist, BEworks 2016-2018

Refereed Publications

Rajsic, J., **Ouslis, N.E.,** Wilson, D.E., & Pratt, J. (2017). Looking sharp: Becoming a search template boosts precision and stability in visual working memory. *Attention, Perception, & Psychophysics, 1-9.*

Lowe, M.X., Stevenson, R.A., Wilson, K.E., **Ouslis, N.E.**, Barense, M.D., Cant, J.S., & Ferber, S. (2016). Sensory processing patterns predict the bias of ensemble statistics for items held in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 294-301.

Conference Presentations

Patel, S., Konrad, A., **Ouslis, N.E.,** & Goh, K. (2020, October). *Glass Cliff or Speed Bump? Gender Diversity in U.S. Mutual Fund Management Teams: 1992 – 2016*. Paper presented at the Interdisciplinary Network for Group Research Conference, virtual.

Ouslis, N.E. & Allen, N.J. (2020, October). *The trajectory and impact of team conflict for design project teams* in the symposium *Conflict in Organizations across Time, Levels, and Methods* presented at the Interdisciplinary Network for Group Research Conference, virtual.

Patel, S., Konrad, A., **Ouslis, N.E.,** & Goh, K. (2020, August). *Glass cliff or speed bump? Gender diversity in U.S. mutual fund management teams: 1992 – 2016* in the symposium *Storming the Last Bastions: Women Entering High Prestige, Male-Dominated Occupations* presented at the Academy of Management Meeting, virtual.

Quinn, A., **Ouslis, N.E.,** & Allen, N.J. (2020, August). *A fear of failure fallacy? How team innovation beliefs and performance relate to fear of failure.* Paper presented at the Academy of Management Meeting, virtual.

Ouslis, N.E. & Allen, N.J. (2019, April). *Testing a new model of team interdependence using team conflict.* Poster at the Society for Industrial-Organizational Psychology Conference, National Harbor, MD.

Ouslis, N.E. & Allen, N.J. (2018, July). *Testing a new model of team interdependence.* Poster at the Interdisciplinary Network for Group Research Conference, Washington, DC.

Ouslis, N.E. & Allen, N.J. (2017, July). *Conceptual and measurement challenges: The case of task interdependence.* Poster Presentation at the Interdisciplinary Network for Group Research Conference, St. Louis, MO.

Allen, N.J., Stanley, D.J., Cameron, K., McMenamin, J., **Ouslis, N.E.,** Lee, H., & Woodley, H. (2017, July). *Group Performance: A 10-year Bibliometric Review of Conceptualizations and Assessment.* Talk at the Interdisciplinary Network for Group Research Conference, St. Louis, MO.

Ouslis, N.E. & Allen, N.J. (2017, June). *Team interdependence: Conceptual and measurement challenges.* Poster Presentation at the Canadian Psychological Association Meeting, Toronto, ON.

Rajsic, J., **Ouslis, N.E.,** Wilson, D.E., & Pratt, J. (2016, November). *Looking sharp: Becoming a search template boosts precision and stability in visual working memory.* Psychonomic Society Meeting, Boston, MA

Pereira, B.J., **Ouslis, N.E.,** & Spence, I. (2015, May). *Controlling exposure time in mental rotation reduces gender differences.* Poster Presentation at the Association for Psychological Science meeting, New York City, NY

Ouslis, N.E., Pereira, B.J., & Spence, I. (2015, March). *Gender Differences in Speed and Response Bias of Three-dimensional Mental Rotation*. Poster Presentation at the Women in Science and Engineering Conference, Toronto, ON

Lowe, M.X., Stevenson, R.A., Wilson, K.E., **Ouslis, N.E.**, Azimi, M., Barense, M.D., Cant, J.S., & Ferber, S. (2015, March). *Sensory Processing Patterns Predict the Bias of Ensemble Statistics for Items Held in Visual Working Memory*. Poster Presentation at the Cognitive Neuroscience Society, San Francisco, CA

Ouslis, N.E., Pereira, B.J., & Spence, I. (2015, February). *Gender Differences in Speed and Response Bias of Three-dimensional Mental Rotation*. Poster Presentation at the Lake Ontario Visionary Establishment, Niagara Falls, ON

Ouslis, N. E., Pereira, B. J., Jeong, J. Y., & Spence, I. (2014, July*). Attention and Visuo-spatial Working Memory in Mental Rotation*. Poster Presentation at the Canadian Society for

Brain, Behaviour, and Cognitive Sciences, Toronto, ON

Non-Refereed Publications

Ouslis, N.E. (2021). Do All Employees in the Same Roles Get Equal Compensation? *Orgnostic,* retrieved from https://blog.orgnostic.com/do-all-employees-in-the-same-roles-get-equal-compensation/.

Ouslis, N.E. (2021). What Is the Distribution of Women and Minorities Across Job Levels? *Orgnostic*, retrieved from https://blog.orgnostic.com/women-and-minorities-across-job-levels/.

Ouslis, N.E. (2021). How Long Does It Take to Hire and Onboard a Replacement? *Orgnostic,* retrieved from https://blog.orgnostic.com/replacement-hire-rate/.

Ouslis, N.E. (2021). Are You Retaining Top Performers? *Orgnostic*, retrieved from https://blog.orgnostic.com/talent-turnover-rate/.

Ouslis, N.E. (2021). Why Do Your Employees Leave? *Orgnostic*, retrieved from https://blog.orgnostic.com/why-do-your-employees-leave/.

Ouslis, N.E. (2021). How Many Bad Hires Did We Make? *Orgnostic,* retrieved from https://blog.orgnostic.com/bad-hires/.

Ouslis, N.E. (2021). Impact of Employee Wellbeing on Referral Rates. *Orgnostic,* retrieved from https://blog.orgnostic.com/employee-wellbeing-and-referral-rates/.

Ouslis, N.E. (2021). Why Do Our Offers Get Rejected? *Orgnostic*, retrieved from https://blog.orgnostic.com/offer-rejection-reasons/.

Ouslis, N.E. (2021). What Is Our Offer Acceptance Rate? *Orgnostic*, retrieved from https://blog.orgnostic.com/offer-acceptance-rate/.

Ouslis, N.E. (2021). What Is the Diversity Breakdown of Our Top Performers and High-Potential Employees? *Orgnostic*, retrieved from https://blog.orgnostic.com/what-is-the-diversity-breakdown-of-our-top-performers-and-high-potential-employees/.

Ouslis, N.E. (2021). How Many Critical Hires That We Make Are Women and Minorities? *Orgnostic*, retrieved from https://blog.orgnostic.com/how-many-critical-hires-that-we-make-are-women-and-minorities/.

Ouslis, N.E. (2021). Is Your Approach to Low Performers Hurting Your Other Employees? *Orgnostic*, retrieved from https://blog.orgnostic.com/is-your-approach-to-low-performers-hurting-your-other-employees/.

Ouslis, N.E. (2021). How Can We Visualize the Employee Lifecycle? *Orgnostic,* retrieved from https://blog.orgnostic.com/how-can-we-visualize-the-employee-lifecycle/.

Ouslis, N.E. (2021). Are Women and Minorities Promoted as Often as the Rest? *Orgnostic,* retrieved from https://blog.orgnostic.com/are-women-and-minorities-promoted-as-often-as-the-rest/.

Ouslis, N.E. (2021). What is Your Quality of Hire? *Orgnostic,* retrieved from https://blog.orgnostic.com/what-is-your-quality-of-hire/.

Ouslis, N.E. (2021). How Do Employee Referrals Compare to Other Sourcing Channels? *Orgnostic,* retrieved from https://blog.orgnostic.com/how-do-employee-referrals-compare-to-other-sourcing-channels/.

Ouslis, N.E. (2021). How Many Applicants on Average Do I Need to Make One Hire? *Orgnostic,* retrieved from https://blog.orgnostic.com/how-many-applicants-on-average-do-i-need-to-make-one-hire/.

Ouslis, N.E. (2020). How Important is Culture to Company Success? Five KPIs Tell Different Stories. *Orgnostic,* retrieved from https://blog.orgnostic.com/how-important-is-culture-to-company-success/.

Ouslis, N.E. (2020). Six Conditions for Team Success. *Orgnostic*, retrieved from https://blog.orgnostic.com/six-conditions-for-team-success/.

Ouslis, N.E. (2020). Burnout vs Wellbeing: How Fair Workplace Processes Drive Engagement. *Orgnostic*, retrieved from https://blog.orgnostic.com/burnout-vs-wellbeing/.

Ouslis, N.E. (2020). How Can Leaders Help Teams Succeed? *Orgnostic,* retrieved from https://blog.orgnostic.com/how-can-leaders-help-teams-succeed/.

Ouslis, N.E. (2020). How Can Investors Measure the Market Value of Leadership? *Orgnostic*, retrieved from https://blog.orgnostic.com/how-can-investors-measure-the-market-value-of-leadership/.

Ouslis, N.E. (2020). Growth vs Profitability Culture: Why the Same Culture Won't Work for Both Goals. *Orgnostic,* retrieved from https://blog.orgnostic.com/growth-vs-profitability-culture/.

Ouslis, N.E. & ElMakkaoui, Z. (2020). High-Potential Employee Programs can be Self-Fulfilling Prophecies. *The Decision Lab,* retrieved from https://thedecisionlab.com/insights/business/high-potential-employee-programs-can-be-self-fulfilling-prophecies-here-are-three-ways-firms-can-avoid-this-problem/.

ElMakkaoui, Z. & Ouslis, N.E. (2020). High-Potential Programs Can Help Some Employees and Hurt Others. *The Decision Lab,* retrieved from https://thedecisionlab.com/insights/business/high-potential-programs-can-help-some-employees-and-hurt-others-heres-how-we-can-design-a-fairer-system/.

Ouslis, N.E. (2020). The Diversity Solutions We Already Have. *PeopleScience*, retrieved from https://peoplescience.maritz.com/Articles/2020/The-Diversity-Solutions-We-Already-Have.

Ouslis, N.E. (2020). How to Find a Behavioural Design Job (During a Pandemic). *Behavioural Design Hub,* retrieved from https://medium.com/behavior-design-hub/how-to-find-a-behavioral-design-job-during-a-pandemic-d22ff18ba7a1.

Salzer, S., Ouslis, N.E., & van den Akker, M. (2020). Behavioural Science Graduate Guide. *Behavioural Design Hub*, retrieved from https://medium.com/behavior-design-

hub/behavioral-science-graduate-guide-d096e0866b64.

Ouslis, N.E. (2020). Five Ways to Design a Better Job for Yourself in the Age of Automation. *The Decision Lab*, retrieved from https://thedecisionlab.com/five-ways-to-design-a-better-job-for-yourself-in-the-age-of-automation/.

Ouslis, N.E. (2020). Don't Ask If Your Job Will Be Automated. Ask These Questions Instead. *The Decision Lab*, retrieved from https://thedecisionlab.com/dont-ask-if-your-job-will-be-automated-ask-these-questions-instead/.

Ouslis, N.E. (2020). Automation at Work Will Change Our Home Lives. *The Decision Lab*, retrieved from https://thedecisionlab.com/automation-at-work-will-change-our-home-lives/.

Ouslis, N.E. (2020). Working From Home Can Amp Up Your Team's Communication and Creativity. *The Decision Lab*, retrieved from https://thedecisionlab.com/how-working-from-home-can-amp-up-your-teams-communication-and-creativity/.

Ouslis, N.E. (2020). Want to Innovate? Stop Hiring the Safest Option. *The Decision Lab*, retrieved from https://thedecisionlab.com/want-to-innovate-stop-hiring-the-safest-option/.

Ouslis, N.E. (2020). Beware Dollar Store Behavioral Economics. *PeopleScience*, retrieved from https://peoplescience.maritz.com/Articles/2020/Beware-Dollar-Store-Behavioral-Economics.

Ouslis, N.E. & Apostolidis, S. (2019). Design for Team Innovation: How Leaders Build Psychological Safety. *The Conference Board of Canada*, retrieved from https://www.conferenceboard.ca/insights/blogs/design-for-team-innovation-how-leaders-create-psychological-safety.

Ouslis, N.E. & Apostolidis, S. (2019). Disrupting Diversity and Inclusion: The Promise of Behavioural Design. *The Conference Board of Canada*, retrieved from https://www.conferenceboard.ca/insights/blogs/disrupting-diversity-inclusion-the-promise-of-behavioural-design.

Ouslis, N.E. (2019). Make Change Stick with Behaviour Modeling. *ScienceForWork*, retrieved from https://scienceforwork.com/blog/training-make-change-stick/.

Ouslis, N.E. (2019). Trust in Leadership – One Key Factor During Organizational Change. *ScienceForWork*, retrieved from https://scienceforwork.com/blog/trust-in-leadership-change/.

Invited Talks

Civic Engagement with Behavioural Design. Invited guest lecture at the University of Toronto, October 29, 2020.

Designing a Team-Based Organization with People Analytics. Invited talk at the AnalyzeHR Virtual Conference, June 9, 2020.

Setting Teams up for Success: Building Team Collaboration Skills for Better Virtual and Face-to-Face Performance. Invited talk at the People Analytics and Future of Work Virtual Conference, April 28, 2020.

Closing the Intention-Action Gap: Using Behavioural Science for Social Good. Invited guest lecture at the University of Toronto, November 15, 2019.

How People Really Act: At Work, at the Polls, and More. Invited guest lecture at the University of Toronto, January 17, 2019.

Service

Staff Writer, The Decision Lab	February 2020–July 2021
Chief Behavioural Writer, Habit Weekly Newsletter	February 2020–July 2021
Director and Head of Partnerships, Science for Work	July 2018–Present
Graduate Advisor, Western Undergraduate Psychology Journal	Sept 2016–August 2020
Social and Recruitment Committee Member, Psychology Department	June 2018–August 2021
Colloquium Committee Member, Psychology Department	May 2017–April 2018
Graduate Student Judge, Western Student Research Conference	March 2017
Editor-in-Chief, Inkblot: Psychology Undergraduate Journal	Sept 2015–Sept 2016
Editor, Inkblot: Psychology Undergraduate Journal	Oct 2014–Sept 2015
Academic Coordinator, Psychology Students` Association	April 2014–April 2015

Honours and Awards

Joseph A. Bombardier Canada Graduate Scholarship–Doctoral ($105,000)	Sept 2018–Aug 2021
Ontario Graduate Scholarship ($15,000) - declined	Sept 2018–Aug 2019
Joseph A. Bombardier Canada Graduate Scholarship–Masters ($17,500)	Sept 2017–Aug 2018
Ontario Graduate Scholarship ($15,000) – declined	Sept 2017–Aug 2018
Ontario Graduate Scholarship ($15,000)	Sept 2016–Aug 2017
Douglas N. Jackson Memorial Award ($500), *Western University*	Sept 2016
Association for Psychological Science Student Travel Award ($440)	May 2015
Arts and Science Student Union Travel Award ($400), *University of Toronto*	February 2015
Gail Ferris Sheard Academic Scholarship ($787), *University College*	September 2015
Co-awarded (B. Pereira), PI (I. Spence): Undergraduate Research Fund ($1500)	2014–2015

Exploring Gender Differences in Shepard-Metzler Mental Rotation Performance

Joel Verwegen Undergraduate Research Award ($500), *Toronto Rehabilitation Institute* 2014

Dean's List, *University of Toronto* 2012–2016

Professional Memberships

Society for Industrial/Organizational Psychology (SIOP)

Interdisciplinary Network for Group Research (INGRoup)

Academy of Management (AOM)