Examination of Peer Contextual Performance Ratings in Co-located and Virtual Teams Using the Social Relations Model

by

Kristin Leigh Cullen

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2011

Keywords: social relations model, contextual performance, virtual teams, peer ratings

Approved by

Daniel J. Svyantek, Chair, Professor of Psychology
Ana M. Franco-Watkins, Assistant Professor of Psychology
Allison Jones-Farmer, Associate Professor of Management
Kevin W. Mossholder, Professor of Management

Abstract

Peer- and self-ratings are common components of performance appraisals for many organizations, including companies relying on team-based work. Self- and peer-ratings may represent the best way to monitor performance in both co-located and virtual teams. However, biases may affect performance ratings in the form of perceiver tendencies (e.g., leniency) or relationship-specific effects. In the current study, I examined team members' ratings of task performance (i.e., contribution to the team's project) and contextual performance. These behaviors are relevant to effective team performance, regardless of the particular team or the specific task. Using the social relations model, I examined the extent to which perceiver and target tendencies as well as unique perceiver-target relationships contribute to performance ratings. I also examined several other indicators of the types of processes that underlie peer performance ratings (i.e., reciprocity, true halo, halo bias, and self-other agreement). The results of this study provide further insight into the factors influencing peer performance ratings and whether the contribution of these factors to performance ratings differs for co-located and virtual teams.

Acknowledgments

I would like to express my sincere gratitude to the many people who guide and supported me throughout the dissertation process and my graduate school career. Special thanks to my advisor, Dr. Daniel Svyantek, for allowing me to explore my interests and for providing practical guidance and unyielding support. I would like to thank Dr. Allison Jones-Farmer and Dr. Kevin Mossholder for their helpful feedback on my research ideas and for their continually encouragement. Finally, I owe a special debt of gratitude to Dr. Ana Franco-Watkins and Dr. Bryan Edwards for their personal and professional mentorship. Your collective guidance has been invaluable.

My personal support system has been a source of strength without which I would not have had any measure of success. Thanks to all of my wonderful friends, who may not have always understood what I was doing or why, but who feigned interest nonetheless. It is rare to find someone who truly understands you – I am fortunate to have met that person during my time at Auburn University. Houston, Thank you for loving and supporting me. Finally, to my parents and my brother Joseph, you have supported me all my life and I am blessed to have a family who will always love and care for each other.

Table of Contents

List of Tables

List of Figures

Chapter One: Literature Review

Teams in the Workplace: Implications for Performance Management

Team-based work has a long history within organizations in the United States. The earliest use of teams was documented in the 1950s and 60s (Gibson & Kirkman, 1999) and by the 1990s, team-based work had become a reality for many employees (Osterman, 1994, Lawler, Mohrman, & Ledford, 1995). In organizations, teams consist of two or more employees working together towards a common goal in an interdependent, adaptive, and dynamic manner (Salas, Dickinson, Converse, & Tannenbaum, 1992). Researchers expect organizations to continue utilizing teams (Kozlowski & Bell, 2003) because this strategy capitalizes on the diverse skills, expertise, and experience of a team's members. Teams are able to execute coordinated actions that are required to operate and adapt in complex, dynamic environments (Salas, Stagl, & Burke, 2004).

A wide variety of teams exists in organizations (Salas, Burke, & Fowlkes, 2006). Teams vary in their permanence, their size, the reality of their members (e.g., some incorporate simulations or machines), their geographic distribution, and their leadership allocation (e.g., some are controlled by a hierarchy and others are self-managing). Despite their widespread use, questions remain regarding the application of traditional human resource practices to teams, especially when teams operate over distance and communicate through technology (Gibson & Kirkman, 1999). The prevailing use of teams as an organizational strategy makes team performance vital to organizational success. Performance management systems are designed to measure, monitor, and maximize performance. Unfortunately, many organizations do not use

performance management systems to assess team effectiveness (Salas et al., 2006) or diagnose the causes of ineffective teams (Kendall & Salas, 2004).[1] Teams require a special approach to performance management, including assessments at the individual, sub-team, and team level (Gibson & Kirkman, 1999; Klein & Kozlowski, 2000), and an expansion in the types of behaviors and the rater perspectives that are considered relevant (Murphy & Cleveland, 1995). This study will provide evidence regarding the usefulness of peer performance ratings as a form of performance measurement in teams.

A broader definition of performance: Taskwork and teamwork. The transition to team-based work has broadened the everyday activities of employees. Employees still engage in behaviors directly related to task objectives and operations (Morgan, Salas, & Glickman, 1993); however, employees also engage in behaviors that help teams function effectively by facilitating collective action (Cannon-Bowers, Tannenbaum, Salas & Volpe, 1995; Morgan et al., 1993). Thus, the actions of team members can be categorized as taskwork or teamwork (McIntyre & Salas, 1995). Both types of behaviors are essential and thus should be measured as part of team performance management systems (Cannon-Bowers & Salas, 1997; Gibson & Kirkman, 1999; Salas et al., 2006).

Researchers have proposed many different models of teamwork behavior (i.e., essential team processes). For example, the Team Work Components Model (TCM; Dickinson et al., 1992; Dickinson & McIntyre, 1997; Rosenstein, 1994) contains four essential components of team coordination: (1) communication, (2) mutual performance monitoring, (3) intrateam

[1] The terms performance management, performance review, and performance appraisal are often used interchangeably; however, assessment (appraisal) is only part of a larger performance management system (Banks & May, 1999; Salas, Weaver, & Smith-Jentsch, 2009).

feedback, and (4) backup behaviors. Salas, Sims, and Burke (2005) offered their big five of teamwork: (1) team leadership, (2) team orientation, (3) mutual performance monitoring, (4) backup behavior, and (5) adaptability. Common to both of these models are backup behaviors. Within the broader organizational literature, this type of extra-role behavior has been referred to as contextual performance (Borman & Motowidlo, 1993; Borman & Motowidlo, 1997; Motowidlo & Van Scotter, 1994), organizational citizenship behaviors (OCBs; Organ, 1988; Smith, Organ, & Near, 1983), and citizenship performance (Borman, 2004; Borman, Penner, Allen, & Motowidlo, 2001).[2] These types of behaviors are extra-role or discretionary and help to create an environment that will facilitate the completion of the work task.

Harris and Barnes-Farrell (1997) made a distinction between OCBs, which they contend are directed toward the organization and teamwork behaviors, which are directed toward the team. However, when employees work as a team, it is likely that many of their extra-role behaviors will be directed towards their teammates in order to facilitate the team's task and ultimately organizational performance. Teamwork behaviors have been described as discretionary because they are not included in the employee's job description or as part of the planned work assignment (George & Brief, 1992). Further, many studies of teamwork and team processes have examined behaviors that fit the classic definition of contextual performance, including helping behaviors (De Dreu & Van Vianen, 2001; Janz, Colquitt, & Noe, 1997), backup behaviors (McIntyre & Dickinson, 1992; McIntyre & Salas, 1995; Salas et al., 2005), supporting behaviors (Smith-Jentsch, Johnston, & Payne, 1998), and extra-role behaviors

[2] Researchers have demonstrated that these constructs are largely isomorphic and have used the terms interchangeably (Coleman & Borman, 2000; Conway, 1999; Motowidlo, 2000; Organ, 1997; Podsakoff, MacKenzie, Paine, & Bachrach, 2000; Hanson & Borman, 2006).

(Weldon, Jehn, & Pradhan, 1991). Behaviors such as filling in for a team member, correcting performance mistakes of another team member, helping teammates who fall behind in their work, and demonstrating dedication and perseverance, allow individuals, teams, and organizations to function effectively by supporting the context in which job tasks are performed (Marks, Mathieu, & Zaccaro, 2001; Podsakoff & MacKenzie, 1997; Porter et al., 2003; Reilly & Aronson, 2009). Thus, for all practical purposes teamwork behaviors and contextual performance refer to the same actions of team members.

Contextual performance behaviors are increasingly relevant for companies to consider because of the emphasis on team-based work (Fletcher, 2001; Tippins & Coverdale, 2009). This emphasis has been reflected by the incorporation of extra-role information into employee evaluations. For example, Borman and Motowidlo (1997) concluded that overall ratings of employee worth are influenced by contextual performance (Borman, White, & Dorsey, 1995; Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo, 1996). Contextual performance has also been incorporated into the competencies assessed in many 360-degree feedback systems (Beehr, Ivanitskaya, Hansen, Erofeev, & Gudanowski, 2001; Atwater, Ostroff, Yammmarino, & Fleenor, 1998; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998) and identified as a criterion for performance appraisal and selection research and practice (Borman & Motowidlo, 1993). Even when extra-role behaviors are not explicitly included in performance evaluations, peers and supervisors often take these types of behaviors into consideration (e.g., MacKenzie, Podsakoff, & Fetter, 1991; Werner, 1994, Van Dyne & LePine, 1998). For example, when employees work in teams, their level of teamwork has been found to influence performance evaluations (Harris & Barnes-Farrell, 1997). Levy and Williams (2004) proposed that performance evaluations in a team context should be broad enough to include nontraditional performance criteria. These

assessments should include appraisals of both taskwork and teamwork competencies (Salas et al., 2005; Porter et al., 2003; Murphy & Cleveland, 1995; Salas, Weaver, Rosen, & Smith-Jentsch, 2009). Thus, in the current study I examine team members' contribution to the team's project and contextual performance—two types of behavior that are relevant to effective team performance, regardless of the particular team or the specific task (Salas et al., 2006).

Change in perspectives considered for performance appraisals. Performance monitoring and evaluation is a vital human resource function often fulfilled by the leader or supervisor of a team (Mathieu, Maynard, Rapp, & Gilson, 2008). However, Mathieu et al. suggested that leadership functions, including performance assessment, should be performed by the most appropriate person. Reductions in the managerial workforce, globalization of business, the popularity of multisource performance and feedback systems, and the implementation of flatter and team-based work structures have made traditional, top-down performance appraisal more difficult by limiting supervisors' day-to-day contact with employees (e.g., Barnes-Farrell, 2001; Brutus & Derayeh, 2002; Conway, Lombardo, & Sanders, 2001; Drexler, Beehr, & Stetz, 2001; Fletcher, 2001; Gibson & Kirkman, 1999; Greguras, Robie, & Born, 2001; London & Smither, 1995; London & Tornow, 1998; Reilly & McGourty, 1998; Saavedra & Kwun, 1993; Viswesvaran, Schmidt, & Ones, 2002). Thus, organizations have turned to alternative sources (e.g., peers, subordinates, customers) and these perspectives account for incremental validity over supervisor ratings (Conway et al., 2001).

There are generally low levels of agreement between performance ratings provided by different sources (e.g., supervisors, subordinates, peers, self; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988) and a large rater factor exists in multitrait–multirater studies (i.e., a large proportion of variance in performance ratings is attributed to the role of the rater; Klimoski &

London, 1974; Scullen, Mount, & Goff, 2000). However, these findings are reasonable to expect (Bozeman, 1997; Lance, Baxter, & Mahan, 2006) because employees at different levels of the organizational hierarchy observe different types of performance behaviors (Conway & Huffcut, 1997; Lance et al., 2006), have different goals when completing performance evaluations (e.g., identifying weakness, providing fair ratings; Murphy, Cleveland, Skattebo, & Kinney, 2004), and provide ratings based on the workplace behaviors that they feel are most relevant (Pulakos, Schmitt, & Chan, 1996). Thus, ratings from different perspectives offer valid information regarding employee performance and address many of the limitations of traditional performance appraisals (Fletcher, 2001; Barnes-Farrell, 2001).

The transition to team-based work has intensified the daily interactions of coworkers (Druskat & Wolff, 1999; Toegel & Conger, 2003) and shifted discretion over how the work is performed and responsibility for monitoring performance to the team members (Banks & May, 1999). This is especially true in self-managed teams (SMTs)—leaderless groups that are given the autonomy to determine how their work is completed, including determining their pace of work, selection, recruitment, training (Bishop & Scott, 2000; Manz & Sims, 1987). The use of SMTs represents a major trend in team-based work. As of the 1997 census, over 34% of U.S. companies reported using SMTs (Blasi & Kruse, 2006). In Fortune 1000 companies, this percentage was much higher (i.e., 68%; Lawler, Mohrman, & Ledford, 1994). Participation in SMTs has been linked to increased performance (Guzzo & Dickson, 1996; Cohen & Bailey, 1997) and many other positive outcomes for employees (e.g., increased decision autonomy, feelings of employment security, and job satisfaction; Batt, 2004). Leadership functions in SMTs are shared (Carson, Tesluk, & Marrone, 2007); including the responsibility for monitoring their own work (Bell & Kozlowski, 2002). The members of SMTs are encouraged to develop their

own performance standards, conduct self-evaluations, and regulate their own behaviors in order to fulfill the performance management function of leadership (Manz & Sims, 1986). Peer performance ratings may represent an additional option, beyond self-ratings for performance monitoring and evaluation in SMTs (Saavedra, & Kwun, 1993; Bretz, Miklovich, & Reed, 1992).

Peers provide a unique and valuable perspective from which to evaluate performance (Greguras et al., 2001; Maurer, Raju, & Collins, 1998; Murphy & Cleveland, 1995; Reilly & McGourty, 1998; Toegel & Conger, 2003). Their close working relationships provide opportunities to observe a wide range of behaviors, including actions employees may try to conceal from their supervisor (Lawler, 1967; Klimoski & London, 1974). Peer ratings are moderately correlated with individual and organizational outcomes, including workload sharing, voice, cooperation, performance, satisfaction (Erez, Lepine, & Elms, 2002), productivity, salary change, and promotions (Kane & Lawler, 1978; Norton, 1992; Shore, Shore, & Thornton, 1992). Information from peer ratings also influences supervisors' performance judgments (Makiney & Levy, 2002).

The most appropriate rating source depends on the type of performance behavior (Kane & Lawler, 1979) and the position of the rater in the organizational hierarchy. A rater's position determines the opportunity to observe and the appropriateness of the perspective (Barrett, 1966). Most studies of contextual performance have relied on ratings of single supervisors (cf., Organ, 1988; 1990). However, within an organization there are likely to be multiple recipients of contextual performance (McNeely & Meglino, 1994; Organ & Ryan, 1995). Supervisors are likely most concerned with task performance and may not be in the best position to evaluate behaviors outside of the work task (e.g., helping and training coworkers, being good team

players; Fedor, Bettenhausen, & Davis, 1999). Many extra-role behaviors may escape the attention of, or are not performed in the presence of the supervisor (Organ & Konovsky, 1989). These behaviors may also be directed at individuals besides the supervisor (Allen, Barnard, Rush, & Russell, 2000). Fletcher (2001) identified peers as a potentially optimal source for assessments of contextual performance because peers pay more attention to extra role behaviors when making overall performance ratings than supervisors who focus on task performance (Conway, 1999). Further, peers may be an especially appropriate rating source in SMTs due to the lack of external supervision.

Although the use of alternative rating sources has been largely embraced by organizations, when resistance exists it is often due to a lack of trust in peer raters (Brutus & Derayeh, 2002; Fedor et al., 1999). Employees have expressed concerns that peer ratings may be biased by friendship and popularity (Cederblom & Lounsbury, 1980; Doll & Longo, 1962; Landy & Farr, 1983; Love, 1981). For example, Bettenhausen and Fedor (1997) found that only those employees with good coworker relationships supported the use of peer ratings for both developmental and administrative purposes. Somewhat surprisingly, these same concerns were not expressed regarding the use of supervisor ratings. The personal relationships that peers develop may lead to more subjectivity in ratings, especially when used for administrative purposes (Van Hooft, Van der Flier, & Minne, 2006). Peers may also be less invested than supervisors in providing performance evaluations because rating others' performance is not a typical job requirement (Conway & Huffcutt, 1997). Further, peers maybe reluctant to differentiate among their coworkers (Murphy & Cleveland, 1995).

In sum, peer ratings offer a different, useful perspective and are widely used in performance evaluations. Peer ratings may be most useful in team context where intensive

interactions occur between team members and external supervision is limited. Peers are likely appropriate rater's of their teammates' contributions to the team's project and contextual performance. However, relatively little is known about which factors contribute to peers' ratings, especially within work teams. Team members are by definition interdependent (Van der Vegt & Van de Vliert, 2002; Mathieu et al., 2008), thus, it is likely that the relationships between team members will influence performance ratings. Researchers have proposed that a more thorough understanding of peer appraisals is essential and investigations into peer ratings should include the influence of peer relationships (Pulakos et al., 1996). Bretz et al. (1992) also called for future research examining the cognitive processes underlying peer performance ratings. They suggested that peer performance appraisals might differ from traditional supervisor ratings due to differences in power dynamics and social interactions. The extent to which employees are able to observe each other's behaviors directly may affect the factors that contribute to peers' ratings. This study answers these calls for more research by examining peer performance ratings in co-located and virtual teams.

Co-located and virtual teams: Implications of observing team members. Advances in technology allow teams to function not only through face-to-face interactions, but also at a distance. A team's architecture (i.e., structure) is determined by team member proximity, communication modality, and the allocation of functions to team members (Urban, Bowers, Cannon-Bowers, & Salas, 1995). Distributed teams differ from co-located teams on two architectural determinates (i.e., member proximity and communication modality). Teams operating from distributed geographical locations must communicate through some form of technology (e.g., mobile phones, text messaging, two-way pagers, e-mail). For this reason,

distributed teams are often called virtual teams (VTs)[3]. Communication via technology is essential for VTs because the geographic distribution of team members prevents face-to-face interactions (Townsend, DeMarie, & Hendrickson, 1998). Co-located teams may also communicate through technological mediums; however, these electronic means of communication are supplemental and often not necessary because their shared geographic location allows co-located team members to have daily face-to-face interactions (Bell & Kozlowski, 2002).

Approximately 60% of professional employees are members of VTs (Jones, 2004; Kanawattanachai & Yoo, 2002; Yoo, Kanawattanachai, & Citurs, 2002). By using VTs, companies are able to assemble team members regardless of their geographic location. This strategy allows organizations to capitalize on expertise distributed throughout their organization and be responsive, adaptive, and flexible (Bell & Kozlowski, 2002). The task, goals, and mission of virtual and face-to-face teams do not differ; however, the processes they use to achieve their goals may differ because of constraints faced by VTs (Bell & Kozlowski, 2002; Salas et al., 2006). Many studies make comparisons within a sample of virtual or co-located teams, but do not make compare these different types of teams (MacDuffie, 2008). MacDuffie suggested that researchers use theoretical frameworks developed for co-located teams to understand VTs and that it is imperative for studies to compare co-located and virtual teams that share the same context. Given the prevalence and apparent advantages to organizations of utilizing VTs, it is important for researchers to understand the processes that occur in VTs and compare these processes to traditional co-located teams.

[3] The term virtual team has also been used to refer teams that include teammates that are synthetic or machines (Salas et al., 2006).

The responsibility for performance assessment in SMTs, shifted from leaders to team members because team members have a better opportunity to observe each other's performance (Manz & Sims, 1987; Tippins & Coverdale, 2009). VTs often operate as SMTs and rely on shared leadership (Balthazard, Waldman, & Warren, 2009, Conger & Pearce, 2003; Pearce, Yoo, & Alavi, 2004). The regulation of performance is essential for the successful performance of SMTs (Bell & Kozlowski, 2002; Kozlowski, Gully, McHugh, Salas, & Cannon-Bowers, 1996; Weisband & Atwater, 1999). High performing VTs teams displayed more leadership behaviors directed at keeping track of the group's work during a semester-long project than low performing VTs (Carte, Chidambaram, & Becker, 2006). As in other SMTs, virtual team members have intensive interactions with peers, thus, peer ratings may be an appropriate mechanism for performance measurement. However, working in virtual teams provides fewer opportunities to observe performance directly (Paris, Salas, & Cannon-Bowers, 2000). A better understanding of the appraisal process underlying alternative performance appraisal sources (i.e., self- and peer-ratings) in the light of contextual constraints (e.g., co-located and virtual teams) is needed (Bretz et al., 1992). The factors contributing to the effectiveness of face-to-face and virtual teams are not necessarily the same (Potter & Balthazard, 2002) and researchers have questioned whether technology will facilitate or inhibit performance management (e.g., Latham & Mann, 2006). Thus, the current study will examine whether communication via technology and over distance fundamentally alters self-ratings and peer-ratings of performance.

An Interpersonal Perception Perspective on Performance Appraisal

Interpersonal perception occurs when one person (i.e., the perceiver) makes judgments about or forms impressions of another person (i.e., the target). Interpersonal perception is a complex process because perceptions are often of other people, thus, they are two-sided (i.e.,

individuals act as both perceivers and targets; Back & Kenny, 2010). Back and Kenny explain that individuals are aware of the two-sided nature of interpersonal perception. They know that others form perceptions of them and they, in turn, form their impressions regarding what they think others think about them (i.e., meta-perceptions). Their perceptions of others are also influenced by what they think about themselves (i.e., self-perceptions). In sum, an individual's perceptions differ depending on whom the person is rating and people's perceptions of the same target will also differ. These complexities highlight an important tenant of interpersonal perception: The formation of a perception involves an interaction between the perceiver and target. In other words, the perceiver and the target are co-producers of the perception (Klimoski & Donahue, 2001).

The complexities of interpersonal perception have important implications for the performance appraisal process—both raters (i.e., perceivers) and ratees (i.e., targets) contribute to performance ratings. In traditional performance appraisal systems, the supervisor has acted as the perceiver and the subordinate of the supervisor has acted as the target. The transition to team-based work and peer performance appraisals has shifted the role of the perceiver to peers (i.e., team members) and the target (i.e., person whose performance is being evaluated) is a coworker or team member of the perceiver (Barnes-Farrell, 2001). Team performance management systems may require each team member to rate the performance of every other member of their team. This type of round-robin peer rating system creates a reciprocal situation in which each team member serves as both the perceiver and the target (Kenny, Kashy, & Cook, 2006; Kenny & Albright, 1987). The dual role served by each employee makes peer-ratings unique from ratings provided by other sources (e.g., supervisors, subordinates; Fedor et al., 1999). When individuals engage in dyadic interactions, such as rating each other's performance, the ratings

they provide are likely to be related (Kenny & Judd, 1986). For example, if Alice and Bill are team members, Alice's rating of Bill's performance is likely to be more similar to Bill's rating of Alice's performance than to ratings of performance provided by other individuals. The independence assumption presumes that data provided by individuals will be unrelated except for variance due to the independent variable of interest. The dependence in Alice and Bill's ratings is likely created by their membership on the same team and also their co-production of each rating (i.e., the rating Alice provides of Bill and the rating Bill provides of Alice). Thus, the dependence in these performance ratings likely violates the independence assumption of many traditional statistical techniques (Kenny et al., 2006). To make valid conclusions the analyses used by researchers must account for the interdependence within peer ratings.[4]

Interdependence is a defining characteristic of work teams (Salas et al., 1992; Van der Vegt & Van de Vliert, 2002; Mathieu et al., 2008). Thus, interdependence should not be treated as a "statistical nuisance that should be controlled, but rather as an important social psychological phenomenon that should be studied" (Kashy & Kenny, 2000, p. 452; Kenny & Judd, 1986). Variance-partitioning approaches have been used to evaluate performance appraisal systems; however, these traditional approaches are limited in the information they can provide regarding the sources that contribute to variance in performance ratings. Further, these traditional

[4] When individuals are mistakenly treated as the unit of analysis and data are not independent the bias can be manifested as increased Type I or Type II error. The type of error depends on the direction of correlation between group members' scores on the outcome variable and the type of independent variable (i.e., between or within). In the present study, I have a between independent variable (i.e., whether teams interact face-to-face or virtually) and the direction of nonindependence is likely positive; thus failure to account for dependence in the ratings would result in inferential statistics that are too large, *p* values that are too small, and increased probability of making a Type I error (Kashy & Kenny, 2000; Kenny et al., 2006).

approaches do not correspond to reciprocal designs in which individuals act as both perceivers and targets. The following sections describe: (1) the application of traditional variance partitioning approaches to performance ratings, (2) an extension of these approaches that is applicable to peer performance ratings based on a round-robin rating design, and (3) a number of fundamental interpersonal perception questions that will provide insight into the appropriateness of peer performance ratings.

Traditional variance partitioning methods for evaluating performance ratings. Classical test theory (CTT) attributes variance in ratings to "true score" variance and unsystematic measurement error. In accordance with CTT, Schmidt and Hunter (1996) defined true score variance as the "worker's average evaluation across a population of raters" (p. 210). However, Schmidt and Hunter's definition does not recognize that other shared factors among perceivers influence their level of agreement regarding a target's performance (Murphy & DeShon, 2000). Examinations of performance ratings have led researchers to reject the CTT approach and acknowledge that systematic variance in performance ratings remains after controlling for true score variance (i.e., variance due to the target of the performance appraisal).

Generalizability theory (GT; Cronbach, Gleser, Nada, & Rajaratnam, 1972) offers an alternative approach to variance partitioning, such that variance in ratings is separated into components beyond true score variance and random error. Wherry's theory of rating (Wherry & Bartlett, 1982) and Scullen et al.'s (2000) later work identified three factors that contribute to variance in performance ratings: (1) actual job performance (demonstrated by the target), (2) perceptual and recall bias that is connected to the perceiver, and (3) measurement error.

A considerable proportion of the variance in GT studies has been attributed to perceiver effects (e.g., Greguras, Robie, Schleicher, & Goff, 2003). Evidence from multi-source rating

systems demonstrates that perceivers in different roles appear to evaluate the same target differently (Woehr, Sheehan, & Bennett, 2005). This perceiver variance is not unsystematic error, but instead represents agreement or disagreement in ratings due to shared characteristics that are unrelated to the target's performance (i.e., the perceiver main effect). Perceivers may agree because they share common goals and biases, have similar frames of reference (e.g., perceptions of the organization and the appraisal system), or share similar relationships with the target (Murphy & DeShon, 2000). Perceivers may also disagree in systematic ways because of differences in the behaviors they observe, their access to information unrelated to performance, their expertise, and their evaluative tendencies (Murphy & DeShon, 2000). Some researchers (Mount et al., 1998; Scullen et al., 2000) have suggested that perceiver variance is generally due to the tendencies of individual raters as opposed to their level with the organizational hierarchy (e.g., some raters are more lenient than other raters). Thus, it is likely that a large portion of the variance in performance ratings will be accounted for by perceiver effects even when raters are at the same organizational level (e.g., all peers). Rater disagreement does not necessarily invalidate the use of peer ratings. For example, Kasten and Nevo (2008) found that although peer ratings show relatively low levels of agreement, the criterion-related validity of peer ratings was only low when the interrateragreement was less than or equal to .4. Based on this criterion-related validity information, peer ratings appear to be a viable source for performance appraisal information. However, the large proportion of the variance attributed to raters does indicate a need to better understand the factors that contribute to peers' performance ratings.

The substantial proportion of perceiver variance found in GT studies may be due to (1) different tendencies among raters—*perceiver main effects* or (2) unique ratings produced by rater-ratee combinations—*perceiver-by-target interaction* (Murphy & DeShon, 2000). The

traditional design of GT studies involves multiple raters rating a single target (i.e., the ratee; Cronbach et al. 1972). This one-sided design is also typical of many traditional job performance evaluations that were carried out by supervisors (Viswesvaran, Ones, & Schmidt, 1996). However, the nested nature of these studies (i.e., targets-within-perceivers) prevents the separation variance due to the perceiver main effect from variance due to the perceiver-by-target interaction (e.g., Conway, 1998; Greguras et al., 2003; Greguras & Robie, 1998; Scullen et al., 2000). A major limitation of the top-down design of traditional performance appraisals and the use of GT as a variance partitioning approach is the inability to separate these two effects, which makes the results of these studies difficult to interpret (Murphy & DeShon, 2000; Woehr et al., 2005). Further, failure to examine both perspectives and the resulting dependence in reciprocal designs results in a loss of information and may even lead to biased statistical tests (Kenny et al., 2006; Kenny & Judd, 1986). These limitations are overcome by using the *social relations model* (SRM; Kenny, 1994; Kenny & La Voie, 1984; Warner, Kenny, Stoto, 1979; Kenny et al., 2006) to examine round-robin ratings of performance (Greguras, Robie, Born, & Koenigs, 2007).

Social relations model. The SRM is a conceptual framework that may be used to capture the complexities of dyadic phenomena, including interpersonal perception (Back & Kenny, 2010).[5] The SRM is also a statistical method designed to handle dependence within round-robin ratings (Kenny et al., 2006). The application of variance-partitioning techniques to performance ratings is based on the assumption that interpersonal perceptions (or any dyadic process) may be divided into components (Kenny et al., 2006; Kenny, West, Malloy, & Albright, 2006). The SRM has been described as a special case of GT (Greguras et al., 2001) that is applied to data

[5] The SRM may also be used to estimate variance components in actual observed behavior or to make comparisons between behavior and perceptions of behavior (Kashy & Kenny, 2000; Kenny et al., 2006).

gathered from reciprocal designs (i.e., designs in which participants are both the perceiver and the target; Kenny, 1994; Malloy & Kenny, 1986). As in other applications of GT, the SRM is used to estimate sources (i.e., components) of variance in the performance ratings. However, as previously stated, traditional GT approaches are only able to isolate variance attributable to the perceiver or the target. These factors only partially accounted for the variance in performance ratings, thus Vance, Winne, and Wright (1983) suggested that researchers should use a dyad-based analysis that includes the assessment of the perceiver–target interaction to explain the process of performance appraisal fully.

Description of the social relations model effects and variances. SRM analyses use a 2-factor, random effects, analysis of variance (ANOVA; Kenny, 1994), perceiver is one factor, target is the other factor, and a perceiver-by-target interaction represents the relationship. Using the SRM analysis, each performance rating is partitioned into the *perceiver effect* (i.e., the component of the rating due to the general tendency of the person providing the rating), the *target effect* (i.e., the component of the rating due to the general tendency of a person to be perceived in a certain way), and the *relationship effect* (i.e., a unique component of the rating due to the perceiver-target pair).

For the purpose of demonstrating the meaning of each of the SRM components, assume that X_{ij} is Alice's ratings of Bill's contextual performance. Several factors may contribute to Alice's rating. The first component is the *mean level* of contextual performance behavior perceived in the group (m). In other words, the average level at which all perceivers view all targets on the performance construct (Kenny, 1994). In some groups, perceivers (as a whole) may have a tendency to see targets as displaying more or fewer contextual performance behaviors. The second component is the *perceiver effect* (a), which represents Alice's tendency to

see others as displaying contextual performance behaviors. For example, Alice may see all people as going out of their way to help her. The third component is the *target effect* (b), which is Bill's tendency to be seen as displaying contextual performance behaviors. All of Bill's teammates may perceive that he goes out of his way to help them. The perceiver and target effects are individual level effects. The fourth component is the *relationship effect* (g), which is Alice's unique perception of Bill's contextual performance behaviors. Alice might think Bill is especially helpful (i.e., more helpful than Bill is viewed by their other teammates). The relationship effect is at the dyad level and is directional (Kenny et al., 2006). In other words, the component of the performance score due to Alice's unique rating of Bill's contextual performance behaviors (g) does not have to match the component of the performance score due to Bill's unique rating of Alice's contextual performance behaviors (g'). Relationship effects are calculated after controlling for target and perceiver effects and thus, represent systematic variance that cannot be explained by either of the perceiver or the target main effects (Kenny, Mohr, & Levesque, 2001). Finally, Alice's rating of Bill's contextual performance may also be influenced by random measurement error. Multiple indicators or multiple measurements over time are required to separate the relationship component of an individual score from random measurement error (Kenny et al., 2006). Table 1 provides a description of each of the SRM components.

The SRM links statistical tests related to variance of perceiver, target, and relationship components and correlations between these different components and between these components and self-ratings to theoretical concepts. Researchers use the SRM to examine the extent to which characteristics of the perceiver, target, and perceiver-target relationship contribute to peers' performance ratings by determining the amount of variance in the perceiver effect (s_a^2), target

effect (s_b^2), relationship effect (s_g^2), respectively. A statistically significant amount of variance in the perceiver, target, and relationship component within a team allows researchers to infer that characteristics of the perceiver, target, or the perceiver-target relationship contribute to performance ratings. The variance associated with each of these components can also be divided by the overall variance to provide an indication of their relative importance. The correlation between the SRM relationship components provides evidence regarding the extent of reciprocity in peer performance ratings (i.e., the extent to which individuals' perceptions mirror each other). The correlations between the SRM target component of multiple variables provides an indication of the extent to which targets perform similarly across performance dimensions and the correlations between the SRM perceiver component of multiple variables provides an indication of the extent to which perceivers rate targets similarly across performance dimensions. Finally, the correlations between the SRM components and self-ratings provide insight into the extent to which peer-ratings are related to self-ratings (i.e., an individual's self-concept; Kenny et al., 2006; Back & Kenny, 2010). The statistical details of the SRM are provided in Chapter 3. The meaning and relevance of these variances and correlations will be discussed in more detail in the following sections.[6]

Application of the social relations model to peer performance ratings. The SRM has been used to study many topics including the interpersonal perception of personality traits (e.g., Kenny, 1994; Markey & Wells, 2002), status within a group (e.g., Anderson, Srivastava, Beer, Spataro, & Chatman, 2006), affect (Park & Flink, 1989; Kenny, 1994), behaviors (e.g., Robins,

[6] Kenny et al. (2006) warn that a common mistake in the presentation of SRM analyses is to confuse effect estimates and variances. The term *effect* should be used to refer to a particular component of a person's score. For example, the perceiver, target, and relationship effects (described above) refer particular components of a person's score. When referring across participants the term *variance* should be used.

Mendelsohn, Connell, & Kwan, 2004; Kenny et al., 2001), attributions (Robins et al., 2004), leadership (Albright & Forziati, 1995; Campbell, Simpson, Stewart, & Manning, 2003; Livi, Kenny, Albright, & Pierro, 2008; Kenny & Zaccaro, 1983; Malloy & Janowski, 1992; Peters, Kinsey, & Malloy, 2004), and participation in groups (Bonito, 2003).

The social relations model is an appropriate method for examining peer performance ratings because it accounts for the dependence created by (1) the reciprocal nature of the round-robin design and (2) the nested nature of individuals within work teams. The SRM also allows researchers to determine if characteristics of the perceiver, target, or relationship are contributing to variability in peer performance ratings and answer questions that will illuminate the interpersonal perception processes involved in performance appraisals. However, few studies have applied the SRM to performance ratings (Sullivan & Reno, 1999; Boldry & Kashy, 1999; Greguras, et al., 2001; 2007) and all of these studies have been conducted with co-located teams. These studies examined a variety of performance dimensions that vary in terms of how prototypical the measures are of task or teamwork performance. Sullivan and Reno (1999) asked students to predict their group members' scores on individual quizzes, rate the amount of credit attributed to each team member for the score achieved on a group quiz, and rank each individual's responsibility. Boldry and Kashy (1999) had students in the Corps of Cadets assess performance criteria, including motivation, leadership, and character. Greguras et al. (2001) had participants rate their teammate's cooperation, ideas, effort, and reliability. Finally, Greguras et al. (2007) had employees from a variety of organizations and occupations complete ratings for a six-dimension group process measure (i.e., teamwork orientation, individualism, rule compliance, dedication, affiliation, and motivation). As previously stated, I chose to use measures of taskwork (i.e., contribution to the team project) and teamwork (i.e., contextual

performance) because researchers have recommended including both competencies in team performance measurement systems (e.g., Murphy & Cleveland, 1995; Porter et al., 2003; Salas et al., 2009; Salas, Sims, & Burke, 2005). To develop my hypothesis, I draw on the small number of studies that have examined performance ratings using the SRM in co-located teams and other SRM studies of traits, affect, and behavior conducted in face-to-face and virtual settings.

Consensus, assimilation, and uniqueness in co-located and virtual teams. In this section, I address three questions that are fundamental to understanding of peer performance appraisal as a form of interpersonal perception (Kenny, 1994). First, do team members agree in their ratings of a teammate? Second, do individual team members tend to rate the performance of all of their teammates in the same way? Third, do individuals form unique impression of their teammates' performance? The answers to these questions are determined by examining the amount of variance in peer performance ratings due to the target, perceiver, and relationship components. As previously stated, the statistical significance of the target variance (s_b^2), perceiver variance (s_a^2), or relationship variance (s_g^2) denotes whether characteristics of the person providing the rating, the person being rated, or the idiosyncratic relationship between the rater and the ratee influences performance ratings. Further, the proportion of total variance due to the target, perceiver, and relationship may be used to determine which factors have the strongest influence on performance ratings. These theoretical questions are summarized in Table 2.

Consensus: How much do targets differ in how their performance is generally seen by others? A high level of agreement (i.e., consensus) in perceivers' ratings is an indication that the variance in the ratings is largely determined by characteristics of the target. The extent of consensus is evaluated by examining the magnitude of the variance in the target effect (Kenny, 1994). Agreement among raters is important because consensus is a necessary, but not sufficient

condition for accuracy (i.e., similar to the relationship between reliability and validity). Consensus generally implies accuracy; however, raters can agree in their ratings of a target and not be accurate (Kenny, 1994). Further, ideally, characteristics of the targets (e.g., their relevant workplace behaviors) will drive ratings in a performance appraisal system. When perceivers agree on who is a high, medium, or low performer, there is a high level of consensus in peer performance ratings. For example, if a team consists of four individuals (Alice, Bill, Charlie, and Devon), and Alice, Bill, and Charlie agree that Devon contributes her fair share to the group project, if all others agree that Alice is a slacker, Bill goes above and beyond, and Charlie does the bare minimum, then there is a high degree of consensus in peer ratings of performance. If raters differ in their assessment of their teammates' performance then there is a low level of consensus.

The SRM was originally used to examine consensus in raters' perceptions of others' traits (e.g., Big Five personality characteristics). Trait perception studies generally found low levels of consensus. Kenny (1994) concluded that 15% of the variance in trait ratings is attributable to characteristics of the target. Even among highly acquainted individuals, target variance accounts for less than one-third of the total variance in ratings. Further, Kenny found that only 10% of the variance in ratings of affect (i.e., liking) was attributed to the target. However, there is reason to expect a higher level of consensus for perceptions of behaviors, including contribution to work and contextual performance. Kenny reported that consensus was highest for perceptions of extroversion (i.e., .27-.32). This finding is likely due to the observable nature of acts related to this trait. Further, Marcus (1998) reported that considerable consensus has been found in perceptions of leadership (e.g., 28-44% in Campbell et al., 2003; 49% in Kenny & Zaccaro, 1983; 40% in Livi et al., 2008; 60% in Malloy & Janowski, 1992; 37% in Vargas, 1996). The

previous four SRM studies of performance ratings found significant target variance (Boldry & Kashy, 1999; Greguras et al., 2001; 2007; Sullivan & Reno, 1999). On average target effects accounted for 24.5% of the variance in peer ratings (Table 3). The significant target variance indicates that characteristics of the target drive performance ratings. I expect to replicate these findings for contribution to work and contextual performance in co-located teams.

Hypothesis 1a: Perceivers will demonstrate consensus in their ratings of peer performance in co-located teams (i.e., the average amount of target variance ($s_b^2{}_{\text{co-located}}$) will be statistically significant).

Virtual interactions remove many physical cues (e.g., attractiveness) that may detract from consensus because raters may be influenced by these cues differently. For example, Weisband and Atwater (1999) found that in electronic groups, the most salient factor in determining performance ratings was a ratee's actual input as opposed to how much the rater liked the ratee. Further, Markey and Wells (2002) found that participants were able to reach consensus for traits of extraversion, agreeableness, and openness when engaging in virtual chat room conversations. Thus, I expect perceivers in VTs to reach a significant level of consensus.

Hypothesis 1b: Perceivers will demonstrate consensus in their ratings of peer performance in VTs (i.e., the average amount of target variance ($s_b^2{}_{\text{virtual}}$) will be statistically significant).

In addition to removing distracting physical appearance cues, virtual communication reduces the general number of shared stimuli. Previous research has found stronger consensus for trait perceptions when individuals interacted as a group as opposed to one-on-one (Kenny, 1994). This finding is likely due to the larger amount of similar stimulus information observed by all group members during the group interactions (Malloy, Agatstein, Yarlas, & Albright, 1997).

Face-to-face interactions occur between team members who can see and hear one another, receive and send messages in real time, and communicate information both simultaneously and in sequence (Driskell, Radtke, & Salas, 2003). In VTs, there are fewer visual and contextual communication cues (MacDuffie, 2008). Participants also reported fewer performance monitoring behaviors when collaborating using audio means (Fletcher & Major, 2006). The loss of these cues may lead to more varied interpretations of behaviors and reduced consensus. Co-located teammates also made more situational attributions for behaviors because of their shared knowledge of the constraints influencing their teammates; however, virtual teammates made more dispositional attributions for their teammates behaviors (Cramton, Orvis, & Wilson, 2007). Robins et al. (2004) compared ratings of the extent to which teammates were viewed as talkative, warm, and nervous while engaging in a separate get-to-know you exercises with each of their teammates. Robins et al. found greater consensus in the face-to-face condition for all behaviors than in the phone or computer mediated conditions. Based on the existing research, I expect to find greater consensus for peer ratings in co-located SMTs than in VTs.

Hypothesis 1c: Consensus will be higher in co-located than in VTs (i.e., the average amount of target variance ($s_b^2{}_{\text{co-located}}$) in co-located teams will be greater than the average amount of target variance ($s_b^2{}_{\text{virtual}}$) in virtual teams).

Assimilation: How much do perceivers differ in how they generally see their teammates' performance? Researchers have extensively documented that a considerable proportion of the variance in performance ratings is attributable to either perceiver effects or the perceiver-by-target interaction (e.g., Greguras et al., 2003; Woehr et al., 2005; Murphy & DeShon, 2000). A portion of this variance is likely due to the tendency of individual perceivers to see all targets in the same way. Kenny (1994) called this effect *assimilation*. The amount of

assimilation in peer performance ratings is assessed by examining the magnitude of the variance related to perceiver effects. A large amount of perceiver variance would indicate that perceivers use different general standards when rating their teammates' performance (Back & Kenny, 2010). In other words, characteristics of the perceiver (irrespective of the target) are contributing to the ratings. Despite teammates sharing the same organizational role (or rater perspective), it is reasonable to expect substantial perceiver variance due to the tendencies of individual raters (Mount et al., 1998; Scullen et al., 2000). For example, this variability may be due to differences in leniency or other rater tendencies. If a team consists of four individuals (Alice, Bill, Charlie, and Devon), and Alice rates Bill's, Charlie's, and Devon's contextual performance high, but Bill rates Alice's, Charlie's and Devon's contextual performance low, and Charlie rates everyone else's performance kind of high, and Devon rates everyone else's performance kind of low then there will be a large amount of perceiver variance (i.e., a high degree of assimilation).

Previous SRM studies have found that perceiver variance accounts for approximately 20% of variance in trait perceptions and liking (Kenny, 1994). In general, all four of the previous studies on perceptions of performance ratings found smaller, but significant levels of perceiver variance (Boldry & Kashy, 1999; Greguras et al., 2001; 2007; Sullivan & Reno, 1999). On average, the perceiver variance accounted for 9.8% of the total variance in peer ratings (Table 3). The significant perceiver variance indicates rater characteristics play a role in performance ratings and may provide evidence of potential bias. Variance attributable to perceiver bias is generally referred to as nonrandom errors (e.g., a rater may be lenient or harsh for all ratees in comparison to other raters), but may also reflect differences in perspectives or opportunities to observe behavior (Woehr et al., 2005). I expect to replicate these findings for contribution to work and contextual performance in co-located teams.

Hypothesis 2a: Perceivers will demonstrate assimilation in their ratings of peer performance in co-located teams (i.e., the average amount of perceiver variance ($s_a^2_{co\text{-}located}$) will be statistically significant).

Kenny (1994) explained that perceiver variance represents a group stereotype, the perceiver's ignorance, or the perceiver's best guess of what a typical person is like. There is strong evidence to support this position as perceiver variance declines when participants know each other for a longer period of time (Kenny, 1994; Kenny, Horner, Kashy, & Chu, 1992; Park & Judd, 1989). Thus, as more information about the other person becomes available the perceiver variance component should decrease. Communication via technology often filters out visual and contextual cues and thus, decreases the information provided by a message between teammates (Driskell et al., 2003). Robins et al. (2004) found relatively similar amounts of perceiver variance in their face-to-face, phone, and computer-mediated conditions (on average 22-26%). However, their study examined an interaction of extremely short duration on a task specifically designed for teammates to get to know each other. Thus, I expect to see a more drastic difference between the magnitude of perceiver variance from members in co-located and virtual teams when they work together for a longer time and are focused on completing a task as opposed to getting to know each other. Markey and Wells (2002) found significant perceiver variance in ratings of their chat room conversation partners' openness, conscientiousness, and agreeableness. At least for trait perceptions, in a virtual chat room environment, perceivers tend to see their targets in a similar way. Thus, I expect to find that perceivers tend to see their virtual teammates in the same way (i.e., provide similar ratings of performance for all team members). Further, assimilation will likely be greater in virtual than in co-located teams because of the reduction in shared information in VTs.

Hypothesis 2b: Perceivers will demonstrate assimilation in their ratings of peer performance in VTs (i.e., the average amount of perceiver variance ($s_a^2{}_{virtual}$) in VTs will be statistically significant).

Hypothesis 2c: Assimilation will be higher in virtual than in co-located teams (i.e., the average amount of perceiver variance ($s_a^2{}_{virtual}$) in VTs will be greater than the average amount of perceiver variance ($s_a^2{}_{co\text{-}located}$) in co-located teams).

Uniqueness: How much do individual raters differ in their unique perceptions of their specific teammates' performance? The idiosyncratic tendencies of specific raters (i.e., the perceiver-by-target interaction) may also contribute to the substantial rater effect found in previous studies on traditional performance ratings (e.g., Greguras et al., 2003; Woehr et al., 2005; Murphy & DeShon, 2000); however, the nested-nature of targets (employees) within perceivers (supervisors) has prevented researchers from examining the variance attributable to the perceiver-target relationship. Perceivers are likely to provide idiosyncratic ratings when they use information that is not available to all raters, attach different meanings to the same behavior, or become influenced by nonverbal information (e.g., liking of the ratee; Kenny, 1994). Statistically, the relationship effect represents the extent to which a perceiver's view of the target cannot be explained by the perceiver or target main effects. Thus, there will be a large amount of relationship variance when ratings provided by perceivers are consistently not explained by the general tendency of perceivers to rate targets in a certain way or the general tendency of targets to be rated in a certain way (Back & Kenny, 2010). For example, Alice might rate Bill high on performance, even though she typically rates everyone else low and all of their other teammates typically rate Bill as low on performance. If other perceiver-target pairs also demonstrate this type of idiosyncratic rating then there would be a large amount of relationship variance. This

variance in peer ratings has been referred to by the term *uniqueness* (Kenny, 1994) and the extent of uniqueness in peer performance ratings may be assessed by examining the magnitude of the variance in the relationship effect.

As previously stated, a major limitation of the traditional application of GT approaches to supervisor performance ratings of subordinates was the inability to separate the variance due to the perceiver effect from the variance due to the perceiver-target interaction (i.e., the relationship effect). In other words, the inability of this approach to separate evidence supporting assimilation in peer ratings from evidence of uniqueness in peer ratings. Kenny (1994) concluded that uniqueness (i.e., relationship variance) will be presented in trait ratings at a "respectable level" and documented that relationship variance generally accounts for 20% of the total variance in trait perception and 40% of the total variance in ratings of affect. On average, the relationship effect accounted for 22.7% of the total variance in past studies of peer ratings (Table 3). The significant relationship effect indicates that perceivers view targets in unique ways. It is clear that interpersonal relationships are an important factor that accounts for the variance of many behaviors (Marcus, 1998; Miller & Kenny, 1986), including performance (Nathan, Mohrman, & Milliman, 1991). I expect to replicate these findings for contextual performance in co-located teams. Contribution to work was only measured with one item, thus, the relationship effect cannot be separated from measurement error (Kenny, 1994; Kenny et al., 2006). It would be inappropriate to interpret the magnitude of the combined variance component as an unknown and potentially large portion of the variance may be due to random measurement error (Ingraham & Wright, 1986).

Hypothesis 3a: In co-located teams, perceivers will see targets in idiosyncratic ways (i.e., demonstrate uniqueness; the average amount of relationship variance in co-located teams ($s_g^2{}_{\text{co-located}}$) will be statistically significant).

Unfortunately, the two previous studies using the SRM model in a virtual environment (i.e., Markey & Wells, 2002; Robins et al., 2004) were unable to separate relationship from error variance. Relationship characteristics are likely to have a stronger influence on performance ratings in co-located teams because team members have had more opportunity to interact and develop close relationships. As previously mentioned, liking a team member impacted evaluations of member contributions in face-to-face groups, but not in electronic groups (Weisband & Atwater, 1999). Further, the opportunity exists for face-to-face groups to develop stronger relational ties as evidence by decreased cohesion in VTs (Straus, 1997; Warkentin, Sayeed, & Hightower, 1997). Driskell et al. (2003) also proposed that reduced sharing of personal information and the weakening of social cues in VTs might lead to decreased intimacy and weaken emotional bonds (Straus, 1997; Weisband & Atwater, 1999). However, members of VTs may also develop unique relationships or have access to unique information that will result in systematic relationship variance. Given the differences in the social interactions in co-located and VTs, I expect perceivers to demonstrate uniqueness in VTs, but the amount of uniqueness will be larger in co-located than in virtual teams.

Hypothesis 3b: In virtual teams, perceivers will see targets in idiosyncratic ways (i.e., demonstrate uniqueness; the average amount of relationship variance in VTs ($s_g^2{}_{\text{virtual}}$) will be statistically significant).

Hypothesis 3c: The amount of uniqueness will be larger in co-located teams than in VTs (i.e., the average amount relationship variance ($s_g^2{}_{\text{co-located}}$) will be greater in co-located teams than in virtual teams ($s_g^2{}_{\text{virtual}}$).

What is the relative contribution of consensus, assimilation, and uniqueness to peer ratings? Previous research using the SRM to analyze peer performance ratings (Sullivan & Reno, 1999; Boldry & Kenny, 1999; Greguras et al., 2007) has found that consensus (target variance) and uniqueness (relationship variance) were larger than assimilation (perceiver variance; Table 3). Greguras et al. (2007) found uniqueness (relationship variance) in performance ratings was particularly strong (i.e., relationship variance was 2.13 times the perceiver variance and 1.31 times the target variance). These findings are in contrast to the results of Sullivan and Reno (1999), which found that target variance, was 1.32 times larger than relationship variance. In general, I expect to find consensus (target variance) and uniqueness (relationship variance) are similar in magnitude and much larger than assimilation (perceiver variance) in co-located teams.

Hypothesis 4a: In co-located teams, raters will demonstrate similar amounts of consensus and uniqueness and smaller amounts of assimilation (i.e., the average amount of target variance ($s_b^2{}_{\text{co-located}}$) ≈ the average amount of relationship variance ($s_g^2{}_{\text{co-located}}$) > the average amount of perceiver variance ($s_a^2{}_{\text{co-located}}$)).

There is likely to be a different pattern of relative variance magnitude in ratings provide by peers in VTs. In VTs, I would expect uniqueness (relationship variance) to be relatively small and assimilation (perceiver variance) to be relatively large. Previous research has demonstrated that peers in VTs are able to identify their coworkers' contributions (e.g., Weisband & Atwater, 1999). Thus, I would expect the consensus (target variance) to be relatively large in VTs.

Hypothesis 4b: In VTs, raters will demonstrate similar amounts of consensus and assimilation and smaller amount of uniqueness (i.e., the average amount of target variance ($s_b^2{}_{virtual}$) ≈ the average amount of perceiver variance ($s_a^2{}_{virtual}$) > the average amount of relationship variance ($s_g^2{}_{virtual}$)).

Reciprocity, bias, and self-other agreement in co-located and virtual teams. The results from the SRM variance-partitioning analysis provide an initial understanding of peer performance ratings. A deeper understanding of peer ratings of performance may be reached by examining (1) the extent of influence peers have on each other ratings (i.e., whether reciprocity exists in peer ratings), (2) the extent to which peer raters demonstrate halo bias (i.e., have a tendency to rate individuals in the same way across constructs), (3) the extent to which targets demonstrate true halo (i.e., have a tendency to perform at similar levels across performance constructs), and (4) the extent to which individuals' self-ratings relate to how others see their performance (self-other agreement) and also the ratings they provide of others. These relationships may be examined by correlating different SRM components within and across performance dimensions and by correlating SRM components with self-ratings. These correlations should only be analyzed if the components of interest account for a non-trivial amount of variance (Back & Kenny, 2010; Kenny et al., 2006). Kenny (1994) recommends at least 10% of the total variance. The SRM correlations provide truer measures of these associations than traditional correlations based on raw scores (Marcus, 1998) because SRM correlations are not contaminated by irrelevant sources of variance. The theoretical questions addressed by examining the SRM intercorrelations are presented in Table 2 and will be discussed throughout the following sections.

Reciprocity: Do pairs of teammates see each other's performance uniquely? The focus of the SRM is on dyadic behavior and interpersonal perceptions, thus, reciprocity in behavior or perceptions is allowed (Kenny et al., 2001). Examinations of reciprocity require a two-sided process (Kenny, 1994). Reciprocity as conceptualized in the SRM can exist at both the individual and dyad level (Kenny et al., 2006; Miller & Kenny, 1986). Studies that used designs besides the SRM have generally failed to find reciprocity in attraction, perceptions, or behaviors because these studies failed to distinguish between *generalized* and *dyadic reciprocity* correlations (Kenny et al., 2006).

Generalized reciprocity operates at the individual level and indicates the extent to which people who are seen by others as behaving in a certain way or having a certain trait see others as behaving in the same way or as possessing the same trait. The existence of generalized reciprocity is assessed by computing the correlation between actor and partner effects (r_{ab}). For example, in the current study, the magnitude of this correlation determines to what extent seeing others as engaging in contextual performance is related to being seen by others as engaging in contextual performance. Previous research found limited support for generalized reciprocity correlations (Kenny, 1994; Greguras et al., 2001; 2007).

Dyadic reciprocity indicates the extent to which the two partners of the dyad see each other as behaving in the same way or having the same trait. For example, do Alice and Bob see each other's contextual performance in a different way than they see the contextual performance of other members of their group? If there is evidence of reciprocity then if Alice sees Bill's level of contextual performance as especially high, then Bill also sees Alice's level of contextual performance as especially high. Dyadic reciprocity is more analogous to the typical definition of reciprocity (Gouldner, 1960) used in organizational research. Thus, dyadic reciprocity is

examined by correlating the relationship component of each member of the dyad ($r_{gg'}$). Kenny (1994) found a high dyadic reciprocity for liking (.26 for short-term acquaintance and .61 for long-term association) and weak dyadic reciprocity in trait perceptions. Kenny (1999) suggested that researchers should examine prosocial constructs for evidence of reciprocity. Kenny et al. (2001) concluded that many prosocial behaviors (e.g., laughing, gesturing, and friendliness) have a high degree of dyadic reciprocity. Further, Greguras et al. (2007) found evidence of dyadic reciprocity only for the performance dimensions of team orientation and affiliation. These were the most interpersonal dimensions of performance assessed. Reciprocity has been theoretically linked to extra-role behaviors in the form of paying back the actions of others and in terms of expecting others to reciprocate altruistic behaviors that are performed towards them (Korsgaard, Meglino, Lester, & Jeong, 2010; Love & Forret, 2008). Subscribing to the norm of reciprocity has also predicted displays of this type of behavior (Yao & Wang, 2008; Coyle-Shapiro, 2002). Thus, I expect dyadic reciprocity to exist for ratings of the contextual performance because these behaviors are discretionary and therefore more likely to operate according to the norm of reciprocity (Kenny et al., 2001).[7]

Hypothesis 5a: Dyadic reciprocity will exist for ratings of contextual performance in co-located teams (i.e., $r_{gg',\text{co-located}}$ will be statistically significant).

The quality of exchange relationships has been linked to the performance of extra-role behaviors (Love & Forret, 2008). In high-quality relationships (i.e., those in which employees go beyond contractual obligations and share mutual liking and trust) the norm of reciprocity promotes engagement in behaviors related to interpersonal facilitation. Interpersonal processes

[7] The dyadic reciprocity of work contribution cannot be examined because this construct has a single indicator and thus the relationship effects cannot be estimated separately from error.

have rarely been studied within VTs (Martins, Gilson, & Maynard, 2004). Thus, the examination of dyadic reciprocity in VTs will help address this gap because reciprocity is a universal, interpersonal process. In VTs it is more difficult for employees to develop close interpersonal relationships; thus, I expect to see smaller levels of dyadic reciprocity in VTs.

Hypothesis 5b: Dyadic reciprocity will exist for ratings of contextual performance in VTs (i.e., $r_{gg'\ virtual}$ will be statistically significant).

Hypothesis 5c: Higher levels of dyadic reciprocity will exist for ratings of contextual performance in co-located teams than in virtual teams (i.e., $r_{gg'\ co\text{-}located} > r_{gg'\ virtual}$).

Halo bias and true halo: Do perceivers see targets the same regardless of the performance dimensions or do targets perform similarly across dimensions? Factor analyses of performance ratings generally reveal the existence of a large general performance factor (Viswesvaran, Schmidt, & Ones, 2005). Researchers have debated whether this factor exists because raters do not discriminate across dimensions of performance (i.e., halo bias) or because there are shared elements among job performance dimensions (i.e., true halo; Bretz et al., 1992). When multiple constructs are measured, the SRM analysis provides a unique opportunity to separate true similarity in performance across dimensions from similarity in ratings across the performance dimensions due to the rater characteristics (i.e., bias).

The extent of true halo in peer ratings of performance is assessed by correlating the target effect from one variable with the target effect from another variable (Kenny et al., 2006). If this correlation is statistically significant and positive then there is evidence that a target is seen similarly across performance dimensions and this similarity is due to characteristics of the target. For example, a significant target-target correlation ($r_{wl\ cp}$) in the current study would indicate that those individuals who contribute to the group project also demonstrate contextual performance.

Greguras et al. (2001; 2007) generally found significant and large target-target correlations, which supports the notion of a general performance factor. The target-target correlations will always be larger than the correlations between average peer ratings because average peer ratings are confounded by rater variance, relationship variance, and residual error (Greguras et al., 2007).

Hypothesis 6: Individuals in co-located teams and virtual teams who contribute to the group project will also demonstrate contextual performance behaviors (i.e., the target-target correlation ($r_{\text{wl cp}}$) will be statistically significant).

The extent of halo bias in peer ratings of performance is assessed by correlating the perceiver effect from one variable with the perceiver effect from another variable (Kenny et al., 2006). If this correlation is statistically significant and positive, then there is evidence that perceivers see targets similarly across performance dimensions and this similarity is due to characteristics of the perceiver. For example, a significant perceiver-perceiver correlation in the current study would indicate that perceivers who see targets as contributing to the group project also see the same targets as engaging in contextual performance. This correlation does not reflect tendencies of the targets to engage in these performance behaviors, but instead tendencies of the raters to develop an overall performance impression. Greguras et al. (2007) generally found significant perceiver-perceiver correlations. These correlations were smaller in magnitude than the target-target correlations; however, their presence still indicates halo bias may influence peer performance ratings.

Hypothesis 7: Perceivers in co-located teams and virtual teams will have a general tendency to see targets similarly for the performance dimensions of contributing to the

team task and contextual performance (i.e., the perceiver-perceiver correlation ($r_{wl\ cp}$) will be statistically significant).

To the best of my knowledge, researchers have not examined the extent of true halo and halo bias in virtual team members' peer performance ratings. Given the reduced information available to teammates of a VT, there is reason to suspect that the magnitude of true halo and halo bias may differ in VTs. For example, halo bias will likely play a larger role in ratings of team members because they do not have the same opportunity to observe behavior and thus, must rely on general impressions when making their ratings. Given the lack of prior research, these proposals are speculative and will be explored, but no formal hypotheses will be offered.

Self-other agreement: Do targets see themselves as others see them? Self-evaluations have been suggested as a mechanism for performance management in SMTs (Elloy, 2004; Manz & Sims, 1987). Elloy suggested self-regulation via self-assessment is essential in SMTs because of the lack of supervision by an external source. This form of performance management has been promoted further in VTs because of decreased opportunity for observation by not only outside observers (e.g., management), but also one's teammates (Bell & Kozlowski, 2002; Martins et al., 2004). Further, self-assessments are often used for development purposes. Self-awareness refers to the degree to which people see themselves as others see them (Wohlers & London, 1989). Poor self-awareness is generally indicated by a discrepancy between a self-rating and the mean of others' ratings or low correlations between self-ratings and the mean of others' ratings (e.g., Carless & Roberts-Thompson, 2001; Church, 1997). In general, correlations between self- and other-ratings are weak or non-significant (Harris & Schaubroeck, 1988; Mabe & West, 1982). Conway and Huffcutt (1997) reported a meta-analytic correlation of .19 between self- and peer-ratings.

Not only are self-ratings different from other-ratings, they are generally higher (Allen et al., 2000; Mabe & West, 1982; Taylor & Brown, 1988). This finding extends to ratings provided in face-to-face groups (e.g., Atwater & Yammarino, 1997; Harris & Schaubroeck, 1988; Mabe & West, 1982). A lack of insight into one's own performance is a cause for concern. Self-ratings of contextual performance were not related to selection or performance criteria (Beehr et al., 2001; Hui, Lam, & Law, 2000). Further, Anderson et al.'s (2006) study of perceptions of status in face-to-face groups (using the Kwan index[8]) revealed that people who had self-enhanced concepts of their status were liked less by their peers. Self-enhancement may be due to a greater knowledge of the domain of behavior (i.e., some extra-role behaviors are not visible to others). Further, raters may not have the ability to observe the ratee's behavior. For example, task performance may be more visible than certain extra-role behaviors. These behaviors are discretionary and likely relational, thus, teammates may not be equal recipients. Another reason for the low correlations between self- and peer-ratings is due to contamination of peer ratings in the form of perceiver, relationship, and random error variance in ratings.

Self-other agreement correlations calculated using the SRM target component are a more precise way to assess self-other agreement because the target variance components used to calculate this correlation are not confounded by perceiver, relationship, or random error variance. Self-perceptions cannot be partitioned into perceiver, target, and relationship effects because

[8] Kwan, John, Kenny, Bond, and Robins (2004) developed a self-enhancement index from SRM components: SE = S-T- P-G, where SE is self-enhancement, S is the self-rating, T is the target score, P is the perceiver score, and G is the group mean. This approach integrates self-insight (represented by the subtraction of target scores) and social-comparison (represented by the subtraction of perceiver scores). The subtraction of the group mean adjusts the self-enhancement score so that the zero scale-point indicates an unbiased self-perception and makes self-enhancement scores statistically independent of group membership.

they are not dyadic measures (Kenny, 1994).[9] Target-other agreement (r_{ob}) correlations generally show more favorable results than the correlations calculated using traditional methods. Self-other agreement for trait perceptions is relatively high (i.e., ranges from .42 to .70 for the Big Five; Kenny, 1994). The highest agreement is for traits that are associated with highly observable behaviors (e.g., extroversion, Kenny, 1994; talkative, Robins et al., 2004). There is also substantial agreement in ratings of leadership (r =. 50; e.g., Livi et al., 2008). The average self-target correlation found by Greguras et al. (2007) for perceptions of performance was .43. This correlation is much larger than average target-peer rating correlation found by Greguras et al. (2001). However, Greguras et al. (2001) did find large self-other correlations for the dimensions of cooperation. In general, these estimates are much larger than the meta-analytic estimates found by Conway and Huffcutt (1997, r = .19). People do appear to see themselves as others do – albeit a little better than others see them.

Weisband and Atwater (1999) found inflation in self-other ratings in VTs, but not in face-to-face teams. Interpersonal interactions and feedback influence self- and other-ratings (Ashford, 1989). Individuals may be more self-aware in a team environment because they have an opportunity to interact with each other and receive feedback on their actions. However, Weisband and Atwater (1999) suggest that self-enhancement in VTs may result from reduced feedback. Markey and Wells (2002) compared self-other agreement in personality ratings for people interacting in a virtual chat room using the SRM. They found significant agreement for extroversion (corrected r of .32) and openness (corrected r of .46) when in engaging in one-on-

[9] Self-ratings may be partitioned when a self-rating is provided for interactions with each dyadic partner separately. This approach may be useful when studying one-on-one interactions (e.g., Robins et al., 2004, Study 1), but is not appropriate for examining self-other rating agreement based on group interactions.

one interactions. They did not find significant agreement when engaging in group interactions because it was more difficult for the participants to distinguish the contributions of each team member and form separate impressions of each teammate's personality traits. Robins et al. (2004) compared self-other agreement in behavior after dyadic face-to-face, phone-mediated, and computer-mediated interactions. Agreement was higher in the face-to-face condition than in the phone or computer-mediated communication condition. To the best of my knowledge, no study has compared self-other agreement for performance ratings based on co-located and virtual team interactions. Based on the existing research, I expect participants to be more self-aware in face-to-face than in virtual teams and thus show stronger self-target agreement and less self-enhancement. Although in both types of teams participants are likely to display some level of self-enhancement.

Hypothesis 8a: In co-located and virtual teams, participants will display self-other agreement (i.e., the correlations between self-ratings and the SRM target effect ($r_{\text{ob co-located}}$ and $r_{\text{ob virtual}}$) will be statistically significant).

Hypothesis 8b: Self-target agreement is likely to be greater in co-located teams than in virtual teams (i.e., $r_{\text{ob co-located}} > r_{\text{ob virtual}}$).

Hypothesis 8c: Self-enhancement is likely to be greater in virtual than in co-located teams (i.e., the magnitude of the self-enhancement index will be larger in virtual teams than in co-located teams; $SE_{\text{virtual}} > SE_{\text{co-located}}$).

Chapter Two: Method

Participants and Procedure

Male and female students (n = 402) were recruited[10] in fall 2007, 2008, 2009, and 2010 from traditional and online sections of an MBA course at a large southeastern university. Participants were assigned by the course instructor to mixed gender teams of 3-5 people. Based on their course enrollment, participants were members of either virtual or co-located teams. These students were recruited because they worked with their teammates to complete an interdependent project. Participants interacted with other members of their team during the completion of the project over the course of several months. Those participants, who agreed to participate, completed Survey 1, which included demographic information before beginning the group project. Survey 2 was completed after the project deadline. As part of Survey 2, participants rated each of their teammates (i.e., completed round-robin ratings). Participants also provided self-ratings. These surveys were completed online. Ratings were confidential and used for research proposes.

Of these students, 375 provided data for the current study (93% response rate). One limitation of using the SRM is that if one team member fails to provide ratings the entire team must be excluded from the analysis (Kenny et al., 2006). Marcus (1998) stated that when only a couple of data points are missing from a round robin design these data could be replaced. The mean score was used for each dimension of the contextual performance construct, thus, when

[10] Permission was obtained from the course instructor to contact the participants and offer them the opportunity to participate in this study in exchange for course credit.

only a couple of data points are missing from a round robin design, the mean was calculated using the items that the participants did complete (Marcus, 1998). When participants failed to provide ratings of contribution (i.e., a one-item scale) or self-ratings the sample mean was used to replace the missing data.

If participants failed to provide round-robin ratings of their teammates' performance, they had to be dropped from the study. When possible I created a smaller size team (e.g., a five-member team reduced to four members or a four-member team reduced to three members). If the number of members in a team dropped below three then the whole team had to be dropped from the study. In total, 20 teams (59 participants) were removed from the study and 27 teams were reduced in size (25 participants were removed). The final sample size was 291 participants in 89 teams. More specific sample size information for the original and final samples is provided in Table 5. The number of participants (n) and the number of groups (g) is broken down by data collection session and type of team.

The round-robin design provides statistical advantages over other designs that generate fewer data points within a group (Lashley & Kenny, 1998; Kenny et al., 2006). Kenny (1994) suggested that a minimum of six groups is necessary to obtain stable estimates for groups containing four to eight members. However, more groups are likely necessary to obtain adequate levels of statistical power. Lashley and Kenny developed a method to determine the power of SRM test of variance and covariance.[11] In general, Lashley and Kenny found that few large

[11] Bond and Lashley (1996) derived formulas that provided theoretical variances for the round-robin design based on the assumption that the only source of variance in the SRM parameter is sampling error. If there is variance in the parameter across groups then Bond and Lashley's formula provides slight underestimates and when there is not variance across groups the formulas provide overestimates. Lashley and Kenny (1998) found that the SRM

groups are preferable to many small groups (i.e., more data per group leads to smaller variance in SRM parameter estimates and thus, greater power for significance tests). However, this study group sizes range from 3 to 5. Using estimates based on the overall findings from Greguras et al. 2007 [actor (perceiver) variance = .098, partner (target) variance = .245, relationship variance = .227, actor-partner covariance = .085, and dyadic covariance = .14], I estimated the number of 4-person groups needed to reach a .80 level of power using the statistical power computer program, AID-SRM, which is based on the article by Lashley and Kenny (1998). This program is available on the web at http://davidakenny.net/srm/srmp.htm. These estimates are provided in Table 6. The current sample size information for co-located and virtual teams of 3-members and of 4 or 5 members is presented in Table 7. According to the power analysis, there are an adequate number of teams to detect the perceiver, target, and dyadic covariance effects in each group when using 4 or 5 person teams. Further, my current sample size is similar to and in many cases larger than the sample size traditional found in SRM studies (e.g., Greguras et al., 2001).

For the final sample 69% of participants were male (n = 195) and 31% were female (n = 89). Most participants were Caucasian (80%; n = 226). Approximately 12% of participants were Asian/Pacific Islander and the rest of the participants self-reported their race as American Indian, African American, Hispanic, Multiracial, or chose to identify as a race not listed on the survey. The average age of participants was 27 years (*SD* = 5.6 years) and the participants ranged in age from 20 to 54 years. On average participants reported having approximately 5 years of work experience (*SD* = 5.5 years). The work experience of participants ranged from 0 to 28 years. These participants typically have work experience and are of a more similar age to employees

---

parameters do not appear to vary across groups based on the slight underestimations of Bond and Lashley's formulas.

in the workforce than typical student samples. Thus, this sample likely enhances the generalizability of these results to real world business settings more than other typical student samples. More specific demographic information for the final sample is provided in Table 8. Demographic information is provided for 3-person and 4/5-person co-located and virtual teams and for participants in co-located and virtual teams as a whole. A comparison of the demographic information for virtual and co-located teams indicated that co-located team members, on average, are younger, t (282) = -10.39, $p < .0001$, and less experienced, t (282) = -12.10, $p < .0001$, than virtual team members.

Measures

The Appendix contains the items for each of the measures used in this study. The instructions for each measure and description of the round-robin rating entry format are also provided.

Demographics. The following information was obtained from participants: gender, age (in years), race, and years of professional work experience. The following information was obtained from the instructors: team composition (i.e., the names of participants in each team for the purposes of sending out the survey), team size, and classroom structure (i.e., whether participants were enrolled in a traditional or distance-learning course).

Interdependence. Stewart and Barrick (2000) described interdependence as the "extent to which team members cooperate and work interactively to complete tasks" (p. 137). Researchers have called for measures of interdependence to be included in studies of teams (Mathieu et al., 2008; Kozlowski & Bell, 2003). In this study, interdependence was measured to ensure teams were engaging in an interactive task that would require team members to work together and thus, have a basis for formulating their ratings of contextual performance. Van de

Ven, Delbecq, and Koenig (1976) described four basic workflow processes based on Thompson (1967) typology. These workflow processes, in order from low to high task interdependence, are pooled/additive (i.e., work activities are performed separately and then combined), sequential (i.e., work activities flow in one direction between teammates), reciprocal workflows (i.e., work activities flow back and forth between teammates), and *intensive* (work activities are performed simultaneously through collaboration). The measure of team workflow was adapted from Tesluk, Mathieu, Zaccaro and Marks (1997) by Arthur, Edwards, Bell, Villado, and Bennett (2005).

Contextual performance. The two-dimensional Van Scotter and Motowidlo (1996) contextual performance scale was used. The scale dimensions were interpersonal facilitation and job dedication. The items for each dimension were rated using a five-point Likert scale (1 = to a very little extent; 5 = to a very great extent). Participants rated themselves and each of their teammates on the extent to which they performed the actions described in each statement. Interpersonal facilitation items described helpful, considerate, and cooperate actions by team members and job dedication items described persistence, effort, and self-discipline. An average score was calculated for each dimension, for the participants' self-rating and their rating of each of their teammates. For example, in a four-person team one participant would have eight contextual performance ratings (self-rating of interpersonal facilitation, self-rating of job dedication, rating of teammate 1's interpersonal facilitation, rating of teammate 1's job dedication and so on). A larger score indicates a higher level of contextual performance. Other ratings of interpersonal facilitation and job dedication will be used as indicators of an overall contextual performance construct for the SRM analyses. The internal consistency of both dimensions was high. For the self-data, the overall α for interpersonal facilitation was .83 (.82 in co-located teams and .83 in virtual teams) and the overall α for job dedication was .88 (.86 in co-located

teams and .89 in virtual teams). For the data provided by teammates (other ratings), the overall α for interpersonal facilitation was .83 (.85 in co-located teams and .82 in virtual teams) and the overall α for job dedication was .92 (.92 in co-located teams and .92 in virtual teams).

Contribution to Workload. Using a five-point Likert scale (1 = to a very little extent; 5 = to a very great extent), participants rated themselves and each of their teammates in terms of their contribution to the group project. Specifically, this item asked, "To what extent did each person do his/her fair share of the work?"

Chapter Three: Overview of Data Analysis

Establishing Measurement Invariance Using Confirmatory Factor Analysis

Many of the hypotheses in the current studies require comparisons between the ratings provided by participants who are part of co-located and virtual teams and also between ratings provided by the targets (self-ratings) and ratings provided by the perceivers (other-ratings). Kenny (1994) warned that self-ratings should not be included in SRM calculations because self-ratings may be qualitatively different from other-ratings. However, researchers using the SRM still make comparisons between self- and other-ratings (e.g., Kenny, 1994; Greguras et al., 2007). Researchers have also recognized the virtual and co-located teams may operate in fundamentally different ways (Bell & Kozlowski, 2002; Potter & Balthazard, 2002; Salas et al., 2006). These differences may lead virtual and co-located team members to differ in their understanding of what it means to engage in contextual performance behaviors. In a summary of the state of the group comparison literature, Vandenberg and Lance (2000) concluded that researchers rarely if ever test whether an instrument of measurement is operating in the same way for each group and whether the underlying construct being measured has the same theoretical structure for each group. However, to make meaningful comparisons, it is necessary to establish that the different groups of individuals, who are the focus of the comparison, interpret the contextual performance measure in the same way.

Before examining the invariance of ratings provided by co-located and virtual team members and for self- and other-ratings, I conducted a CFA using the full data set of self- and other-ratings (i.e., the participants from both virtual and co-located teams were included in the

same data set, but the data provided by self and other raters was analyzed separately). These analyses were necessary to first establish adequate fit of the contextual performance measurement model. After examining the fit of the measurement model using the full data set, I then examined the configural invariance and metric invariance for ratings provided by the target participant (i.e., self-ratings) and for ratings provided by the perceivers (i.e., other-ratings). Finally, I examined configural invariance and metric invariance for ratings provided by co-located and virtual team members. The establishment of configural and metric measurement invariance allows for comparisons between the ratings provided by the different groups of individuals. I examine only configural and metric invariance because it is widely accepted that testing for invariance of error variance and covariances is overly restrictive (Vandenberg & Lance, 2000).

Invariance tests for self-ratings and other-ratings. Woehr et al. (2005) described a procedure to establish invariance across ratings sources, while appropriately modeling performance ratings as a direct function of the performance dimension (i.e., what is being rated), rating source (i.e., who is providing the rating), and unique variance (error). Previous research used a multiple-group CFA methodology (e.g., Cheung, 1999; Maurer et al., 1998, Facteau & Craig, 2001; Lance & Bennett, 1997). When using a multiple-group CFA analysis, performance ratings are modeled as a function of the performance dimension and unique variance, and then the moderating effect of rating source on item loadings for each performance dimensions is examined. Woehr et al. (2005) proposed that because rating source also contributes to systematic and meaningful variance in performance ratings, a more appropriate method for invariance testing is based on the multitrait-multirater (MTMR) framework. When using MTMR CFA analysis there are latent factors for each trait (i.e., performance dimension) and for each rating

source. In the current study, I created a latent factor for the interpersonal facilitation and job dedication performance dimensions. I also created latent factors for self and other, which represent the two rating sources used in the current study. Each item is then specified to load onto the appropriate performance dimension and rater source. I used this approach to examine configural and metric invariance for self- and other-ratings.

Invariance tests for ratings provided by virtual and co-located team members. I used a multiple-groups CFA approach to examine the invariance of ratings provided by co-located and virtual team members. To establish configural invariance, the fit of the two-dimensional model of contextual performance is examined in each group separately (i.e., co-located team members and virtual team members). To establish metric invariance of the contextual performance scale, I fit the two-dimensional model to the data simultaneously for both groups and constrained the factor loadings to be equal across both groups. The establishment of configural and metric measurement invariance allows for comparisons between co-located and virtual team members.

Social Relations Model Calculations

In Appendix B of his book *Interpersonal Perception,* Kenny (1994) provides the statistical details for how the SRM components, variances, and correlations are calculated. In this section I borrow heavily from the examples and equations provided in that appendix. Equation 1 describes the decomposition of an individual's performance rating according to the SRM. X_{ij} is the performance score for person i's rating of person j (Kenny, 1994; Kenny et al., 2006).

$$X_{ij} = m + a_i + b_j + g_{ij} + e_{ijl} \quad (1)$$

m is the mean performance rating in the group of which i and j are members. All individuals in a group have the same mean score. The SRM components represent the portion of each person's rating that is attributed the general tendency of the perceiver, the general tendency

of the target, and the unique combination of the perceiver and target (i.e., their relationship). These components are computed for each team member in a group. For example, consider team members i and j:

a_i is person i's perceiver effect. The perceiver effect is the component of the performance rating that is due to the general tendencies of i (i.e., the person providing the rating) to see others in a certain way. This effect may be calculated by using Equation 2 in which $M_{..}$ equals the mean across all $n(n-1)$ scores, n equals the number of people in the group, $M_{i.}$ equals the mean score for perceiver i across $n-1$ targets, and $M_{.i}$ equals the mean score for target i across $n-1$ perceivers (Kenny, 1994):

$$\hat{a}_i = \frac{(n-1)^2}{n(n-2)} M_{i.} + \frac{n-1}{n(n-2)} M_{.i} - \frac{n-1}{n-2} M_{..} \qquad (2)$$

b_j is person j's target effect. The target effect is the component of the performance rating that is due to the general tendency of j (i.e., the person being rated) to be seen in a certain way. This effect may be calculated by using Equation 3 in which $M_{..}$ equals the mean across all $n(n-1)$ scores, n equals the number of people in the group, $M_{j.}$ equals the mean score for perceiver j across $n-1$ targets, and $M_{.j}$ equals the mean score for target j across $n-1$ perceivers (Kenny, 1994):

$$\hat{b}_j = \frac{(n-1)^2}{n(n-2)} M_{.j} + \frac{n-1}{n(n-2)} M_{j.} - \frac{n-1}{n-2} M_{..} \qquad (3)$$

g_{ij} is the relationship effect. The relationship effect is the component of the performance rating (X_{ij}) that is due to tendency of the perceiver to see a particular target in a unique way. The relationship effect for the performance score in which person i rates person j (g_{ij}) does necessarily have to match the relationship effect for the performance score in which person j

rates person I (g_{ji}). The relationship effect g_{ij} may be calculated by using Equation 4 in which $M_{..}$ equals the mean across all $n(n-1)$ scores, a_i is person i's perceiver effect, and b_j is person j's target effect (Kenny, 1994):

$$\hat{g}_{ij} = X_{ij} - \hat{a}_i - \hat{b}_j - M_{..} \quad (4)$$

e_{ijl} is the error in performance measure l for perceiver i and target j. The error in a performance rating can only be separated from the relationship effect if multiple indicators are used for the performance construct (Kenny, 1994; Kenny et al., 2006).

The SOREMO software program was used to analyze the round-robin data (Kenny, 1998). Calculating the SRM components is the first step towards conducting SRM analyses that are of primary interest to researchers. The SRM analysis can be thought of as a two-step process: (1) variance partitioning and (2) calculation of correlations. SOREMO estimates perceiver and target effects for each person in a group. The relationship effects are also computed for each person's performance rating of every other person in the group (i.e., two relationship effects are calculated for each pair of teammates). Thus, Kenny (1994) states that the perceiver and target effects are at the individual level and the relationship effects are at the dyadic level. SOREMO then calculates the variance in the perceiver effects, target effects, and relationship effects for each group. Finally, the mean for each variance estimate is computed across groups and tested for statistical significance.

The variables *a*, *b*, and *g* are random variables with a mean of zero and variances of σ_a^2, σ_b^2, $\sigma_{g.}^2$. Estimates of these variances are s_a^2, s_b^2, s_g^2, respectively (Kenny, 1994). The perceiver (s_a^2), target (s_b^2), and, relationship variance (s_g^2) are of interest to researchers using the SRM because they represent the portion of the variance in performance ratings that is attributable to the perceiver, target, and perceiver-by-target relationship, respectively. The SRM also includes a

covariance between perceiver and target (σ_{ab}) effects and a covariance between the relationship effects for each teammate pair ($\sigma_{gg'}$). All other covariances in the model are assumed to equal zero (Kenny, 1994). Estimates of these covariances are denoted as s_{ab} and $s_{gg'}$, respectively. For ease of interpretability, the correlation between the actor-partner effects (r_{ab}) and between the relationship effects for each teammate pair ($r_{gg'}$) are generally interpreted to determine the amount of generalized reciprocity and dyadic reciprocity, respectively. The generalized reciprocity correlation indicates the extent to which people who are seen by others as behaving in a certain way also see others as behaving in that same way. Dyadic reciprocity indicates the extent to which two individuals (e.g., a pair of teammates) sees each other as behaving in the same way.

As previously described, SOREMO computes effect estimates for each individual in a team; then, using these estimates, the mean squares and cross-products are calculated so that the parameters of interests (i.e., SRM variances; SRM correlations) may be estimated. A two-way random effects ANOVA is used to calculate the following mean squares estimates and ultimately estimate s_a^2, s_b^2, s_g^2 (Kenny, 1994). In contrast to a fixed-effects ANOVA, the purpose of using a random-effects ANOVA is not to estimate the difference between means, but instead to estimate the variance of the perceiver, target, and relationship effects (Kenny, 1994; Kenny et al., 2006).

The following equations are used to compute the mean squares for each of the factors (i.e., perceiver and target) and for the interaction (i.e., perceiver-by-target). The summations were across *n* persons (Kenny, 1994).

$$\text{Mean squares perceiver } (MS_p) = A = \frac{\sum \hat{a}_i^2}{n-1} \tag{5}$$

$$\text{Mean squares target } (MS_T) = B = \frac{\sum \hat{b}_i^2}{n-1} \tag{6}$$

$$\text{Mean squares for perceiver x target interaction} = C = \frac{\sum \hat{a}_i \hat{b}_i}{n-1} \quad (7)$$

SOREMO also estimates the average and difference for the relationship effects for each teammate pair (Kenny, 1994). In Equation 8 and 9, $\hat{g}_{ij}$ is person i's relationship effect with person j, and $\hat{g}_{ji}$ is person j's relationship effect with person i.

$$\text{Average} = \hat{e}_{ij} = .5(\hat{g}_{ij} + \hat{g}_{ji}) \quad (8)$$

$$\text{Difference} = \hat{d}_{ij} = \hat{g}_{ij} - \hat{g}_{ji} \quad (9)$$

Then, summing across $n(n\text{-}1)/2$ dyads (i.e., pairs of teammates), the mean squares between (MSB) and the mean squares within (MSW) are computed (Kenny, 1994).

$$\text{Mean squares between (MSB)} = D = \frac{2\sum \hat{e}_{ij}^2}{[(n-1)(n-1)/2] - 1} \quad (10)$$

$$\text{Mean squares within (MSW)} = E = \frac{\sum \hat{d}_{ij}^2}{(n-1)(n-1)} \quad (11)$$

The MSB and MSW were used to compute the relationship variance and covariance (Kenny, 1994). The relationship covariance is also known as the dyadic reciprocity covariance.

$$\text{Estimate of the relationship variance} = s_g^2 = (\text{MSB} + \text{MSW})/2 \quad (12)$$

$$\text{Estimate of the relationship covariance} = s_{gg'} = (\text{MSB-MSW})/2 \quad (13)$$

Once the relationship variance and covariance are estimated, then the perceiver-target covariance, perceiver variance, and target variance were estimated (Kenny, 1994).

$$\text{Estimate of the perceiver-target covariance} = s_{ab} = C - \frac{s_{gg'}(n-1)}{n(n-2)} - \frac{s_g^2}{n(n-2)} \quad (14)$$

$$\text{Estimate of the perceiver variance} = s_a^2 = A - \frac{s_g^2 (n-1)}{n(n-2)} - \frac{s_{gg'}}{n(n-2)} \quad (15)$$

$$\text{Estimate of the target variance} = s_b^2 = B - \frac{s_g^2 (n-1)}{n(n-2)} - \frac{s_{gg'}}{n(n-2)} \quad (16)$$

The variance estimates (s_a^2, s_b^2, s_g^2) are computed for each group, then average estimate across groups is tested using a one-sample t-test (*df* = g -1) to determine whether the mean effect is statistically different from zero (Kenny, 1994; Kenny et al., 2006).[12] This procedure was used to test hypotheses 1a, 1b, 2a, 2b, 3a, and 3b, which proposed that perceiver, target, and relationship variance would be significant for co-located and virtual teams. Hypothesis 1c, 2c, and 3c used an independent sample t-test (*df* = G -2) to determine if the mean of the variance estimates (i.e., perceiver, target, and relationship variance) for co-located and virtual teams are statistically different from each other. Hypothesis 4a and 4b used a series of dependent groups t-tests (*df* = g - 1) to determine if the mean variance estimates (i.e., perceiver, target, and relationship variance) are significantly different from each other for co-located and virtual teams.

Using the previously described equations, the SOREMO program calculated dyadic reciprocity covariance (Equation 13) and perceiver-target covariance (Equation 14), but for ease of interpretation, the correlation as opposed to the covariance is interpreted. Hypothesis 5a and 5b were examined using a one-sample *t* test ($g - 1$) to determine if the dyadic reciprocity correlation was statistically different from zero for co-located and virtual teams. Hypothesis 5c was examined using an independent sample *t* test ($G - 2$) to determine if the dyadic reciprocity covariance (correlation) was statistically different in co-located and virtual teams. Although no formal hypotheses were proposed, the same test was used to examine the significance of the perceiver-target correlation.

[12] When group sizes vary, the estimates are weighted by $n - 1$.

To estimate all of the parameters of the SRM a minimum team size of 4 is needed for the round-robin design (Kenny, 1994; Kenny et al., 2006). Three participant teams may be used when dyadic reciprocity is constrained to be zero. Using the data set consisting of teams with four or more members the extent of dyadic reciprocity was examined. If the dyadic reciprocity correlation is significant then only those teams with four or more members will be used to examine the hypotheses. If the dyadic reciprocity correlation is not significant then teams with three or more members will be used to examine the hypotheses.

Once the group mean is removed, the other components are independent of group (Kenny et al., 2006). The variability of the target and perceiver component remains the same regardless of the number of raters (i.e., estimates are calculated with a correction based on the number of perceivers and the relationship variance). Thus, the perceiver, target, and relationship variance components are theoretical estimates based on an infinite number of raters (Kenny et al., 2006). The SOREMO program also disattenuates all of the correlations to correct for the reliability of the components (Marcus, 1998).

To this point, the hypotheses have involved univariate analyses (i.e., only one construct of interest). The SRM may also be applied used as a multivariate analysis technique to examine tendencies in perceivers and targets across constructs. The SOREMO program was used to calculate the target-target covariance (correlation) and the perceiver-perceiver covariance (correlation) for the constructs of workload contribution and contextual performance including interpersonal facilitation and job dedication. To compute these correlations both constructs are decomposed, as previously described, using the SRM analysis (Kenny, 1994). The target-target correlation and the perceiver-perceiver correlation are tested using the group as the unit of analysis. To test hypothesis 6 and 7 the estimate is computed for each group. A one-sample *t* test

(g – 1) is used to test whether the mean correlation for virtual and co-located teams is statistically different from zero.

The SOREMO program may also be used to calculate the self-target correlations. The target and perceiver effects are first calculated for the round-robin ratings provided by other teammates (Kenny, 1994). Then to test hypothesis 8a, a Pearson product-moment correlation was computed between target effects and self-ratings, when groups are partialed out. This correlation is tested for statistical significance (with $df = N – G – 1$), N = number of people and G = number of groups. An independent sample t test ($G – 2$) was used to test hypothesis 8b and determine if self-target agreement was greater in co-located teams than in virtual teams. Finally the average magnitude of the self-enhancement index was compared using an independent sample t test ($G – 2$) to examine hypothesis 8c, which predicted that self-enhancement is greater in virtual than in co-located teams.

Chapter Four: Results

I first present the results of the confirmatory factor analyses to establish the fit of the measurement model of contextual performance, the invariance of self- and other-ratings, and the invariance of ratings provided by co-located and virtual team members. I then present descriptive statistics that are relevant to understanding the data. Finally, I present the results of the SRM analysis for the hypotheses.

Confirmatory Factor Analysis

I conducted a separate CFA on the self-ratings and other-ratings provided by all participants from virtual and co-located teams to establish adequate fit of the measurement model before conducting the invariance tests for self-ratings and other-ratings and for co-located and virtual team members. I imported the raw data into Mplus (Muthén & Muthén, 1998-2010). For all CFAs, I fixed the variance of the factors to 1.0 in order to scale the latent variables. For the self-ratings, each rating was used as a manifest indicator of its respective latent construct. For ratings provided by others, the average of the ratings provided by others on each item for a participant was calculated and used as a manifest indicator of latent construct. This procedure was necessary due to the round-robin rating format in the current study. To the best of my knowledge a statistical method for conducting a confirmatory factor analysis and invariance testing for round-robin data has not been developed.[13]

[13] The self- and other-ratings in the current data sets are nested with team; however, there are too few groups (i.e., level-two data points) to conduct a multilevel CFA. Thus, all CFAs were conducted without accounting for group membership and the results should be interpreted with some caution because standard errors may be underestimated

I compared a two-factor model in which items were manifest indicators of two latent variables (i.e., interpersonal facilitation and job dedication) to a one-factor model in which items were indicators of a single latent variable (i.e., contextual performance). In both models, all factor loadings and error variances of the indicators were freely estimated. In the two-factor model, the covariance between the latent constructs was also freely estimated. In the one-factor model the correlation between the latent constructs was constrained to equal one. An inspection of the distributional properties of the indicators demonstrated univariate and multivariate normality. Very small amounts of data were missing from the covariance matrix (i.e., on average 99% coverage). Thus, the model was estimated using full information maximum likelihood imputation. Model fit was assessed with the chi-square statistic and several fit indices such as the root mean square error of approximation (RMSEA; Steiger, 1990), the comparative fit index (CFI; Bentler, 1990), and the standardized root mean residual (SRMR). Kline (2005) describes CFI values greater than .90 and SRMR values below .10 as favorable. Hu and Bentler (1999) suggested more stringent values for the CFI (< .95) and SRMR (< .08) as well as a cutoff of close to .06 for RMSEA.

The fit indices for the one-factor and two-factor model solutions are presented in Table 9 for the self-ratings and Table 10 the other-ratings. In general, the one and two-factor models of contextual performance did not meet the desired fit statistics. The model fit was improved by removing one item from the interpersonal facilitation sub-dimension (i.e., Treated others fairly). The removal of this one item will not substantially reduce the construct validity of scale, as there

---

and an inflated type I error rate (Bovaird, 2007). Also, using the average of other-ratings provided by the round-robin data collection procedure may affect the results of the CFA analysis. However, these actions were deemed most appropriate given the sample size limitations and no established procedure for conducting a CFA with round-robin data.

are still six remaining items that provide substantial coverage the content domain. Even after removing one item, the fit of the two-factor model of contextual performance can be considered marginal, at best, in the current data set. Specifically, the CFI and SRMR fit indices for other ratings were favorable according to Kline (2005), but the RMSEA value is above the acceptable cut-off. Similarly, the SRMR value for the two-factor model using self-ratings was favorable; however, the CFI value was below and RMSEA value was above the acceptable cutoff. However, for both the self-ratings and other-ratings, a χ^2 difference test revealed that the two-factor solution is a more appropriate solution than the one-factor solution in the current data set (see Table 9 and Table 10). Thus, I chose to retain the two-factor model for both self- and other-ratings.

Establishing invariance of ratings provided by self and other raters. A MTMR CFA (Woehr et al., 2005) was used to assess the invariance of self-ratings and other-ratings of interpersonal facilitation and job dedication. Each item was specified to load on a performance dimension (i.e., job dedication or interpersonal facilitation) and a rater source (i.e., self or other). Figure R1 presents the final MTMR model. An initial model was assessed in which all items (except the one item dropped as a result of the earlier CFAs) loaded on the appropriate performance dimension and rater source latent factors. As can be seen in Figure 1 the performance dimensions were correlated as were the rater source dimensions (Woehr et al., 2005). All other correlations between latent factors were set equal to zero. Because the exact same items were completed by self and other raters the error terms for these manifest indicators were correlated.

Configural Invariance. The establishment of the same factor structure (i.e., items loading onto the same performance dimensions) for self- and other-ratings demonstrates

configural invariance (Woehr et al., 2005). The factor loadings for the configural invariance CFA model are presented in Table 11. As can be seen by examining the factor loadings in the first half of the table (Model 1), the self-ratings for two items on the job dedication scale (Put in extra hours to get work done on time; Worked harder than necessary) were not statistically significant; however, the other-ratings for the same items were statistically significant. These results do not support configural invariance for the original scale (Model 1). I removed these two items and reran the analysis. All factor loadings for the revised scale (Model 2) significantly loaded onto the specified performance factors. These results support the configural invariance of self-and other-ratings for the revised scale. Model fit was assessed using the chi-square statistic, CFI, and SRMR (see Table 12). The fit indices for the configural models also support the use of the revised model.

Metric Invariance. The MTMR CFA approach for testing metric invariance involves constraining the factor loading for self-rated and other-rated items to be equal. As can be seen in Table 12 the $\Delta\chi^2$ significance test was not significant, which indicates support for metric invariance. However, an examination of the other fit indices reveals that the best fitting model is the configural invariance model. In sum, for the revised measure (Model 2) the items load onto the factors of interpersonal facilitation and job dedication consistently for self-ratings and other-ratings. The $\Delta\chi^2$ significance test also provides some support that the factor loadings for items completed by self and other raters are of similar magnitude; however, given the deterioration of the other fit indices this conclusion should be made with caution.

In the next section, the invariance of ratings provided by co-located and virtual team members will be examined. In these analyses all but one item of the interpersonal facilitation scale (item 6) and all but two items (item 1 and item 3) of the job dedication scale will be used as

indicators of the latent performance constructs. Given that these items are assumed to be parallel indicators of a content domain, the removal of these items should not substantially reduce the content validity as six items remain for each scale.

Establishing invariance of ratings provided by co-located and virtual team members.

Configural Invariance. I used a multiple-groups CFA approach to examine the invariance of ratings provided by co-located and virtual team members. I established the configural invariance of the two-dimensional contextual performance measure for each group (i.e., co-located team members and virtual team members) by examining the fit of the model in each group separately. This analysis was conducted using both the self-rating data set and the other-rating data set. Model fit was assessed with the chi-square statistic, RMSEA, CFI, and SRMR (see Table 13). The fit statistics for the two-dimensional model of self-ratings in virtual teams met the recommendation provided by Hu and Bentler (1999) for SRMR, but did not meet the recommended cut-offs for CFI or RMSEA. The fit statistics for the two-dimensional model of self-ratings for co-located teams met recommended cut-off for SRMR provided by Hu and Bentler and the recommended cut-off for CFI provided by Kline (2005), but just missed the recommended value for RMSEA. The fit statistics for other-ratings for co-located and virtual teams met the recommendation by Hu and Bentler for SRMR (1999) and the CFI recommendation provided by Kline, but missed the recommended cut-off for RMSEA. Taken together these results provide support for comparison of other-ratings provided by virtual and co-located team members. Self-ratings provided by co-located team members did not demonstrate adequate fit in the current data set, thus, comparisons between the self-ratings provided by virtual and co-located team members should be made with caution.

Metric Invariance. I examined the metric invariance of the two-dimensional contextual performance model by fitting the model simultaneously for both groups. The baseline model, which hypothesized no constrained factor loadings, is used as the model against which all subsequent models are compared. In the second model, all factor loadings were constrained to be equal across the groups. As seen in Table 14, the change in the $\Delta\chi^2$ tests were not statistically significant, which supports the conclusion of metric invariance (Vandenberg, 2002).Thus, the appropriateness of comparisons between other-ratings provided by co-located and virtual team members is supported. However, given the initial poor fit of the two-dimensional configural invariance model in the co-located, comparisons between self-ratings provided by co-located and virtual team members should be made with caution.

Descriptive Analyses

Data were collected from at four different points in time (the fall semester of four consecutive years). Participants recruited in years 1 and 2 were taught by the same instructor. A different instructor taught the same course in years 3 and 4. There were significant mean differences in other ratings of workload contribution, interpersonal facilitation, and job dedication based on course instructor (see Table 15). Mean differences based on course instructor for other-rated task interdependence and for all self-ratings were not statistically different. Bonferroni post-hoc comparisons reveal that only statistically significant mean difference based on semester of data collection is for the construct of workload contribution, such that the mean level of workload contribution in semester 2 is different from the mean level of workload contribution in semester 4 (p = .028). Further, the mean level differences based on instructor and semesters are of minimal concern because the SRM analysis removes the group mean when calculating the other SRM components. The removal of the group mean also

indirectly addresses any differences in the mean level of the constructs based on instructor or semester.

Descriptive information (i.e., means, standard deviations, and correlations) for the study variables (both self-ratings and other-ratings) are provided in Table 16. The alpha coefficients for each of the scales are presented on the diagonal of Table 16. All measures demonstrated adequate internal consistency. The primary variables in the current study are all forms of performance (i.e., contribution to the team's workload, interpersonal facilitation behaviors, and job dedication behaviors), thus, it is not surprising that these variables are significantly and positively correlated (for both self- and other-ratings). Further, the extent to which individuals perceive the task as requiring interdependence is significantly and positively correlated with the various types of performance.

Table 17 presents the means, standard deviations, and correlations for the study variables by team type (i.e., co-located or virtual teams). An examination of the level of interpersonal facilitation and job dedication ratings provided by both the self and others in co-located and virtual teams revealed that ratings are higher in co-located than in virtual teams. Interestingly, there was not a statistically significant difference in the task interdependence ratings provided by participants in co-located and virtual teams. However, task interdependence was statistically and positively related to all performance dimensions in virtual teams, but in co-located teams, task interdependence was only statistically and positively related to workload contribution and job dedication ratings provided by others. Although, not the primary focus of this study, this descriptive information does indicate that there are statistically significant mean level differences in performance ratings in co-located and virtual teams. Further, the differential pattern of relationships between task interdependence and the performance dimensions in co-located and

virtual teams provides some evidence that interpersonal interactions differ in co-located and virtual teams and these differences may influence performance ratings. In the next section, I use the SRM to examine peer performance ratings as an interpersonal perception process, including the impact of the perceiver, target, and perceiver-target relationship on peer ratings. I also make comparisons between SRM components in co-located and virtual teams and between SRM components and self-ratings.

Social Relations Model Analysis

The separation of relationship variance from random error variance requires the use of multiple indicators of an overall performance construct (Kenny et al., 2001). Given the high correlation between the interpersonal facilitation dimension and job dedication dimension, these two dimensions were used as indicators of the overall contextual performance construct. Because multiple scales were used as indicators of the overall performance construct the variance in these construct scale scores can be separated into stable and unstable variance. Stable variance is the variance that is consistent across replications and unstable variance is the variance that is unique to each replication (Kenny, 1998). For co-located teams, .74 of the variance in the overall contextual performance construct was stable and for virtual teams, .76 of the variance was stable. The relative portion of the stable variance in the overall contextual performance construct was then examined and divided into target, perceiver, relationship, and error variance. For the dimensions of interpersonal facilitation, job dedication, and workload contribution, the variance cannot be separated into stable and unstable components (because only one indicator was used for each construct). Thus, the relative portion of the total variance is examined and divided into target, perceiver, and combined relationship and error variance. In the following sections, I

present the results for the overall performance construct as well as the interpersonal and job dedication dimensions, and workload contribution.[14]

Variance partitioning. The variance partitioning analyses were calculated for performance ratings provided by co-located and virtual team members separately. The relative estimates for the target, perceiver, and relationship variance are presented in Table 18.The target and perceiver variance is meaningful to interpret for all constructs; however, the relationship variance cannot be separated from residual variance for workload contribution, interpersonal facilitation, and job dedication because there is only one indicator for each of these performance dimensions. The combined relationship and error variance is presented in Table 18 for each of the scales. However, as recommend by Ingraham and Wright (1986), I did not assess the significance of the combined relationship and error variance because it is unclear what portion of the variance is attributable to the relationship and what portion is attributable to random error. I only assessed the significance of the relationship variance component for the overall contextual performance construct.

A one-sample t-test (*df* = g -1) was used to determine whether the average variance estimates (s_a^2, s_b^2, s_g^2) across groups is statistically different from zero (Kenny, 1994; Kenny et al., 2006). One-tailed significance tests are used because theoretically variances should always be positive (Kenny, 1994). The relative variance components were presented for each performance construct for ease of interpretation; however, all tests of significance were performed on the absolute variance components. This procedure was used to test several hypotheses, which proposed that target (hypothesis 1a, 1b), perceiver (hypothesis 2a, 2b), and

[14] As documented in a later section of the results, there is evidence of dyadic reciprocity, thus, all analyses were conducted on the subset of teams that had four or more members.

relationship variance (hypothesis 3a, 3b) were significant for co-located and virtual teams, respectively. As can be seen by examining the top half of column two in Table 18, Hypothesis 1a was not supported because the target variance component ($s_b^2{}_{co\text{-}located}$) was not statistically significant. Perceivers in co-located teams did not demonstrate a significant level of consensus in their ratings of overall contextual performance, interpersonal facilitation, job dedication, or workload contribution. However, this amount of variance did exceed the recommendation by Kenny (1994; i.e., 10%), thus, it may be considered non-trivial. Hypothesis 1b was supported because the target variance ($s_b^2{}_{virtual}$) was statistically significant in virtual teams as perceivers demonstrated consensus in their ratings of peer overall contextual performance, interpersonal facilitation, job dedication, and workload contribution (see bottom half of column two in Table 18). Hypothesis 2a and hypothesis 2b were supported because the perceiver variance in co-located ($s_a^2{}_{co\text{-}located}$) and virtual teams ($s_a^2{}_{virtual}$) was statistically significant. Perceivers in both types of teams demonstrated assimilation in their ratings of peer overall contextual performance, interpersonal facilitation, job dedication, and workload contribution (see column three in Table 18). Hypothesis 3a and hypothesis 3b were supported because the relationship variance in co-located ($s_g^2{}_{co\text{-}located}$) and virtual teams ($s_g^2{}_{virtual}$) was statistically significant. Perceivers in both teams demonstrated uniqueness (i.e., saw targets in idiosyncratic ways) for the overall contextual performance measure.[15]

An independent sample t-test (df = G -2) was used to test hypothesis 1c, 2c, and 3c and determine if the mean of the variance estimates (i.e., target, perceiver, and relationship variance) for co-located and virtual teams are statistically different from each other. As can be seen in

[15] Only the magnitude of relationship variance for the overall contextual performance construct was tested for significance because this was the only construct for which there were multiple indicators.

Table 19 (column 2) there were not significant differences in the amount of consensus in co-located and virtual teams, thus, Hypothesis 1c which proposed that consensus would be higher in co-located than in VTs (i.e., $s_b^2{}_{\text{co-located}} > s_b^2{}_{\text{virtual}}$) was not supported. Although not a statistically significant difference, the magnitude of the target variance was actually higher in virtual teams (d = .41 for work load contribution, d = .10 for interpersonal facilitation, d = .41 for job dedication, and d = .16 for the overall contextual performance construct), which indicates that virtual team members may actually demonstrate more consensus in their ratings than co-located team members. Further, these differences are of a moderate size for the dimensions of workload contribution and job dedication based on Cohen's d effect size estimates (Cohen, 1988). Differences in the amount of assimilation in co-located and virtual teams were also not statistically significant (see Table 19, column 3), thus, Hypothesis 2c which proposed that assimilation would be higher in virtual than in co-located teams ($s_a^2{}_{\text{virtual}} > s_a^2{}_{\text{co-located}}$) was not supported. However, again the effect size estimates revealed moderate effects for the dimensions of workload (d = .46) and job dedication (d = .42) and smaller effects for interpersonal facilitation (d = .14) and the overall contextual performance construct (d = .30). Finally, there were not statistically significant differences in the amount of assimilation in co-located and virtual teams (see Table 19 column 4), thus, Hypothesis 3c which proposed that amount of uniqueness will be larger in co-located teams than in VTs (i.e., $s_g^2{}_{\text{co-located}} > s_g^2{}_{\text{virtual}}$) was not supported. Again, an examination of the effect size estimates revealed that there were moderate differences for the interpersonal facilitation dimension (d = .49) and smaller differences for the other dimensions (d ranged from .19 to .31).

Dependent group t-tests (df = g -1) were used to test hypothesis 4a and 4b and determine if the mean estimates for target, perceiver, and relationship variance are significantly different

from each other. Table 20 presents the t-test results. Hypothesis 4a proposed that in co-located teams, raters demonstrate similar amounts of consensus and uniqueness and smaller amounts of assimilation (i.e., $s_b^2_{\text{co-located}} \approx s_g^2_{\text{co-located}} > s_a^2_{\text{co-located}}$). The pattern of results across all four performance constructs shows that in co-located teams raters demonstrated larger amounts of assimilation (perceiver variance) than consensus (target variance). Examination of the overall contextual performance construct reveals that in co-located teams, raters demonstrate larger amounts of assimilation (perceiver variance) than uniqueness (relationship variance), but similar amounts of consensus (target variance) and uniqueness (relationship variance). These results (i.e., $s_b^2_{\text{co-located}} \approx s_g^2_{\text{co-located}} < s_a^2_{\text{co-located}}$) only partially support the proposed pattern of variance magnitudes in Hypothesis 4a. Target variance ($s_b^2_{\text{co-located}}$) and relationship variance ($s_g^2_{\text{co-located}}$) are similar in magnitude, but these variance components are smaller (not larger) than perceiver variance ($s_a^2_{\text{co-located}}$). This pattern of results is further demonstrated by a relative comparison of the magnitude of the variances: The perceiver variance is 4.15 times larger than the target variance and 7.71 times larger than the relationship variance for the overall performance measure. Further, target variance was 1.86 times larger than the relationship variance.

Hypothesis 4b proposed that raters in VTs would demonstrate similar amounts of consensus and assimilation (i.e., $s_b^2_{\text{co-located}} \approx s_a^2_{\text{co-located}}$). The results support this hypothesis for three out of the four constructs as there were not statistically significant differences in the magnitude of the perceiver and target variance components for the overall contextual performance construct, workload contribution, or job dedication. However, there was a statistically significant difference in the magnitude of the perceiver and target variance components for interpersonal facilitation. Further, examination of the overall performance construct indicates that the magnitude of the perceiver variance, but not the target variance was

significantly larger than the relationship variance. Taken together, these findings partially support the proposed pattern of variance magnitudes in Hypothesis 4b. Target variance ($s_b^2{}_{virtual}$) and perceiver variance ($s_a^2{}_{virtual}$) are similar in magnitude; however, only the magnitude of the perceiver variance ($s_a^2{}_{virtual}$) is statistically larger than the magnitude of the relationship variance ($s_g^2{}_{co\text{-}located}$). Again, this pattern of results is further illustrated by a relative comparison of the magnitude of the variances: In virtual teams, the perceiver variance is only 1.50 times larger than the target variance and 3.55 times larger than the relationship variance for the overall contextual performance measure. Further, target variance was 2.36 times larger than the relationship variance.

Correlations between SRM components.

Dyadic reciprocity. The examination of the correlations between SRM components is only advisable when the variance of each of the SRM components that are being correlated is statistically significant or exists at a non-trivial level (Kenny, 1994; Kenny et al., 2006). The relationship variance for the overall contextual performance measure was statistically significant for both co-located and virtual teams, thus, it is appropriate to examine whether the dyadic reciprocity covariance (correlation) is statistical significant. The significance of these relationships was tested using a one-sample *t* test ($g - 1$) to determine if they were statistically different from zero. Hypothesis 5a proposed that dyadic reciprocity would exist for ratings of contextual performance in co-located teams. The magnitude of the dyadic reciprocity relationship in co-located teams ($r_{gg'\ co\text{-}located} = .34$) was statistically significant, $t\ (22) = 2.11$, $p = .046$. Thus, hypothesis 5a was supported. Hypothesis 5b proposed that dyadic reciprocity would exist for ratings of contextual performance in virtual teams. The magnitude of the dyadic reciprocity relationship in virtual teams ($r_{gg'\ virtual} = .07$) was not statistically significant, $t\ (22) =$

0.38, $p = .706$. Thus, hypothesis 5b was not supported. Hypothesis 5c proposed that higher levels of dyadic reciprocity would exist for ratings of contextual performance in co-located teams than in virtual teams. This hypothesis was examined using an independent-sample *t* test ($G - 2$). Despite the significance of the dyadic reciprocity relationship in only the co-located teams, hypothesis 5c was not supported by statistical significance tests, $t(44) = 1.18$, $p = .244$. However, the average magnitude of the dyadic reciprocity correlation was larger in co-located teams than in virtual teams (i.e., $r_{gg',\text{co-located}} > r_{gg',\text{virtual}}$) and the effect size estimate indicates that the difference is of small to medium size ($d = .35$).

Generalized reciprocity. The target variance for the overall contextual performance, workload contribution, interpersonal facilitation, and job dedication was not statistically significant for co-located teams. However, the level of variance is above the recommendation provided by Kenny (1994). The target and perceiver variances were statistically significant for the virtual teams. Thus, the perceiver-target correlation (i.e., r_{ab}) was examined for both co-located and virtual teams using a one-sample *t* test ($g - 1$) to determine if the generalized reciprocity correlation was statistically different from zero. The results indicate that the generalized reciprocity was not statistically significant for any of the performance constructs in co-located and virtual teams. As this analysis was not a focus of the current paper, the specific results are not presented here.

Halo bias and true halo. Hypothesis 6 proposed that individuals in co-located teams and virtual teams who contribute to the group project will also demonstrate contextual performance behaviors (i.e., the target-target correlations ($r_{\text{wl cp}}$) will be statistically significant). Comparisons were made between the dimensions of workload contribution, job dedication, and interpersonal facilitation because these are the basic performance dimensions of the study (i.e., the overall

contextual performance dimensions is a combination of interpersonal facilitation and job dedication). To test hypothesis 6 correlations between the target effects for each performance dimension are computed for each group and a one-sample *t* test is used to determine whether the mean correlation for virtual and co-located teams is statistically different from zero. The target-target correlations between the performance dimensions are statistically significant, which supports hypothesis 6 (see Table 21). Further, a comparison of the target-target and the other-other correlations in Table 17 reveals that the target-target correlations are much higher.

Hypothesis 7 proposed that perceivers in co-located teams and virtual teams will have a general tendency to see targets similarly for the performance dimensions of workload contribution, interpersonal facilitation, and job dedication (i.e., the perceiver-perceiver correlations ($r_{\text{wl cp}}$) will be statistically significant). To test hypothesis 7 correlations between the perceiver effects for each performance dimensions are computed for each group and a one-sample *t* test is used to determine whether the mean correlation for virtual and co-located teams is statically different from zero. The perceiver-perceiver correlations between the performance dimensions are statistically significant, which supports hypothesis 7. As can be seen by reviewing Table 21 the perceiver-perceiver and target-target correlations were not substantially different in co-located and virtual teams. Thus, the magnitude of halo bias and true halo appear to be similar.

Self-other comparisons. Hypothesis 8a proposed that in both co-located and virtual teams, participants would display self-other agreement (i.e., the correlations between self-ratings and the SRM target effects ($r_{\text{ob co-located}}$ and $r_{\text{ob virtual}}$) would be statistically significant). A Pearson product-moment between the self-rating and target effect was calculated for each group. This correlation was tested for statistical significance (with $df = N - G - 1$; N = number of people and

G = number of groups). As previously stated, the target variance was not significant for co-located teams, thus, it is only appropriate to interpret the significance of the self-target correlations for virtual teams. Kenny et al. (2006) warned that when target effects do not account for substantial variance in performance ratings, correlations between the target effects and other variables (e.g., self-ratings) might be inflated. The magnitudes of the self-target correlations are presented in Table 22. Hypothesis 8a was supported for virtual teams. Hypothesis 8b proposed that self-target agreement would be greater in co-located teams than in virtual teams (i.e., $r_{\text{ob co-located}} > r_{\text{ob virtual}}$). An examination of the results in Table 22, indicates that the opposite pattern of effects (i.e., $r_{\text{ob virtual}} > r_{\text{ob co-located}}$). Thus, hypothesis 8b was not supported.

The means for self-ratings and other-ratings for the dimensions of workload contribution, interpersonal facilitation, and job dedication are presented in Table 23 for both co-located and virtual teams. The average magnitude of the self-enhancement index is also presented. The self-enhancement index was calculated by subtracting the perceiver effect, target effect, and group mean from the self-rating. A positive value is an indication of self-enhancement. An independent samples t-test (*df* = G -2) was used to examine Hypothesis 8c, which predicted that self-enhancement is larger in virtual than in co-located teams. The independent sample t-tests does not support a significant difference in the magnitude of the self-enhancement index in virtual and co-located teams: workload contribution, t (44) = -*.87, p = .39,* interpersonal facilitation, t (44) = .28, p = .783, job dedication, t (44) = -.363, p = .72. Thus, hypothesis 8c was not supported.

Exploratory Analyses. As can be seen in Table 17 and Table 18, participants rated the interdependence of the task at approximately 4 out of 5, which indicates a reciprocal level of interdependence (i.e., work activities flow back and forth between teammates; Arthur et al., 2005; Tesluk et al., 1997 Thompson, 1967). Further, the level of task interdependence was not

statistically different from co-located and virtual teams. What is interesting to note is that in co-located teams, participants' ratings of task interdependence were statistically, positively related to perceiver effects for the dimensions for workload contribution (r = .31) and job dedication (r = .29). However, the relationship between participants' ratings of task interdependence and interpersonal facilitation was not statistically significant (r = .15). In virtual teams, participants' ratings of task interdependence were not statistically related to perceiver effects for any of the performance dimensions (r ranged from .02 to .19). Further, in both co-located and virtual teams, participants' ratings of task interdependence were not statistically related to target effects for any of the performance dimensions (r ranged from -.09 to .12).

Chapter Five: Discussion

Peer ratings have become an essential source for performance assessment in many organizations. The prevalent use of SMTs (Bell & Kozlowski, 2002) among other organizational conditions (e.g., larger span of supervision) has reduced the opportunity for supervisors to observe employees' performance. Teammates offer an intriguing alternative perspective for performance appraisal because they are able to evaluate taskwork (i.e., workload contribution) and teamwork (i.e., contextual performance; Fletcher, 2001) behaviors. Self-evaluations have also been proposed as a mechanism to regulate team performance (Manz & Sims, 1986). However, potential concerns exist regarding leniency in self-ratings and the influence of bias in peer ratings due to workplace friendships (Cederblom & Lounsbury, 1980; Doll & Longo, 1962; Landy & Farr, 1983; Love, 1981). Thus, researchers have called for thorough investigations into the use of self- and peer-ratings (Bretz et al., 1992; Pulakos et al., 1996).

Examining performance appraisal as a person perception process allows researchers to gain insight into which factors contribute to the variance in performance ratings. Previous research examining the variance in peer performance ratings has documented the presence of large amounts of perceiver variance, but has generally been unable to explore if this variance is attributable to differences in tendencies between perceivers or due to specific perceiver-target relationships[16] (i.e., most studies used a data collection design in which targets were nested within perceivers; Murphy & DeShon, 2000). The current study avoids this major limitation by

[16] For exceptions, see previous SRM analyses of peer performance ratings (i.e., Boldry & Kashy, 1999; Greguras et al., 2001; 2007; Sullivan & Reno, 1999).

using round-robin data and the SRM (Kashy & Kenny, 2000) to provide insight into the influence of peer relationships, as well as individual perceiver and target tendencies on performance ratings (Pulakos et al., 1996). Further, most of the research on self-ratings and all of the existing research on other-ratings has focused on co-located teams. Researchers have called for comparisons between co-located and virtual teams (e.g., Latham & Mann, 2006; MacDuffie, 2008; Potter & Balthazard, 2002) and more broadly for examination of the situational factors that are likely to influence person perception (Kenny et al., 2001; Martins et al., 2004). Unlike co-located team members, virtual team members are distributed across geographic locations and thus, they are forced to communicate through technology-mediums (Bell & Kozlowski, 2002; Townsend et al., 1998; Urban et al., 1995). The results of the current study contribute to the existing literature on self-ratings and peer-ratings of performance by examining whether the team architectural structure influences the usefulness of self- and peer-ratings.

Factors Contributing to Peer Performance Ratings

The results of the SRM variance-partitioning indicate that each source of variance accounts of a significant amount of variance in performance ratings or at least meets the level of non-trivial variance suggested by Kenny (1994). The relative amount of variance in performance ratings was compared for the overall contextual performance construct (i.e., the only construct with multiple indicators). Most of the variance in the overall contextual performance ratings was attributable to perceiver (54% in co-located teams/ 39% in virtual teams), followed by the residual (26%/24%), target (13%/26%), and the relationship (7%/11%). This general pattern of findings was the same for co-located and virtual teams, although the magnitude of the variance components differed to some extent. This pattern of findings is inconsistent with previous research, which found that the residual accounted for most of the variance in performance

ratings, followed by the relationship, target, and perceiver effects (e.g., Greguras et al., 2007). The smaller amount of residual variance, in comparison to previous studies, is encouraging because it indicates that factors not modeled in the current study account for a smaller proportion of the variance in performance ratings. However, the smaller target variance and larger perceiver variance is somewhat discouraging. The variance partition results for the target, perceiver, and relationship components are discussed in more detail in the following sections.

Consensus. Although virtual team members do not interact face-to-face and thus, are less likely to have opportunities to monitor each other's performance directly, these factors do not appear to adversely affect the ability of team members to reach consensus in their peer performance ratings. In fact, in the current study, peers demonstrated significant amounts of consensus (i.e., 26-48% target variance) in virtual teams, but a smaller and not statistically significant amount of variance in co-located teams (i.e., 12-18% target variance). On average, target variance was 2.13 times higher in virtual teams than in co-located teams. Thus, there is evidence that in virtual teams, characteristics of the target contribute to peers' ratings of performance and perceivers' agree regarding the targets' level of performance. Consistent with previous research on virtual teams (e.g., Markey & Wells, 2002; Weisband & Atwater, 1999) the ratees' actual behavior is a salient factor in determining peer ratings.

Agreement among raters is important to establish for peer performance ratings because it is necessary for raters to agree, if there is the possibility that the ratings they provide are accurate assessments of the targets' performance (Kenny, 1994). Thus, these results provided evidence in favor of the usefulness of peer ratings in virtual teams, but raise some questions regarding the usefulness of peer-ratings in co-located teams. The average amount target variance in previous SRM studies of co-located teams was 24.5%, but ranged from 13%-42%. Thus, although the

amount of consensus is lower than the amount found in previous studies, these results are not completely inconsistent with previous research. Further, the proportion of variance accounted for in virtual teams is actually much higher than previous estimates. It is also interesting to note that in both co-located and virtual teams, perceivers demonstrated the greatest consensus for workload contribution and job dedication. Thus, there appear to be some dimension-specific differences in consensus, such that perceivers are able to research higher levels of agreement for task-related dimensions (i.e., workload contribution and job dedication) than for more relationship-related dimensions, such as interpersonal facilitation.

Assimilation. Consistent with previous research (e.g., Greguras et al., 2003; Woehr et al., 2005), there was a large amount of perceiver variance (i.e., assimilation) in both co-located (45-76 %) and virtual teams (14-67%; see Table 18). Thus, for both types of teams, there is evidence that individual perceivers tend to see all targets in the same way (Kenny, 1994) and this assimilation is due to characteristics of the perceiver and not the target. However, on average, perceiver variance is 1.88 times higher in co-located teams than in virtual teams. There may be fewer opportunities to observable teammates' performance directly in virtual teams, but team members may feel more comfortable differentiating between teammates in this situation (i.e., they do not feel the same social pressure to rate everyone similarly). The perceiver variance found in the current study for virtual teams and especially for co-located teams is much greater than the perceiver variance found in the four previous studies on perceptions of performance ratings (i.e., 9.8%; range of 8%-30%; Boldry & Kashy, 1999; Greguras et al., 2001; 2007; Sullivan & Reno, 1999).

Back and Kenny (2010) explain that large amounts of perceiver variance indicate the use of different standards by perceivers (i.e., tendencies to see all teammates as high or low

performers). Previous research on 360-degree performance ratings has found that the rating source (e.g., supervisors, peers, subordinates) has a large impact on performance ratings because perceivers have access to different types of information, observe different types of behaviors, and have different goals (e.g., Bozeman, 1997; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Klimoski & London, 1974; Lance et al., 2006; Murphy et al., 2004; Pulakos et al., 1996; Scullen et al., 2000). However, because all raters are at the same level (i.e., peers) it is likely that they shared a common perspective and had the opportunity to observe similar behaviors, thus, the large amount of perceiver variance may represent a reluctance to distinguish between peers (Mount et al., 1998; Scullen et al., 2000). It appears that in co-located teams teammates were particularly reluctant to distinguish between their peers. In sum, the findings regarding perceiver variance in the current study indicate that perceiver characteristics affect performance ratings. These characteristics may take the form of differences in leniency, subjectivity, investment in the rating system, and willingness to differentiate among coworkers (Conway & Huffcutt, 1997; Murphy & Cleveland, 1995).

An examination of the dimensions of performance in the current study indicates that across all dimensions there are high levels of perceiver variance; however, the amount of perceiver variance was lower in virtual teams for the workload dimension. This finding may indicate that in virtual teams, general perceiver tendencies have a relatively smaller impact on evaluations of contribution to the team project than for other extra-role performance dimensions. This type of performance may be more observable and objective and thus, teammates in virtual teams may be more willing to distinguish between team members on this construct. In the current study, the ratings were used for research purposes; however, based on previous research there is

also reason to expect large amounts of perceiver variance when ratings are used for administrative purposes (Van Hooft et al., 2006).

Uniqueness. In addition to characteristics of the perceiver, the idiosyncratic relationships between teammates (i.e., relationship variance; uniqueness) also contributed to peer's rating of performance. In both co-located and virtual teams, the amount of uniqueness is relatively low: 7% in co-locate teams and 11% in virtual teams (see Table 18). Thus, although it does appear that individual relationships have an impact on peer performance ratings in both co-located and virtual teams; this impact appears to be minimal in comparison to the impact of general perceiver tendencies. Unfortunately, in the current study multiple indicators (i.e., interpersonal facilitation and job dedication) were only used for the overall contextual performance, thus, examination of the relationship variance (separate from error variance) is not possible for the all of the performance dimensions (i.e., workload contribution, interpersonal facilitation, and job dedication). Based on the results for the overall performance construct it appears that relationship variance has the smallest impact on performance ratings. Further, the amount of relationship variance is less than found in the three previous studies of performance ratings in co-located teams that were able to separate relationship and effort variance (i.e., 22.7%; ranged from 17%-32%; Boldry & Kashy, 1999; Greguras et al., 2007; Sullivan & Reno, 1999), personality traits (i.e., 20%, Kenny, 1994) and affect (i.e., 40%, Kenny, 1994). Further, this study provides the first evidence regarding the level of relationship variance in peer ratings provided in a virtual environment (Markey & Wells, 2002; Robins et al., 2004). In sum, concerns that individual relationships are going to drastically affect peer performance ratings are minimal. Further, the relative amount of relationship variance in co-located and virtual teams is similar. This finding is somewhat surprising given that co-located teams likely have more opportunity to interact and

develop interpersonal relationships (Driskell et al., 2003) and especially because liking has been found to affect ratings of performance in co-located, but not virtual teams (Weisband & Atwater, 1999).

Reciprocity in Peer Performance Ratings

Reciprocity in organizational research refers to an exchange in which behaviors between partners in a relationship mirror each other (Gouldner, 1960). Examinations of reciprocity using designs other than the SRM have generally failed to support the existence of reciprocity in attraction, perceptions, or behaviors. Kenny et al. (2006) claimed that researchers' inability to find evidence of reciprocity is due to a failure to distinguish between generalized and dyadic reciprocity correlations. The current study supports this contention as there is no evidence of generalized reciprocity (i.e., seeing others as engaging in workload contribution/contextual performance was not related to being seen by others as engaging in workload contribution/contextual performance). These findings are in line with previous research, which has found very limited support for the existence of generalized reciprocity (Kenny, 1994; Greguras et al., 2001; 2007). Kenny et al. (2001) suggested that researchers generally failed to find generalized reciprocity because of the low levels of perceiver variance in previous studies. Kenny et al. also suggested that when substantial perceiver variance exists then there is also likely generalized reciprocity. However, the current study demonstrates that low levels of target variance can also prevent the existence of generalized reciprocity in peer performance ratings.

The existence of dyadic reciprocity indicates the extent to which pairs of teammates see each other as demonstrating similar levels of contextual performance. Kenny (1994; Kenny et al., 2001) proposed that dyadic reciprocity is likely to be strong for prosocial (i.e., helping) behaviors. The only previous study examining the level of dyadic reciprocity in peer

performance ratings found evidence for only the constructs of team orientation and affiliation (Greguras et al., 2007). The level of dyadic reciprocity for contextual performance behaviors is much lower than the amount of dyadic reciprocity found for interpersonal liking (Kenny, 1994) and lower than would be expected given the interpersonal nature of contextual performance behaviors. Further, there is only evidence of dyadic reciprocity in co-located teams. These findings extend previous research, which has only examined dyadic reciprocity in co-located teams by identifying a potential boundary condition for the existence of reciprocity. Specifically, teammates in virtual teams do not appear to reciprocate in their ratings of contextual performance behaviors. These findings are consistent with previous research, which has demonstrated stronger interpersonal relationships in co-located teams than virtual teams (Driskell et al., 2003; Love & Forret, 2008; Straus, 1997; Warkentin et al., 1997; Weisband & Atwater, 1999). These findings also address the scarcity of research on interpersonal processes in virtual teams (Martins, Gilson, & Maynard, 2004) by demonstrating that virtual team members develop unique relationships that influence their performance ratings; however, this influence does not take the form of reciprocity.

Halo in Peer Performance Ratings

The current study provides evidence that both halo bias and true halo contribute to the significant positive relationship between performance dimensions. Thus, there is evidence that perceivers do not differentiate across dimensions of performance (i.e., halo bias) and there is also evidence that there are shared elements among job performance dimensions (i.e., true halo; Bretz et al., 1992). The use of a multivariate SRM analysis allowed for the separation of true similarity in performance across dimensions from similarity in ratings across the performance dimensions due to the perceiver characteristics. The removal of the perceiver, relationship, and error

components from and individuals score provides a purer estimation of the relationship between performance dimensions. Those individuals who contribute to the group project also demonstrate contextual performance and this relationship is stronger when estimated using target-target correlations than correlations between other-ratings. These findings are consistent with previous research by Greguras et al. (2001; 2007) and support the notion of a general performance factor. There is also evidence that perceivers see targets similarly across performance dimensions and this similarity is due to characteristics of the perceiver. Thus, it appears that perceivers develop an overall performance impression (i.e., halo bias). Consistent with Greguras et al. (2007), the significant perceiver-perceiver correlations were smaller in magnitude than the target-target correlations; however, these results still indicate that halo bias may influence peer performance ratings. This study is the first examination of the extent of true halo and halo bias in virtual teams. There are not substantial differences in the magnitude of the target-target correlations (i.e., true halo) and perceiver-perceiver correlations (i.e., halo bias) for co-located and virtual teams. These findings should be replicated given the small amount of target variance in the current study, which is somewhat inconsistent with previous SRM studies examining peer performance.

Self-Other Agreement in Peer Performance Ratings

The use of self-ratings has been suggested as a mechanism for performance management in SMTs because of the lack of regulation by an external supervisor (Elloy, 2004; Manz & Sims, 1987). Self-ratings have also been suggested as a form of performance appraisal in VTs because of decreased opportunity for observation by not only supervisors but also other teammates (Bell & Kozlowski, 2002; Martins et al., 2004). Research on self-ratings has traditionally found low correlations between self- and peer ratings (e.g., meta-analytic r = .19; Conway & Huffcutt,

1997; Harris & Schaubroeck, 1988; Mabe & West, 1982). One reason for the low levels of agreement is the contamination of peer ratings with perceiver, relationship, and random error variance. Self-target agreement correlations are a more precise way to assess self-other agreement because they are not contaminated by these sources of variance. Greguras et al. (2007) concluded that self-target correlations are higher than self-other correlations. Interestingly, the current study did not replicate these results. The self-other correlations ranged from .25 to .89, while the self-target correlations ranged from .11 to .62. The drastically smaller amount of target variance in the current study is likely responsible for these results. The level of self-other agreement based on the average of others ratings were much higher than what has traditionally been found. However, given the relatively small levels of target variance in the ratings it is not clear that this agreement is an indication of accurate performance assessment. Instead, it may be due to the strong perceiver effects.

In addition to the low correlation with other-ratings, self-ratings are often elevated in comparison to other-ratings (Allen et al., 2000; Mabe & West, 1982; Taylor & Brown, 1988). Previous research using the Kwan self-enhancement index (Kwan et al., 2004) found that on average people have self-enhanced concepts of their status in face-to-face groups (Anderson et al., 2006). The self-enhancement index (Kwan et al., 2004) was calculated from SRM components, such that an unbiased self-perception is indicated by zero. The mean self-enhancement index ranged from .13 - .14 in co-located teams and .11 - .24 in virtual teams. The only significant level of self-enhancement was for interpersonal facilitation in co-located teams and for workload contribution in virtual teams. Thus, there appears to be minimal levels of self-enhancement in co-located and virtual teams. These findings are partially consistent with previous research by Weisband and Atwater (1999) as they only found inflation in self-other

ratings in VTs. Although, the self-enhancement index provides researchers with the advantages of removing non-relevant sources of variance, this index is still based on difference scores and thus, may be subject to the limitations of the difference score approach (cf. Fleenor, Smither, Atwater, Braddy, & Sturm, 2010).

Limitations and Future Directions

In the current study, MBA students were recruited to rate the performance of themselves and their teammates with whom they were completing an extended class project. The benefits of this sample were that it allowed for comparisons between co-located and virtual teams and the students (especially in virtual teams) had work experience and were of a similar age to many employees in the workforce. However, there are several limitations of the current sample. First, the sample size for virtual and co-located teams (i.e., 23 teams in each condition) was small for comparisons between the different types of teams. Thus, it is not surprising that the statistical tests were not significant. However, the effect sizes for differences between co-located teams and virtual teams were of moderate size for several of the variance estimates (e.g., perceiver variance for the overall contextual performance measure, relationship variance for the interpersonal facilitation measure, and target variance for the job dedication measure). A larger number of teams per condition are necessary to have enough power to adequately test hypotheses regarding comparisons across co-located and virtual teams. For example, post-hoc power analyses indicated that at least 53 teams in each condition would be necessary to have .80 power when examining the difference in the relationship variance for co-located and virtual teams. Post-hoc power analyses for the other variance components for each performance dimension revealed that even more teams would be need to reach .80 power for comparisons between co-located and virtual teams.

A second limitation of the current sample is that teams completed a project that was part of their class and not an actual work environment. The team member interactions that occur as part of a class project may be different from the interactions of teammates in a work environment. For example, while co-located team members might interact daily in a traditional work environment, co-located student teams might interact on a less frequent schedule. The frequency of interaction was not collected as part of the current study. Examining the frequency of interaction as well as other indicators to more clearly describe the types of interactions between team members is an important direction for future research. These limitations should be considered when evaluating the generalizability of the results from the current study.

In addition to the frequency of interaction, understanding other constraints on the type of interaction is necessary. The current study provides some evidence that communication via technology does not prohibit the usefulness of peer performance ratings; however, differences in the perceiver, target, and relationship variance indicate that different factors may contribute to peers' ratings of performance in co-located and virtual teams. Co-located and virtual team members are distinguished based on the need of virtual team members to communicate via technology (Bell & Kozlowski, 2002; Townsend et al., 1998); however, co-located team members also communicate via technology. Examination of workplace teams that span the range of virtuality may help researchers describe under what conditions peer ratings in co-located teams differ from peer ratings in virtual teams. As previously mentioned, virtual versus co-located teams were identified as one boundary condition for the existence of reciprocity in peer ratings. Tests of significance in the current study were based on the calculation of a variance or correlation for each team and then the mean for that estimate across groups was tested to determine if the value was statistically different from zero. Based on the results there was only

evidence of dyadic reciprocity in co-located teams. However, even across co-located teams there was a large amount of variability in the magnitude of the dyadic reciprocity relationship. Given the variability of dyadic reciprocity correlation from different co-located teams, future research is needed to identify further boundary conditions on when dyadic reciprocity exists.

The cross-sectional design of the current study is a limitation for several reasons. First, with multiple measurements of the performance constructs, researchers could determine if the relative proportion of the perceiver, target, and relationship changes as team members become more acquainted. Relational dynamics within a team change over time, but at this point it is unclear whether increased acquaintance will lead to more appropriate assessment of teammates performance (i.e., larger amounts of target variance), create unique relationships within teams (i.e., larger amounts of relationship variance), or increase raters tendency to rate their team members uniformly (i.e., larger amounts of perceiver variance).

Measurements at multiple points in time could also be used as indicators of an overall performance construct so that relationship variance could be separated from error variance. This approach may provide different variance estimates than those estimates obtained when using different dimensions as indicators of an overall performance construct. Although, the two dimensions of contextual performance are highly related and thus, measurement error was minimal. Multiple measurements of the same dimensions as well as broader expansion of the teamwork behaviors considered would provide a more thorough understanding of peer performance ratings and how those ratings would differ based on the performance dimensions considered. For example, the results of the current study indicate that the magnitude of perceiver and target effects may differ based on whether the performance dimension considered is interpersonal or task-oriented. Future research may consider other teamwork performance

behaviors (e.g., communication, performance monitoring, feedback, leadership, team orientation, adaptability; Dickinson et al., 1992; Dickinson & McIntyre, 1997; Salas et al.,2005) to determine if the relative partitioning of variance differs across dimensions.

In addition to examining other types of teamwork performance, the general inadequate fit of the contextual performance measurement model suggests a need for refinement of the Van Scotter and Motowidlo (1996) contextual performance scale. In particular, configural invariance model demonstrated poor fit for self-ratings from co-located team members and also for comparisons between self-ratings and other-ratings. Thus, refinement is needed to provide researchers with confidence in the use of self-ratings based on this contextual performance scale, in comparisons made between the self-ratings of co-located and virtual team members, and also in comparisons between co-located and virtual team members. In the current study, two items were removed from the job dedication dimension to achieve configural invariance for the self-rating and other-ratings. The differences in these items indicate that targets (i.e., those providing self-ratings) and perceivers (i.e., those providing other-ratings) may have different ideas about the job dedication construct. Future research may explore these differences further.

The large amount of perceiver variance in ratings provided by both co-located and virtual team members is also an indication that there may be room for refinement of the contextual performance measure, as ideally characteristics of the target and not the perceiver would be contributing most to performance ratings. The SRM variance-partitioning approach used in the current study could be used to examine ratings based on a refined scale. This process would allow researchers to determine whether the scale refinement resulted in larger portions of target variance and smaller portions of perceiver variance. In general this SRM

approach could be used to examine attempted refinements to not only the ratings scale, but also other changes in the rating process (e.g., rater training).

In the current study, individual differences in teammate's perceptions of task interdependence were correlated with the SRM target and perceiver components to provide insight into what kinds of factors may be responsible for the ratings provided. There was some support, for the relationship between task interdependence in co-located teams and perceiver effects: When participants felt that their co-located team completed the task in an independent manner they were more likely to rate all of the teammates in the same way. This relationship was not significant for virtual teams, which indicates that task interdependence does not explain perceiver effects in this situation. Further, task interdependence was not related to target effects in either co-located or virtual teams. Thus, the level of interdependence by which individuals completed a task did not influence whether peer ratings were based on the actual behavior of the participant. These findings are interesting because one would think that interdependence would increase the opportunity for observation of teammates' performance and result in a larger portion of the variance in peer ratings being based on characteristics of the target. The examination of other individual difference variables represents and important future direction for research on peer performance ratings. For example, rater goals may predict the magnitude of perceiver and target effects (e.g., maintaining harmony in the group might lead perceivers to rate all targets in the same way).

Conclusion

This study begins to address Latham and Mann's (2006) question regarding whether technology will facilitate or inhibit performance management. The usefulness of peer performance ratings appears to extend to VTs. This study actually found that target

characteristics contribute more to performance ratings and perceiver characteristics contribute less to performance ratings in VTs as compared to co-located teams (although these differences are not statistically significant based on this relatively small sample of teams). The effect size results, which are indicators of practical significance, indicate that the magnitude of the difference between virtual and co-located teams is substantial for many of the comparisons made in the current study. Further, in both types of teams individual relationships appear to contribute to performance ratings. Although the impact of individual relationships appears to be relatively small and reciprocity in peer-ratings only occurred in co-located teams. Further, teammates in both co-located and virtual teams showed relatively low levels of self-enhancement. Given these results there appears to be evidence in support of using both self-ratings and other-ratings of performance. However, there are multiple areas for future research that will be essential to improving our understanding of self- and peer-performance ratings.

References

Albright, L., & Forziati, C. (1995). Cross-situational consistency and perceptual accuracy in leadership. *Personality and Social Psychology Bulletin, 21,* 1269-1276.

Allen, T. D., Barnard, S., Rush, M. C., & Russell, J. E. A. (2000). Ratings of organizational citizenship behavior: Does the source make a difference? *Human Resource Management Review, 10,* 97-114.

Anderson, C., Srivastava, S., Beer, J. S., Spataro, S. E., & Chatman, J. A. (2006). Knowing your place: Self-perceptions of status in face-to-face groups. *Journal of Personality and Social Psychology, 91,* 1094-1110.

Arthur, W., Edwards, B., Bell, S., Villado, A., & Bennett, W., Jr. (2005). Team task analysis: Identifying tasks and jobs that are team based. *Human Factors, 47,* 654-669.

Ashford, S. (1989). Self-assessments in organizations: A literature review and integrative model. *Research in Organizational Behavior, 11,* 133-174.

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology, 51,* 577-598.

Atwater, L., & Yammarino, F. (1997). Self-other rating agreement: A review and model. *Research in personnel and human resources management, Vol. 15* (pp. 121-174). US: Elsevier Science/JAI Press.

Back, M. D., & Kenny, D. A. (2010). The social relations model: How to understand dyadic processes. *Social and Personality Psychology Compass, 4,* 855-870.

Balthazard, P. A., Waldman, D. A., & Warren, J. E. (2009). Predictors of emergence of transformational leadership in virtual decision teams. *Leadership Quarterly, 20,* 651-663.

Banks, C. G., & May, K. E. (1999). Performance management: The real glue in organizations. In A. Kraut & A. K. Korman (Eds.), *Evolving practices in human resource management: Responses to a changing world of work.* (pp. 90-145). San Francisco: Jossey-Bass.

Barnes-Farrell, J. (2001). Performance appraisal: Person perception processes and challenges. *How people evaluate others in organizations* (pp. 135-153). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Barrett, R. (1966). *Performance rating*. Oxford England: Science Research Assoc., Inc.

Batt, B. (2004). Who benefits from teams? Comparing workers, supervisors, and managers. *Industrial Relations, 43,* 183-212.

Beehr, T. A., Ivanitskaya, L., Hansen, C. P., Erofeev, D., & Gudanowski, D. M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior, 22,* 775-788.

Bell, B., & Kozlowski, S. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27,* 14-49.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Bettenhausen, K., & Fedor, D. (1997). Peer and upward appraisals: A comparison of their benefits and problems. *Group & Organization Management, 22,* 236-263.

Bishop, J. W., & Scott, D. K. (2000). An examination of organizational and team commitment in a self-directed team environment. *Journal of Applied Psychology, 85,* 439-50.

Blasi, J. R., & Kruse, D. L. (2006). U.S. high-performance work practices at century's end. *Industrial Relations, 45,* 547-578.

Boldry, J. G., & Kashy, D. A. (1999). Intergroup perception in naturally occurring groups of differential status: A social relations perspective. *Journal of Personality and Social Psychology, 77,* 1200–1212.

Bond, C. F., Jr., & Lashley, B. R. (1996). Round-robin analysis of social interaction: Exact and estimated standard errors. *Psychometrika, 61,* 303-311.

Bonito, J. A. (2003). A social relations analysis of participation in small groups. *Communication Monographs, 70,* 83-97.

Borman, W. (2004). The concept of organizational citizenship. *Current Directions in Psychological Science, 13,* 238-241.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W.C. Borman, & Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass.

Borman, W. C, & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10,* 99-109.

Borman, W., C., Penner, L., Allen, T., & Motowidlo, S. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment, 9,* 52-69.

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80,* 168-177.

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Wainer (Eds.), *Modeling Contextual Effects in Longitudinal Studies* (pp. 151-182). Mahwah, NJ: Lawrence Erlbaum Associates.

Bozeman, D. P. (1997). Interrater agreement in mulit-source performance appraisal: A commentary. *Journal of Organizational Behavior, 18,* 313-316.

Bretz, R., Milkovich, G., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management, 18,* 321-352.

Brutus, S., & Derayeh, M. (2002). Multisource assessment programs in organizations: An insider's perspective. *Human Resource Development Quarterly, 13,* 187-202.

Campbell, L., Simpson, J. A., Stewart, M., & Manning, J. (2003). Putting personality in social context: Extraversion, emergent leadership, and the availability of rewards. *Personality and Social Psychology Bulletin, 29,* 1547-1559.

Cannon-Bowers, J. A., & Salas, E. (1997). A framework for developing team performance measures in training. In M. T. Brannick, E. Salas, & C. W. Prince (Eds.) *Team performance, assessment, and measurement: Theory, methods, and applications* (pp. 45-62). Mahwah, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco: Jossey-Bass.

Carless, S., & Roberts-Thompson, G. (2001). Self-ratings in training programs: An examination of level of performance and the effects of feedback. *International Journal of Selection and Assessment, 9,* 217-225.

Carson, J., Tesluk, P., & Marrone, J. (2007). Shared leadership in teams: An investigation of antecedent conditions and performance. *Academy of Management Journal, 50,* 1217-1234.

Carte, T., Chidambaram, L., & Becker, A. (2006). Emergent leadership in self-managed virtual teams: A longitudinal study of concentrated and shared leadership behaviors. *Group Decision and Negotiation, 15,* 323-343.

Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. *Personnel Psychology, 33,* 567-578.

Cheung, G. W. (1999). Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology, 52,* 1-36.

Church, A. H. (1997). Managerial self-awareness in high-performing individuals in organizations. *Journal of Applied Psychology, 82,* 281-292.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Lawrence Erlbaum Associates.

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness from the shop-floor to the executive suite. *Journal of Management, 23,* 239-260.

Coleman, V., & Borman, W. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resource Management Review, 10,* 25-44.

Conger, J. A., & Pearce, C. L. (2003). A landscape of opportunities: Future research on shared leadership. In C. L. Pearce & J.A. Conger (Eds.), *Shared leadership: Reframing the hows and whys of leadership* (pp. 285–301). Thousand Oaks, CA: Sage.

Conway, J. M. (1998). Understanding method variance in multitrait–multirater performance appraisal matrices: Examples using general impressions and interpersonal affect as measured method factors. *Human Performance, 11*, 29-55.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84*, 3-13.

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331-360.

Conway, J. M., Lombardo, K., & Sanders, K. C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance, 14*, 267-303.

Coyle-Shapiro, J. (2002). A psychological contract perspective on organizational citizenship behavior. *Journal of Organizational Behavior, 23*, 927-946.

Cramton, C., Orvis, K., & Wilson, J. (2007). Situation invisibility and attribution in distributed collaborations. *Journal of Management, 33*, 525-546.

Cronbach, L. J., Gleser, G. C., Nada, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

De Dreu, C. K. W., & Van Vianen, A. E. M. (2001). Managing relationship conflict and the effectiveness of organizational teams. *Journal of Organizational Behavior, 22*, 309-328.

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. *Team performance assessment and measurement: Theory, methods, and applications* (pp. 19-43). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Dickinson, T. L., McIntyre, R. M., Ruggeberg, B. J., Yanushefski, A., Hamil, L. S., & Vick, A. L. (1992). *A conceptual framework for developing team process measures of decision-making performance*. Orlando, FL: Naval Training Systems Center, Human Factors Division.

Doll, R. E., & Longo, A. A. (1962). Improving the predictive effectiveness of peer ratings. *Personnel Psychology*, *15*, 215-220.

Drexler, J. A., Jr, Beehr, T. A., & Stetz, T. A. (2001). Peer appraisals: Differentiation of individual performance on group task. *Human Resource Management, 40*, 333–345.

Driskell, J., E., Radtke, P., H., & Salas, E. (2003). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics: Theory, Research, and Practice, 7*, 297-323.

Druskat, V., & Wolff, S. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, *84*, 58-74.

Elloy, D. F. (2004). The influence of super leader behaviors on organization commitment, job satisfaction and organization self-esteem in a self-managed work team. *Leadership and Organizational Development Journal, 26*, 120-127.

Erez, A., Lepine, J., & Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology, 55*, 929-948.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology, 86,* 215-227.

Fedor, D. B., Bettenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees' dual roles as raters and recipients. *Group and Organization Management, 24,* 92 -120.

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly, 21,* 1005-1034.

Fletcher, C. (2001). Performance appraisal and management: The developing research agenda. *Journal of Occupational and Organizational Psychology, 74,* 473-487.

Fletcher, T., & Major, D. (2006). The effects of communication modality on performance and self-ratings of teamwork components. *Journal of Computer-Mediated Communication, 11*, 557-576.

Gibson, C. B., & Kirkman, B. L. (1999). Our past, present, and future in teams: The role of human resource professionals in managing team performance. In A. Kraut & A. K. Korman (Eds.), *Evolving practices in human resource management: Responses to a changing world of work.* (pp. 90-117). San Francisco: Jossey-Bass.

George, J. M., & Brief, A. P. (1992). Feeling good-doing good: A conceptual analysis of the mood at work-organizational spontaneity relationship. *Psychological Bulletin, 112*, 310-329.

Gouldner, A.W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25,* 161–178.

Greguras, G. J., & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83*, 960–968.

Greguras, G., Robie, C., & Born, M. P. (2001). Applying the social relations model to self and peer evaluations. *Journal of Management Development, 20*, 508-525.

Greguras, G. J., Robie, C., Born, M. P., & Koenigs, R. J. (2007). A social relations analysis of team performance ratings. *International Journal of Selection and Assessment, 15,* 434-448.

Greguras, G. J., Robie, C., Schleicher, D. J., & Goff, M. III (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology, 56,* 1–21.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47,* 307-38.

Hanson, M., & Borman, W. (2006). Citizenship performance: An integrative review and motivational analysis. *Performance measurement: Current perspectives and future challenges* (pp. 141-173). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41,* 43-62.

Harris, T., & Barnes-Farrell, J. (1997). Components of teamwork: Impact on evaluations of contributions to work team effectiveness. *Journal of Applied Social Psychology, 27,* 1694-1715.

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

Hui, C., Lam, S., & Law, K. (2000). Instrumental values of organizational citizenship behavior for promotion: A field quasi-experiment. *Journal of Applied Psychology, 85,* 822-828.

Ingraham, L., & Wright, T. (1986). A cautionary note on the interpretation of relationship effects in the social relations model. *Social Psychology Quarterly, 49,* 93-97.

Janz, B. D., Colquitt, J. A., & Noe, R. A. (1997). Knowledge worker team effectiveness: The role of autonomy, interdependence, team development, and contextual support variables. *Personnel Psychology, 50*, 877-904.

Jones. C. (2004). *Teleworking:* The *quiet revolution.* Report published by The Gartner Group.

Kanawattanachai, P., & Yoo, Y. (2002). Dynamic nature of trust in virtual teams. *The Journal of Strategic Information Systems, 11,* 187-213.

Kane, J., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin, 85,* 555-586.

Kane, J., & Lawler, E. E. (1979). Performance appraisal effectiveness. In B. Staw (Eds.), *Research in organizational behavior*: Vol. 1 (pp. 425-478). Greenwich, Conn.: JAI Press.

Kashy, D., & Kenny, D. (2000). The analysis of data from dyads and groups. *Handbook of research methods in social and personality psychology* (pp. 451-477). New York, NY US: Cambridge University Press.

Kasten, R., & Nevo, B. (2008). Exploring the relationship between interrater correlations and validity of peer ratings. *Human Performance, 21*, 180-197.

Kendall, D. L., & Salas, E. (2004). Measuring team performance: Review of current methods and consideration of future needs. *Advances in Human Performance and Cognitive Engineering Research*, 5, 311-331.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A. (1998). SOREMO [computer program]. Available: http://nw3.nai.net/~dakenny/kenny.htm.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102*, 390-402.

Kenny, D., A., Horner, C., Kashy, D., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88-97.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99*, 422-431.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic analysis.* New York: Guilford Press.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142-182). Orlando, FL: Academic Press.

Kenny, D. A., Mohr, C. D., & Levesque, M. J. (2001). A social relations variance partitioning of dyadic behavior. *Psychological Bulletin, 127,* 128-141.

Kenny, D. A., West, T. V., Malloy, T. E., & Albright, L. (2006). Componential analysis of interpersonal perception data. *Personality and Social Psychology Review, 10*, 282-294.

Kenny, D. A., & Zaccaro, S. J. (1983). An estimate of variance due to traits in leadership. *Journal of Applied Psychology, 68*, 678-685.

Klein, K., T., & Kozlowski, S W (Eds.). (2000) *Multi-level theory, research, and methods, in organizations.* San Francisco: Jossey-Bass Publishers.

Klimoski, R. J., & Donahue, L. M. (2001). Person perception in organizations: An Overview of the field. *How people evaluate others in organizations* (pp. 5-43). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology, 59*, 445-451.

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

Korsgaard, M., Meglino, B., Lester, S., & Jeong, S. (2010). Paying you back or paying me forward: Understanding rewarded and unrewarded organizational citizenship behavior. *Journal of Applied Psychology, 95*, 277-290.

Kozlowski, S. W. J., & Bell, B. (2003). Work groups and teams in organizations. *Handbook of psychology: Industrial and organizational psychology*, Vol. 12 (pp. 333-375). Hoboken, NJ US: John Wiley & Sons Inc.

Kozlowski, S. W. J., Gully, S. M., McHugh, P. P., Salas, E., & Cannon-Bowers, J. A. (1996). A dynamic theory of leadership and team effectiveness: Developmental and task contingent leader roles. In G. R. Ferris (Ed.), *Research in personnel and human resource management* (Vol. 14, pp. 253-305). Greenwich, CT: JAI.

Kwan, V., John, O., Kenny, D., Bond, M., & Robins, R. (2004). Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychological Review, 111*, 94-110.

Lance, C., Baxter, D., & Mahan, R. (2006). Evaluation of alternative perspectives on source effects in multisource performance measures. *Performance measurement: Current perspectives and future challenges* (pp. 49-76). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Lance, C. E., & Bennett, W., Jr. (1997, April). *Rater source differences in cognitive representations of performance dimensions.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Landy, F. J. & Farr, J. L. (1983). *The measurement of work performance*. New York: Academic Press.

Lashley, B. R., & Kenny, D. A. (1998). Power estimation in social relations analyses. *Psychological Methods, 3*, 328-338.

Latham, G., & Mann, S. (2006). Advances in the Science of Performance Appraisal: Implications for Practice. *International review of industrial and organizational psychology* (Vol 21) (pp. 295-337). Hoboken, NJ US: Wiley Publishing.

Lawler, E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51*, 369-381.

Lawler, E., III, Mohrman, S.A., & Ledford, G.E. Jr. (1994). *Employee Involvement and Total Quality Management: Practices and Results in Fortune 1000 Companies*, Jossey-Bass, San Francisco, CA.

Lawler, E. E., III, Mohrman, S. A., & Ledford, G. E., Jr. (1995). *Creating high performance organizations: Practices and results of employee involvement and total quality management in Fortune 1000 companies.* San Francisco: Jossey-Bass.

Levy, P., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management, 30*, 881- 905.

Livi, S., Kenny, D. A., Albright, L., Pierro, A. (2008). A social relations analysis of leadership. *The Leadership Quarterly, 19*, 235-248.

London M., & Smither, J. W. (1995). Can multisource feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology, 48*, 803–839.

London, M., & Tornow, W.W. (1998). 360-degree feedback—More than a tool! In W. Tornow & M. London (Eds.), *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (pp. 1–8). San Francisco: Jossey-Bass.

Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. *Journal of Applied Psychology, 66,* 451-457.

Love, M., & Forret, M. (2008). Exchange relationships at work: An examination of the relationship between team-member exchange and supervisor reports of organizational citizenship behavior. *Journal of Leadership & Organizational Studies, 14*, 342-352.

Mabe, P. A., III., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280-296.

MacDuffie, J. (2008). HRM and distributed work: Managing people across distances. *The Academy of Management annals* (Vol. 1) (pp. 549-615). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

MacKenzie, S., Podsakoff, P., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons' performance. *Organizational Behavior and Human Decision Processes*, *50,* 123-150.

Makiney, J. D., & Levy, P. E. (2002). The influence of self-ratings versus peer ratings on supervisors' performance judgments. *Organizational Behavior and Human Decision Processes, 74*, 212-228.

Malloy, T., Agatstein, F., Yarlas, A., & Albright, L. (1997). Effects of communication, information overlap, and behavioral consistency on consensus in social perception. *Journal of Personality and Social Psychology, 73*, 270-280.

Malloy, T. E., & Janowski, C. (1992). Perceptions and meta-perceptions of leadership: Components, accuracy and dispositional correlates. *Personality and Social Psychology Bulletin, 18,* 700-708.

Malloy, T., & Kenny, D. (1986). The Social Relations Model: An integrative method for personality research. *Journal of Personality, 54,* 199-225.

Manz, C., & Sims, H. P. (1987). Leading workers to lead themselves: The external leadership of self-managing work teams. *Administrative Science Quarterly, 32,* 106–129.

Manz, C., & Sims, H. P. (1986). Leading self-managed groups: a conceptual analysis of a paradox. *Economic and Industrial Democracy, 7,* 141-65.

Marcus, D. K. (1998). Studying group dynamics with the social relations model. *Group Dynamics: Theory, Research, and Practice, 2,* 230-240.

Markey, P. M., & Wells, S. M. (2002). Interpersonal perception in internet chat rooms. *Journal of Research in Personality, 36,* 134-136.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26,* 356-376.

Martins, L. L., Gilson, L. L., & Maynard, M. T. (2004). Virtual teams: What do we know and where do we go from here? *Journal of Management, 30,* 805-835.

Mathieu, J., Maynard, M., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34,* 410-476.

Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83,* 693–702.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 149-203). San Francisco, CA: Jossey-Bass.

McIntyre, R. M., & Dickinson, T. L. (1992). Systemic assessment of teamwork processes in tactical environments (N61339-91-C-0145). Norfolk, VA: Old Dominion University.

McNeely, B. L., & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology, 79,* 836-844.

Miller, L., & Kenny, D. (1986). Reciprocity of self-disclosure at the individual and dyadic levels: A social relations analysis. *Journal of Personality and Social Psychology, 50,* 713-719.

Morgan, B., Salas, E., & Glickman, A. (1993). An analysis of team evolution and maturation. *Journal of General Psychology, 120,* 277-291.

Motowidlo, S. (2000). Some Basic Issues Related to Contextual Performance and Organizational Citizenship Behavior in Human Resource Management. *Human Resource Management Review, 10,* 115-126.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79,* 475-480.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology, 51,* 557-575.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89,* 158-164.

Murphy, K. R., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53,* 873-900.

Muthén, L. K. & Muthén, B. O. (1998-2010). Mplus user's guide: Sixth edition. Los Angeles, CA: Muthén, & Muthén.

Nathan, B., Mohrman, A., & Milliman, J. (1991). Interpersonal relations as a context for the effects of appraisal interviews on performance and satisfaction: A longitudinal study. *Academy of Management Journal, 34,* 352-369.

Norton, S. M. (1992). Peer assessments of performance and ability: An exploratory meta-analysis of statistical artifacts and contextual moderators. *Journal of Business and Psychology, 6,* 387-399.

Organ, D. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10,* 85-97.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome.* Lexington, MA: Lexington Books.

Organ, D. W. (1990). The motivational basis of organizational citizenship behavior. In B. Staw, & L. Cummings (Eds.), *Research in Organizational Behavior*, (Vol. 12, pp. 43-72).

Organ, D. W., & Konovsky, M. (1989). Cognitive versus affective determinants of organizational citizenship behavior. *Journal of Applied Psychology, 74,* 157-164.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology, 48,* 775-802.

Osterman, P. (1994). How common is workplace transformation and who adopts it? *Industrial and Labor Relations Review, 47,* 173-188.

Paris, C., Salas, E., & Cannon-Bowers, J. (2000). Teamwork in multi-person systems: A review and analysis. *Ergonomics, 43,* 1052-1075.

Park, B., & Flink, C. (1989). A social relations analysis of agreement in liking judgments. *Journal of Personality and Social Psychology, 56,* 506-518.

Park, B., & Judd, C. (1989). Agreement on initial impressions: Differences due to perceivers, trait dimensions, and target behaviors. *Journal of Personality and Social Psychology, 56,* 493-505.

Pearce, C. L., Yoo, Y., & Alavi, M. (2004). Leadership, social work, and virtual teams: The relative influence of vertical versus shared leadership in the nonprofit sector. In R. Riggio & S. Smitt-Orr (Eds.), *Improving leadership in nonprofit organizations* (pp. 180–203). San Francisco: Jossey-Bass.

Peters, S., Kinsey, P., & Malloy, T. E. (2004). Gender and leadership perceptions among African Americans. *Basic and Applied Social Psychology, 26*, 93-101.

Podsakoff, P. M., & MacKenzie, S. B. (1997). The impact of organizational citizenship behavior on organizational performance: A review and suggestions for future research. *Human Performance, 10,* 133-152.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management, 26*, 513–563.

Porter, C. O. L. H., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P. J., West, B. J., & Moon, H. (2003). Backing up behaviors in teams: The role of personality and legitimacy of need. *Journal of Applied Psychology, 88*, 391-403.

Potter, R., & Balthazard, P. (2002). Virtual team interaction styles: Assessment and effects. *International Journal of Human-Computer Studies, 56,* 423-443.

Pulakos, E. D., Schmitt, N., & Chan, D. (1996). Models of job performance ratings: An examination of ratee race, ratee gender, and rater level effects. *Human Performance, 9,* 103–120.

Reilly, R. R., & Aronson, Z. H. (2009). Managing contextual performance. In J. W. Smither, M. London, J. W. Smither, M. London (Eds.), *Performance management: Putting research into action* (pp. 297-328). San Francisco, CA US: Jossey-Bass.

Reilly, R. R., & McGourty, J. (1998). Performance appraisal in team settings. In Smither J. W. (Ed.), *Performance appraisal: State of the art in practice* (pp. 244–277). San Francisco: Jossey-Bass.

Robins, R. W., Mendelsohn, G. A., Connell, J. B., & Kwan, V. S. Y. (2004). Do people agree about the causes of behavior? A social relations analysis of behavior ratings and causal attributions. *Journal of Personality and Social Psychology, 86,* 334-344.

Rosenstein, R. (1994). *The Teamwork Components Model: An Analysis Using Structural Equations Modeling. Unpublished doctoral dissertation,* Old Dominion University, Norfolk, VA.

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78,* 450–462.

Salas, E., Burke, C., & Fowlkes, J. E. (2006). Measuring Team Performance 'In The Wild': Challenges and Tips. In W. Jr. Bennett, C. E. Lance, D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges* (pp. 245-272). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Salas, E., Dickinson, T. Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. J. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there "big five" in teamwork? *Small Group Research, 36*, 555-599.

Salas, E., Stagl. K., & Burke, C S. (2004). 25 years of team effectiveness in organizations: Research themes and emerging needs. In C L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 19. pp. 47-91). New York: Wiley.

Salas, E., Weaver, S. J., Rosen, M. A., & Smith-Jentsch, K. A. (2009). Managing team performance in complex settings: Research-based best practices. In J. W. Smither, M. London, J. W. Smither, M. London (Eds.), *Performance management: Putting research into action* (pp. 197-232). San Francisco, CA US: Jossey-Bass.

Schmidt, F., & Hunter, J. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*, 199-223.

Scullen, S. E, Mount, M. K., &Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956–970.

Shore, T. H., Shore, L. M., & Thorton, G. C. (1992). Construct validity of self- and peer evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology, 77*, 42-54.

Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology, 68*, 653-663.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individuals and team training* (pp. 61-87). Washington, DC: American Psychological Association.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Stewart, G. L., & Barrick, M. R. (2000). Team structure and performance: Assessing the mediating role of intrateam process and the moderating role of task type. *Academy of Management Journal, 43*, 135-148.

Straus, S. G. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Human-Computer Interaction, 12*, 227–266.

Sullivan, M. P., & Reno, R. R. (1999). Perceiving Groups Accurately. *Group Dynamics: Theory, Research, and Practice, 3*, 1–10.

Taylor, S., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193-210.

Tesluk, P., Mathieu, J. E., Zaccaro, J. E., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and Measurement: Theory, Methods, and Applications.* Mahwah, NJ: Erlbaum.

Thompson, J. D. (1967). *Organizations in Action*, McGraw-Hill, New York, NY.

Tippins, N. T., & Coverdale, S. H. (2009). Performance management of the future. In J. W. Smither, M. London, J. W. Smither, M. London (Eds.), *Performance management: Putting research into action* (pp. 555-583). San Francisco, CA US: Jossey-Bass.

Toegel, G., & Conger, J. A. (2003). 360-degree assessment: Time for reinvention. *Academy of Management Learning and Education, 2*, 297–311.

Townsend, A. M., DeMarie, S. M., & Hendrickson, A. R. (1998). Virtual teams: technology and the workplace of the future. *Academy of Management Executive, 12,* 17-29.

Urban, J., Bowers, C., Cannon-Bowers, J., & Salas, E. (1995). The importance of team architecture in understanding team processes. Advances in interdisciplinary studies of work teams: Knowledge work in teams, Vol. 2 (pp. 205-228). US: Elsevier Science/JAI Press.

Van de Ven, A. H., Delbecq, A. L., & Koenig, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review, 41*, 322-328.

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5,* 139-158.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4−69.

Van der Vegt, G., & Van de Vliert, E. (2002). Intragroup interdependence and effectiveness: Review and proposed directions for theory and practice. *Journal of Managerial Psychology, 17,* 50-67.

Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: evidence of construct and predictive validity. *Academy of Management Journal, 41,* 108-119.

Van Hooft, E. A. J., Van der Flier, H., & Minne, M. R. (2006). Construct validity of multi-source performance ratings: An examination of the relationship of self-, supervisor-, and peer-ratings with cognitive and personality measures. *International Journal of Selection and Assessment, 14,* 67-81.

Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology, 81,* 525-531.

Vance, R., Winne, P., & Wright, E. (1983). A longitudinal examination of rater and ratee effects in performance ratings. *Personnel Psychology, 36*, 609-620.

Vargas, P. R. (1996). *A social relations analysis of leadership perceptions among athletes. Unpublished master's thesis,* Sam Houston State University, Huntsville.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81,* 557-574.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87,* 345-354.

Viswesvaran, C., Schmidt, F., & Ones, D. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90,* 108-131.

Warkentin, M. E., Sayeed, L., & Hightower, R. (1997). Virtual teams versus face-to-face teams: An exploratory study of a web-based conference system. *Decision Science, 28,* 975–996.

Warner, R., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology, 37*, 1742-1757.

Weisband, S., & Atwater, L. (1999). Evaluating self and others in electronic and face-to-face groups. *Journal of Applied Psychology, 84,* 632–639.

Weldon, E., Jehn, K. A., & Pradhan, P. (1991). Processes that mediate the relationship between a group goal and improved group performance. *Journal of Personality and Social Psychology, 61,* 555-569.

Werner, J. M. (1994). Dimensions that make a difference: Examining the impact of in-role and extra-role behaviors on supervisory ratings. *Journal of Applied Psychology, 79,* 98-107.

Wherry, R. J., Sr, & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology, 35,* 521–551.

Woehr, D. J., Sheehan, M. K., & Bennett, W., Jr. (2005). Assessing measurement equivalence across rating sources: A multitrait–multirater approach. *Journal of Applied Psychology, 90,* 592–600.

Wohlers, A., & London, M. (1989). Ratings of managerial characteristics: Evaluation difficulty, co-worker agreement, and self-awareness. *Personnel Psychology, 42,* 235-261.

Yao, X., & Wang, L. (2008). Socially oriented values and reciprocity norm predict organizational commitment. *Asian Journal of Social Psychology, 11,* 247-252.

Yoo, Y., Kanawattanachai, P., & Citurs, A. (2002). Forging into the wired wilderness: A case study of a technology-mediated distributed discussion-based class. *Journal of Management Education, 26, 139-163.*

Table 1. Components of the Social Relations Model for Study Variables

Round-robin ratings were provided by team members (i.e., each team member rated every other team member on items measuring contribution to the team project and contextual performance). Using the SRM analysis each rating is partitioned into five components: constant (group mean), perceiver effect, target effect, relationship effect, and error. The description of each SRM component provided below is for team member A's rating of team member B. The same interpretation would be made for any combination of team members.

Component	Description
Constant	The average level within a group at which perceivers view targets' contextual performance (or work contribution)
Perceiver effect	The extent to which A sees team members (targets) as exhibiting high or low levels of contextual performance (or work contribution)
Target effect	The extent to which B is seen a exhibiting high or low levels of contextual performance (or work contribution) by the other team members
Relationship effect	The degree to which A sees B as especially high or low on contextual performance (or work contribution)
Error	Chance, inconsistent, or unstable aspects of the rating process

Note: The relationship effect cannot be separated from measurement error for the work contribution or either dimension of contextual performance because only a single indicator was used in the SRM analyses. The perceiver effect is called the actor effect when analyzing actual as opposed to perceived behavior and the target effect is called the partner effect (Kenny, et al., 2006).

Table 2. Issues Investigated by examining SRM Variances and Intercorrelations

Source	Question	Term
Variance components		
Target variance	How much do targets differ in how their performance is generally seen by others?	Consensus
Perceiver variance	How much do perceivers differ in how they generally see their teammates' performance?	Assimilation
Relationship variance	How much do individual raters differ in their unique perceptions of their specific teammates' performance?	Uniqueness
SRM intercorrelations		
Relationship-relationship correlation	Do pairs of teammates see each other's performance uniquely?	Dyadic reciprocity
Target-target correlation	Do targets perform similarly across performance dimensions?	True halo
Perceiver-perceiver correlation	Do perceivers see targets the same regardless of the performance dimensions?	Halo bias
Self-rating-target correlation	Do targets see themselves as others see them?	Self-other agreement

Note: SRM, Social Relations Model. This table is similar to the table included by Greguras et al. (2007) and is a subset of a table originally presented in Kenny (1994). The terms *true halo* and *halo bias* were added for clarity.

Table 3. Summary of SRM Studies on Peer Performance Ratings

Study	Rating Purpose	Sample	Percentage of variance accounted			
			Target effect	Perceiver effect	Relationship effect	Residual effect
Sullivan and Reno (1999)*	research	students	42%	17%	32%	9%
Boldry and Kashy (1999)	research	students	17%	14%	19%	51%
Greguras, Robie, and Born (2001) *	administrative	students	26%	30%	44%	
Greguras, Robie, Born, and Koenigs (2007)*	developmental	applied	13%	8%	17%	63%
Average of all four studies			24.5%	9.8%	22.7%	
Average of three studies using student samples			28.3%	20.3%	25.5%	

Note: *Studies that collected both self and peer evaluations.

Table 4. Summary of Hypotheses

Term		Hypotheses
Consensus	Hypothesis 1a	Perceivers will demonstrate consensus in their ratings of peer performance in co-located teams (i.e., target variance ($s_b^2{}_{\text{co-located}}$) will be statistically significant).
	Hypothesis 1b	Perceivers will demonstrate consensus in their ratings of peer performance in VTs (i.e., target variance ($s_b^2{}_{\text{virtual}}$) will be statistically significant).
	Hypothesis 1c	Consensus will be higher in co-located than in VTs (i.e., $s_b^2{}_{\text{co-located}} > s_b^2{}_{\text{virtual}}$).
Assimilation	Hypothesis 2a	Perceivers will demonstrate assimilation in their ratings of peer performance in co-located teams (perceiver variance ($s_a^2{}_{\text{co-located}}$) will be statistically significant).
	Hypothesis 2b	Perceivers will demonstrate assimilation in their ratings of peer performance in VTs (perceiver variance ($s_a^2{}_{\text{virtual}}$) will be statistically significant).
	Hypothesis 2c	Assimilation will be higher in virtual than in co-located teams ($s_a^2{}_{\text{virtual}} > s_a^2{}_{\text{co-located}}$).

Table 4 (continued).

Term		Hypotheses
Uniqueness	Hypothesis 3a	In co-located teams, perceivers will see targets in idiosyncratic ways (i.e., demonstrate uniqueness; relationship variance ($s_g^2{}_{\text{co-located}}$) will be statistically significant).
	Hypothesis 3b	In virtual teams, perceivers will see targets in idiosyncratic ways (i.e., demonstrate uniqueness; $s_g^2{}_{\text{virtual}}$ will be statistically significant).
	Hypothesis 3c	The amount of uniqueness will be larger in co-located teams than in VTs (i.e., $s_g^2{}_{\text{co-located}} > s_g^2{}_{\text{virtual}}$).
Relative magnitude of variance components	Hypothesis 4a	In co-located teams, raters will demonstrate similar amounts of consensus and uniqueness and smaller amounts of assimilation (i.e., $s_b^2{}_{\text{co-located}} \approx s_g^2{}_{\text{co-located}} > s_a^2{}_{\text{co-located}}$).
	Hypothesis 4b	In VTs, raters will demonstrate similar amounts of consensus and assimilation and smaller amount of uniqueness (i.e., $s_b^2{}_{\text{virtual}} \approx s_a^2{}_{\text{virtual}} > s_g^2{}_{\text{virtual}}$).
Reciprocity	Hypothesis 5a	Dyadic reciprocity will exist for ratings of contextual performance in co-located teams (i.e., $r_{gg'\text{ co-located}}$ will be statistically significant).
	Hypothesis 5b	Dyadic reciprocity will exist for ratings of contextual performance in VTs (i.e., $r_{gg'\text{ virtual}}$ will be statistically significant).

Table 4 (continued).

Term		Hypotheses
Reciprocity (continued)	Hypothesis 5c	Higher levels of dyadic reciprocity will exist for ratings of contextual performance in co-located teams than in virtual teams (i.e., $r_{gg'\ co\text{-}located} > r_{gg'\ virtual}$).
True halo	Hypothesis 6	Individuals in co-located teams and virtual teams who contribute to the group project will also demonstrate contextual performance behaviors (i.e., the target-target correlations ($r_{wl\ cp}$) will be statistically significant).
Halo bias	Hypothesis 7	Perceivers in co-located teams and virtual teams will have a general tendency to see targets similarly for the performance dimensions of contributing to the team task and contextual performance (i.e., the perceiver-perceiver correlations ($r_{wl\ cp}$) will be statistically significant).
Self-other agreement	Hypothesis 8a	In co-located and virtual teams, participants will display self-other agreement (i.e., the correlations between self-ratings and the SRM target effect ($r_{ob\ co\text{-}located}$ and $r_{ob\ virtual}$) will be statistically significant).
	Hypothesis 8b	Self-target agreement is likely to be greater in co-located teams than in virtual teams (i.e., $r_{ob\ co\text{-}located} > r_{ob\ virtual}$).
	Hypothesis 8c	Self-enhancement is likely to be greater in virtual than in co-located teams (i.e., the magnitude of the self-enhancement index will be larger in virtual teams than in co-located teams; $SE_{virtual} > SE_{co\text{-}located}$).

Table 5. Sample Size for Original and Final Sample

		Co-located Teams		Virtual Teams	
Data Collection	*n*	*n*	*g*	*n*	*g*
Original Sample					
2007	113	68	17	45	12
2008	102	60	16	42	11
2009	84	51	15	33	9
2010	76	39	11	37	10
Total	375	218	59	157	42
Final Sample					
2007	95	58	15	37	10
2008	77	41	12	36	10
2009	56	31	9	25	7
2010	63	30	9	33	9
Total	291	160	45	131	36

Note: *n* = number of participants; *g* = number of groups.

Table 6. Estimated Group Size Required to Detected Significant Effects at Power of .80

	Number of groups needed	
	group-level t-test	z-score normal test
Perceiver	25	23
Target	11	10
Dyadic covariance	17	14

Note: Power estimates were based on conducting a one-sample group t-test using the SRM components.

Table 7. Number of 3-person and 4/5-person Co-located and Virtual Teams

Data Collection	Co-located Teams		Virtual Teams	
	3-person team	4/5-person team	3-person team	4/5-person team
2007	3	7	4	11
2008	4	6	7	5
2009	3	4	5	4
2010	3	6	4	3
Total	13	23	22	23
	36		45	

Table 8. Demographic Information for Co-located and Virtual Teams

	Co-located Teams						Virtual Teams					
	3-person		4/5-person		All		3-person		4/5-person		All	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender												
Male	24	61.5	54	60.7	78	60.9	48	73.8	69	75.8	117	75.0
Female	15	38.5	35	39.3	50	39.1	17	26.2	22	24.2	39	25.0
Race												
American Indian	1	2.6	-	-	1	.8	-	-	-	-	-	-
Asian/Pacific Islander	10	26.3	22	24.7	32	25.2	2	3.1	1	1.1	3	1.9
African American	2	5.3	1	1.1	3	2.4	-	-	2	2.2	2	1.3
Hispanic	1	2.6	1	1.1	2	1.6	3	4.6	6	6.6	9	5.8
Caucasian	23	60.5	62	69.7	85	66.9	59	90.1	82	90.1	141	90.4
Other Race	1	2.6	1	1.1	2	1.6	1	1.5	-	-	1	.6
Multiracial	-	-	2	2.3	2	1.6	-	-	-	-	-	-

Note: *n* = number of participants

Table 8. (continued) Demographic Information for Co-located and Virtual Teams

	Co-located Teams						Virtual Teams					
	3-person		4/5-person		All		3-person		4/5-person		All	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	24.1	2.9	24.1	3.0	24.1	3.0	29.8	5.7	30.3	6.6	30.1	5.9
Experience	1.6	2.4	1.7	2.8	1.6	2.7	7.7	5.2	8.3	5.7	8.0	5.5

Note: Four participants in the 4/5-person co-located teams and one participant in the 3-person co-located teams did not provide some of the demographic information. Three participants in the 4/5-person virtual teams and one participant in the 3-person virtual teams did not provide some of the demographic information.

Table 9. Comparison of Overall One-factor and Two-factor Models for Self-Ratings

	Original 1-factor	Original 2-factor	Revised 1-factor	Revised 2-factor
Fit Statistics				
$\chi^2(df)$	533.56 (90)*	432.68 (89)*	459.57 (77)*	327.84 (76)*
CFI	.81	.86	.83	.89
RMSEA	.13 (.12-.14)	.12 (.11, .13)	.13 (.12, .15)	.11 (.10, .12)
SRMR	.071	.069	.070	.061
Standardized Loadings				
Interpersonal Facilitation Item 1	0.66	0.77	0.66	.78
Interpersonal Facilitation Item 2	0.58	0.67	0.59	.68
Interpersonal Facilitation Item 3	0.64	0.68	0.64	.67
Interpersonal Facilitation Item 4	0.70	0.82	0.70	.83
Interpersonal Facilitation Item 5	0.59	0.65	0.60	.67
Interpersonal Facilitation Item 6	0.60	0.54	–	–
Interpersonal Facilitation Item 7	0.67	0.65	0.67	.64

Note: All item loadings and correlations between dimensions were statistically significant at $\alpha = .05$

Table 9 (continued). Comparison of Overall One-factor and Two-factor Models for Self-Ratings

	Original 1-factor	Original 2-factor	Revised 1-factor	Revised 2-factor
Standardized Loadings				
Job Dedication Item 1	0.64	0.66	0.64	.66
Job Dedication Item 2	0.72	0.73	0.71	.73
Job Dedication Item 3	0.65	0.67	0.66	.67
Job Dedication Item 4	0.67	0.68	0.68	.69
Job Dedication Item 5	0.75	0.73	0.74	.73
Job Dedication Item 6	0.75	0.79	0.75	.79
Job Dedication Item 7	0.83	0.84	0.83	.84
Job Dedication Item 8	0.79	0.80	0.78	.80
Average Standardized Loading	0.68	0.80	0.78	0.80
IF and JD correlation	1.00	0.82	1.00	.79
$\Delta\chi^2$ (1)	100.88, $p < .0001$		131.73, $p < .0001$	

Note: All item loadings and correlations between dimensions were statistically significant at $\alpha = .05$

Table 10. Comparison of Overall One-factor and Two-factor Models for Other-Ratings

	Original 1-factor	Original 2-factor	Revised 1-factor	Revised 2-factor
Fit Statistics				
$\chi^2(df)$	500.74 (90)*	385.49 (89)*	450.44 (77)*	314.52 (76)*
CFI	.86	.90	.87	.92
RMSEA	.13 (.12-.14)	.11 (.10, .12)	.13 (.12, .14)	.10 (.09, .12)
SRMR	.064	.060	.063	.056
Standardized Loadings				
Interpersonal Facilitation Item 1	0.67	.80	0.68	0.79
Interpersonal Facilitation Item 2	0.84	.65	0.58	0.67
Interpersonal Facilitation Item 3	0.67	.69	0.65	0.68
Interpersonal Facilitation Item 4	0.75	.85	0.74	0.86
Interpersonal Facilitation Item 5	0.83	.65	0.57	0.67
Interpersonal Facilitation Item 6	0.42	.55	–	–
Interpersonal Facilitation Item 7	0.74	.68	0.70	0.67

Note: All item loadings and correlations between dimensions were statistically significant at $\alpha = .05$

Table 10 (continued). Comparison of Overall One-factor and Two-factor Models for Other-Ratings

	Original 1-factor	Original 2-factor	Revised 1-factor	Revised 2-factor
Standardized Loadings				
Job Dedication Item 1	0.64	.74	0.73	0.74
Job Dedication Item 2	0.65	.80	0.79	0.80
Job Dedication Item 3	0.71	.70	0.71	0.71
Job Dedication Item 4	0.80	.69	0.70	0.69
Job Dedication Item 5	0.74	.83	0.82	0.82
Job Dedication Item 6	0.73	.90	0.88	0.90
Job Dedication Item 7	0.75	.88	0.88	0.88
Job Dedication Item 8	0.74	.84	0.83	0.84
Average Standardized Loading	0.71	0.84	0.83	0.84
IF and JD Correlation	1.00	.85	1.00	0.83
$\Delta\chi^2$ (1)	115.25, $p < .0001$		135.92, $p < .0001$	

Note: All item loadings and correlations between dimensions were statistically significant at $\alpha = .05$

Table 11. Standardized Parameter Estimates for the Factor Loadings and Unique Variance Parameters Derived from the Configural Invariance CFA Model

Performance Dimension		Model 1				Model 2			
		Dim	Self	Peer	Uniqueness	Dim	Self	Peer	Uniqueness
Self	IF1	.69*	.41*		.36*	.73*	.33*		.36*
Other	IF1	.55*		.60*	.34*	.56*		.58*	.35*
Self	IF2	.57*	.33*		.56*	.60*	.29*		.56*
Other	IF2	.57*		.33*	.57*	.57*		.33*	.57*
Self	IF3	.49*	.48*		.54*	.50*	.45*		.55*
Other	IF3	.53*		.46*	.52*	.51*		.47*	.52*
Self	IF4	.72*	.43*		.29*	.75*	.38*		.29*
Other	IF4	.65*		.56*	.25*	.67*		.55*	.25*
Self	IF5	.55*	.25*		.64*	.57*	.21*		.63*
Other	IF5	.53*		.28*	.64*	.54*		.26*	.64*
Self	IF6	–	–	–	–	–	–	–	–
Other	IF6	–	–	–	–	–	–	–	–
Self	IF7	.38*	.56*		.55*	.39*	.53*		.57*
Other	IF7	.56*		.38*	.55*	.52*		.42*	56*
Self	JD1	*.08*	.77*		.40*	–	–	–	–
Other	JD1	.69*		.29*	.44*	–	–	–	–

Note: * $p < .05$; IF = interpersonal facilitation; JD = job dedication; Dim = performance dimension loading.

Table 11 (continued). Standardized Parameter Estimates for the Factor Loadings and Unique Variance Parameters Derived from the Configural Invariance CFA Model

Performance		Model 1				Model 2			
Dimension		Dim	Self	Peer	Uniqueness	Dim	Self	Peer	Uniqueness
Self	JD2	.26*	.71*		.43*	.25*	.72*		.42*
Other	JD2	.71*		.36*	.36*	.67*		.42*	.38*
Self	JD3	*.12*	.72*		.47*	–	–	–	–
Other	JD3	.65*		.27*	.50*	–	–	–	–
Self	JD4	.32*	.59*		.56*	.34*	.56*		.57*
Other	JD4	.61*		.31*	.54*	.59*		.33*	.54*
Self	JD5	.44*	.62*		.43*	.45*	.62*		.42*
Other	JD5	.70*		.45*	.31*	.69*		.48*	.30*
Self	JD6	.23*	.77*		.36*	.22*	.79*		.32*
Other	JD6	.81*		.38*	.20*	.79*		.41*	.20*
Self	JD7	.44*	.72*		.32*	.40*	.73*		.31*
Other	JD7	.77*		.42*	.23*	.78*		.44*	.20*
Self	JD8	.69*	.67*		.35*	.44*	.70*		.32*
Other	JD8	.73*		.41*	.30*	.72*		.44*	.29*
ASL		.54*	.57*	.39*	.43*	.55*	.53*	.43*	.42*
Self-Other Correlation				.64				.60	
IF–JD Correlation				.79				.77	

Note: * $p < .05$; IF = interpersonal facilitation; JD = job dedication; ASL = average standardized loading; Dim = performance dimension loading.

Table 12. Fit Statistics for the Multitrait-multirater Contextual Performance Models

	χ^2	*df*	$\Delta\chi^2$	Δdf	*sig.*	SRMR	CFI
Model 1							
Configural Invariance	762.81	306				.064	.94
Metric Invariance	768.41	320	5.6	14	*ns*	.120	.96
Model 2							
Configural Invariance	495.99	214				.059	.96
Metric Invariance	512.68	226	16.69	12	*ns*	.117	.96

Note: Model 1 contains all items except interpersonal facilitation item 6. Model 2 Contains all items except interpersonal facilitation item 6 and job dedication items 1 and 3.

Table 13. Configural Invariance of the Contextual Performance Ratings Provided by Co-located and Virtual Team Members

	Co-located Self-Ratings	Virtual Self-Ratings	Co-located Other-Ratings	Virtual Other-Ratings
Fit Statistics				
$\chi^2(53)$	163.18*	116.15*	134.94*	131.83*
CFI	.86	.94	.93	.94
RMSEA	.13 (.11, .15)	.09 (.07, .11)	.11 (.09.13)	.10 (.08, .12)
SRMR	.076	.053	.057	.057
Standardized Loadings				
Interpersonal Facilitation Item 1	0.80	0.79	0.83	0.77
Interpersonal Facilitation Item 2	0.67	0.66	0.70	0.62
Interpersonal Facilitation Item 3	0.68	0.67	0.71	0.67
Interpersonal Facilitation Item 4	0.84	0.83	0.86	0.85
Interpersonal Facilitation Item 5	0.68	0.61	0.65	0.66

Note: Separate CFAs were run for co-located and virtual team members in both the self- and other-rating data set. All item loadings and correlations between dimensions were statistically significant at $\alpha = .05$

Table 13 (continued). Configural Invariance of the Contextual Performance Ratings Provided by Co-located and Virtual Team Members

	Co-located Self-Ratings	Virtual Self-Ratings	Co-located Other-Ratings	Virtual Other-Ratings
Interpersonal Facilitation Item 7	0.58	0.64	0.67	0.65
Job Dedication Item 2	0.68	0.74	0.78	0.80
Job Dedication Item 4	0.65	0.68	0.66	0.69
Job Dedication Item 5	0.72	0.75	0.87	0.79
Job Dedication Item 6	0.74	0.80	0.88	0.90
Job Dedication Item 7	0.85	0.84	0.90	0.88
Job Dedication Item 8	0.78	0.85	0.82	0.87
Average Standardized Loading	0.72	0.74	0.77	0.76
IF and JD correlation	0.73	0.83	0.85	0.81

Note: Separate CFAs were run for co-located and virtual team members in both the self- and other-rating data set. All item loadings and correlations between dimensions were statistically significant at $\alpha = .05$

Table 14. Metric Invariance of the Contextual Performance Ratings Provided by Co-located and Virtual Team Members

Model Description	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*Statistical Significance*
Model–Self-ratings						
1. Baseline Model	339.14	128	.000			
2. Factor loadings constrained equal	328.68	116	.000	10.46	12	*ns*
Model–Other ratings						
1. Baseline Model	323.50	128	.000			
2. Factor loadings constrained equal	308.40	116	.000	15.10	12	*ns*

Note: In the baseline model all factor loadings and latent variable correlations are freely estimated. These CFAs are estimated simultaneously for both groups. All model comparisons are made to the baseline model.

Table 15. Comparison of Task Interdependence and Contextual Performance by Instructor

	Instructor 1			Instructor 2			
	n	*M*	*SD*	n	*M*	*SD*	*t*(264)
Self-Ratings	159			107			
Task Interdependence		4.08	.89		4.18	.91	-0.85
Contribution		4.24	.84		4.33	.83	-0.84
Interpersonal Facilitation		3.60	.94		3.73	.92	-1.12
Job Dedication		3.92	.72		3.90	.85	0.16
Other-Ratings	160			119			
Task Interdependence		4.07	.89		4.22	.89	-1.39
Contribution		4.02	.77		4.26	.71	-2.61*
Interpersonal Facilitation		3.46	.92		3.69	.86	-2.12*
Job Dedication		3.69	.77		3.87	.77	-1.98*

Note: List-wise deletion used. *p <.05

Table 16. Means, Standard Deviations, and Correlations for Study Variables

Variable	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.
1. Task Interdependence	4.12	.90	–						
2. Contribution—Self-Ratings	4.28	.84	.13*	–					
3. Contribution—Other-Ratings	4.12	.75	.26*	.25*	–				
4. Interpersonal Facilitation—Self-Ratings	3.65	.93	.19*	.40*	.40*	*.83*			
5. Interpersonal Facilitation—Other-Ratings	3.55	.90	.23*	.25*	.54*	.87*	*.83*		
6. Job Dedication—Self-Ratings	3.91	.77	.23*	.60*	.31*	.73*	.61*	*.88*	
7. Job Dedication—Other-Ratings	3.78	.77	.31*	.20*	.72*	.64*	.77*	.63*	*.92*

*$p < .05$; List-wise deleted n of 265.The interpersonal facilitation and job dedication revised scales are used in the correlation matrix.

Table 17. Means, Standard Deviations, and Correlations for Study Variables by Team Type

	Co-located Teams		Virtual Teams		*t*							
Variable	*M*	*SD*	*M*	*SD*	(263)	1.	2.	3.	4.	5.	6.	7.
1. Task Interdependence	4.15	.96	4.09	.85	0.51	–	.25*	.25*	.29*	.28*	.36*	.32*
2. Contribution—Self-Ratings	4.29	.82	4.27	.86	0.20	.00	–	.18*	.51*	.28*	.66*	.19*
3. Contribution—Other-Ratings	4.19	.74	4.06	.75	1.39	.26*	.34*	–	.38*	.54*	.29*	.71*
4. Interpersonal Facilitation—Self-Ratings	3.91	.83	3.42	.95	4.37*	.07	.25*	.40*	–	.84*	.76*	.62*
5. Interpersonal Facilitation—Other-Ratings	3.80	.84	3.34	.90	4.29*	.18	.22*	.54*	.89*	–	.56*	.77*
6. Job Dedication—Self-Ratings	4.03	.70	3.81	.83	2.25*	.07	.53*	.33*	.66*	.65*	–	.61*
7. Job Dedication—Other-Ratings	3.89	.78	3.67	.76	2.30*	.30*	.22*	.72*	.64*	.77*	.65*	–

* $p < .05$; List-wise deleted *n* of 122 for co-located teams (below the diagonal) and *n* of 143 for virtual teams (above the diagonal).The interpersonal facilitation and job dedication revised scales are used in the correlation matrix.

Table 18. Construct Variance Partitioning

Construct	Target	Perceiver	Relationship	Residual	Total
Co-located 4+ Person Teams					
Contextual Performance	13%	54%*	7%*	26%	0.74
Interpersonal Facilitation	13%	76%*	11%		0.72
Job Dedication	16%	66%*	19%		0.76
Workload	18%	45%*	37%		1.01
Virtual 4+ Person Teams					
Contextual Performance	26%*	39%*	11%*	24%	0.85
Interpersonal Facilitation	16%*	67%*	17%		0.88
Job Dedication	42%*	37%*	22%		0.82
Workload	48%*	14%*	38%		1.18

Note: The relative variances are reported for each of interpretation, but significance of the variance components were conducted on the absolute variance components.

Table 19. Statistical Tests for Differences in the Variance Components between Co-located and Virtual Teams

Construct	Target	Perceiver	Relationship
Contextual Performance	$t(44) = -0.98$, $p = .33$ $d = .41$	$t(44) = 0.54$, $p = .59$ $d = .46$	$t(44) = -1.06$, $p = .29$ $d = .21$
Interpersonal Facilitation	$t(44) = -0.46$, $p = .65$ $d = .10$	$t(44) = -0.29$, $p = .78$ $d = .14$	$t(44) = -1.73$, $p = .09$ $d = .49$
Job Dedication	$t(44) = -1.27$, $p = .21$ $d = .49$	$t(44) = 1.40$, $p = .17$ $d = .42$	$t(44) = -0.65$, $p = .51$ $d = .19$
Workload	$t(44) = -1.54$, $p = .13$ $d = .16$	$t(44) = 1.39$, $p = .17$ $d = .30$	$t(44) = -0.74$, $p = .46$ $d = .31$

Note: Theses analyses were conducted for 4-plus person teams.

Table 20. Statistical Tests for Differences in the Target, Perceiver, and Relationship Components in Co-located and Virtual Teams

Construct	Comparison	*t*(22)	*p*
Co-located 4/5 Person Teams			
Contextual Performance	perceiver - target	2.44	0.023
Contextual Performance	perceiver - relationship	3.26	0.004
Contextual Performance	target - relationship	0.49	0.632
Interpersonal Facilitation	perceiver - target	3.62	0.002
Job Dedication	perceiver - target	2.53	0.019
Workload	perceiver - target	2.18	0.040
Virtual 4/5 Person Teams			
Contextual Performance	perceiver - target	0.76	0.457
Contextual Performance	perceiver - relationship	2.72	0.013
Contextual Performance	target - relationship	1.46	0.157
Interpersonal Facilitation	perceiver - target	3.31	0.003
Job Dedication	perceiver - target	-0.30	0.768
Workload	perceiver - target	-0.48	0.634

Note: The relative variances are reported for each of interpretation, but significance of the variance components were conducted on the absolute variance components.

Table 21. Target-Target and Perceiver-Perceiver Correlations for Co-located and Virtual Teams

	Target-Target Correlations			Perceiver-Perceiver Correlations		
Variable	1.	2.	3.	1.	2.	3.
1. Workload Contribution	–	0.88	1.00	–	.63	.76
2. Interpersonal Facilitation	0.97	–	0.89	.64	–	.77
3. Job Dedication	0.98	1.00	–	.55	.78	–

Notes: Correlations above the diagonal are target-target correlations for co-located teams. Correlations below the diagonal are target-target correlations for virtual teams. All correlations are statistically significant at $\alpha = .05$.

Table 22. Self-Target Correlations for Co-located and Virtual Teams

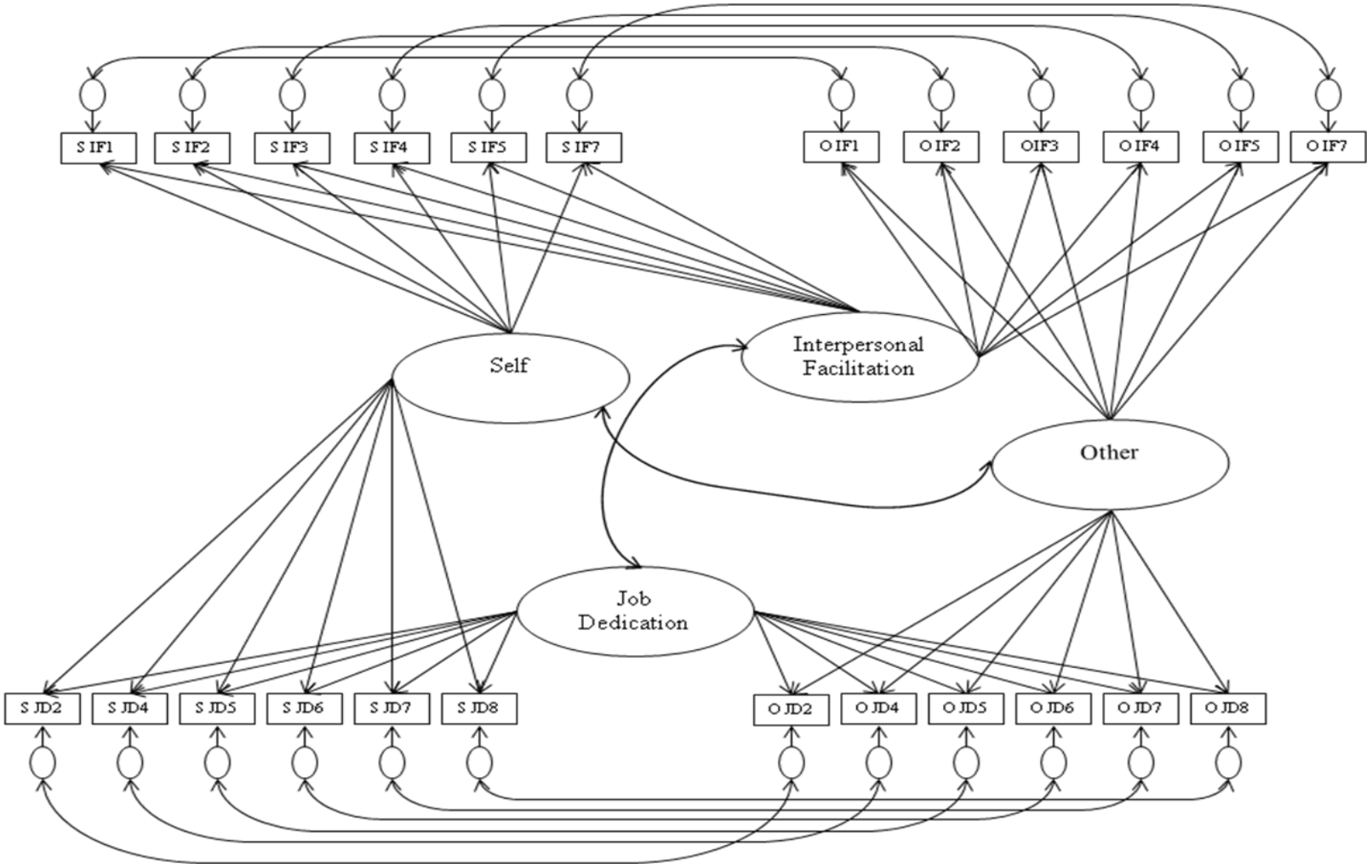
Variable	Co-located Teams	Virtual Teams
1. Workload Contribution	.51*	.62*
2. Interpersonal Facilitation	.13	.52*
3. Job Dedication	.11	.33*

Note: * $p < .05$.

Table 23. Self-Other Comparisons

Construct	Mean Self- Ratings	Mean Other Ratings	T-test (df = 22)	Mean Self-enhancement Index
Co-located Teams				
Workload	4.33	4.20	1.50	.13
Interpersonal Facilitation	3.97	3.84	2.93*	.14*
Job Dedication	4.07	3.90	2.65*	.13
Virtual Teams				
Workload	4.20	3.97	2.53*	.24*
Interpersonal Facilitation	3.40	3.29	1.90	.11
Job Dedication	3.75	3.58	1.77	.17

Figure 1. Multitrait-multirater confirmatory factor analysis model of the two contextual performance dimensions and the two rating sources. For the item labels: S = self-rating, O = other-rating, IF = interpersonal facilitation, and JD = job dedication.



Appendix

Demographics

Gender:

Male

Female

Racial/Ethnic Background:

American Indian

Asian/Pacific Islander

African American

Hispanic-American

Caucasian

Other

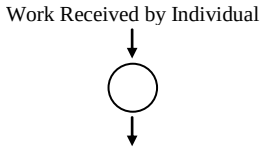
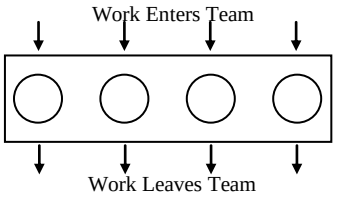
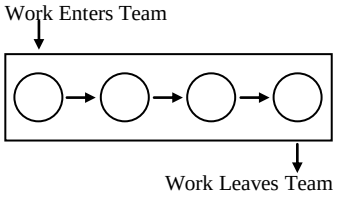
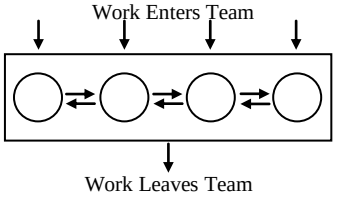
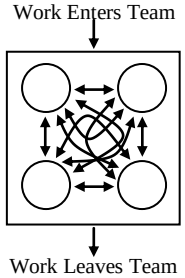
I'd prefer not to answer

Age:

Years of professional work experience:

Task Interdependence

This chart presents four TEAM WORK FLOW PATTERNS as well as a description and illustration of each pattern. Using the chart, please select the team workflow pattern that BEST characterizes the OVERALL pattern in your work group during the group project. That is, identify the team workflow pattern that best characterizes the way that work between team members flowed for the optimal performance of the group.

Team Workflow Pattern	Description	Illustration
1. Not a Team Job	Work and activities are ***NOT*** performed as a member of a team; they are performed alone outside the context of the team. Work and activities are performed by an individual working ***ALONE***, ***NOT*** in a team.	
2. Pooled/Additive Interdependence	Work and activities are performed separately by all team members and work does not flow between members of the team.	
3. Sequential Interdependence	Work and activities flow from one member to another in the team, but mostly in one direction.	
4. Reciprocal Interdependence	Work and activities flow between team members in a back–and–forth manner over a period of time.	
5. Intensive Interdependence	Work and activities come into the team and members must diagnose, problem solve, and/or collaborate as a team in order to accomplish the team's task.	

Workload

The following statements describe interactions between you and your teammate(s). Rate the extent to which each of the following statements is descriptive of you and your teammate(s) with respect to your interactions as a team. Where you see your name, please rate yourself.					
		Name of Teammate 1	Name of Teammate 2	Name of Teammate 3	Name of Teammate 4
1.	did your/their fair share of the work.	PD	PD	PD	PD

Note: The name of each team member in a group was presented at the top of one of the four columns. Where participants saw their own name they rated themselves on that item. Where the participants saw a teammates name they rated that particular teammate on that item Participants provided ratings using a pull down menu. The pull down (PD) options were 1 = to a very little extent, 2 = to a little extent, 3 = to some extent, 4 = to a great extent, and 5 = to a very great extent.

Contextual Performance-Interpersonal Facilitation

The following statements describe interactions between you and your teammate(s). Rate the extent to which each of the following statements is descriptive of you and your teammate(s) with respect to your interactions as a team. Where you see your name, please rate yourself.					
		Name of Teammate 1	Name of Teammate 2	Name of Teammate 3	Name of Teammate 4
1.	praised team members when they are successful.	PD	PD	PD	PD
2.	supported or encouraged a team member with a personal problem.	PD	PD	PD	PD
3.	talked to other team members before taking actions that might affect them.	PD	PD	PD	PD
4.	say things to make people feel good about themselves or the team.	PD	PD	PD	PD
5.	encouraged others to overcome their differences and get along.	PD	PD	PD	PD
6.	treated others fairly.	PD	PD	PD	PD
7.	helped someone without being asked.	PD	PD	PD	PD

Note: The name of each team member in a group was presented at the top of one of the four columns. Where participants saw their own name, they rated themselves on that item. Where the participants saw a teammate's name, they rated that particular teammate on that item. Participants provided ratings using a pull down menu. The pull down (PD) options were 1 = to a very little extent, 2 = to a little extent, 3 = to some extent, 4 = to a great extent, and 5 = to a very great extent.

Contextual Performance-Job Dedication

The following statements describe interactions between you and your teammate(s). Rate the extent to which each of the following statements is descriptive of you and your teammate(s) with respect to your interactions as a team. Where you see your name, please rate yourself.					
		Name of Teammate 1	Name of Teammate 2	Name of Teammate 3	Name of Teammate 4
1.	put in extra hours to get work done on time.	PD	PD	PD	PD
2.	paid close attention to important details.	PD	PD	PD	PD
3.	worked harder than necessary.	PD	PD	PD	PD
4.	asked for a challenging work assignment.	PD	PD	PD	PD
5.	exercised personal discipline and self-control.	PD	PD	PD	PD
6.	took the initiative to solve a problem.	PD	PD	PD	PD
7.	persisted in overcoming obstacles to complete a task.	PD	PD	PD	PD
8.	tackled a difficult assignment enthusiastically.	PD	PD	PD	PD

Note: The name of each team member in a group was presented at the top of one of the four columns. Where participants saw their own name, they rated themselves on that item. Where the participants saw a teammate's name, they rated that particular teammate on that item. Participants provided ratings using a pull down menu. The pull down (PD) options were 1 = to a very little extent, 2 = to a little extent, 3 = to some extent, 4 = to a great extent, and 5 = to a very great extent.