Team Performance: The Effect of Team Conscientiousness and Task Type
Andrew English, Richard L. Griffith and Lisa A. Steelman
Small Group Research 2004 35: 643
DOI: 10.1177/1046496404266320

The online version of this article can be found at:
http://sgr.sagepub.com/content/35/6/643

Published by:
SAGE
http://www.sagepub.com

Additional services and information for Small Group Research can be found at:
Email Alerts: http://sgr.sagepub.com/cgi/alerts
Subscriptions: http://sgr.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://sgr.sagepub.com/content/35/6/643.refs.html

Version of Record - Nov 2, 2004
What is This?
The purpose of this research was to examine the relationship between measures of conscientiousness at differing levels of analysis and team performance in the context of task type. It was hypothesized that a team referent measure of conscientiousness would have more predictive power than an aggregate of individual-level measures and that task type would moderate the relationship between team conscientiousness and overall team performance. Participants were 30 cockpit crews, made up of three pilots each. Both an individual-level and team referent measure of conscientiousness were administered to the pilots, and crew performance appraisal data were collected that separated performance into overall, additive, disjunctive, and conjunctive task types. Results indicated that the team referent measure of conscientiousness predicted overall team performance greater than did the individual-level measure. The results also partially supported the moderating effect of task type.

Keywords: teams; personality; frame of reference; bandwidth; levels of analysis

The continuing utilization of teams in the workplace has become a catalyst for an explosion of research in the area (Kirkman, Tesluck, & Rosen, 2001; Marks, Sabella, Burke, & Zaccaro, 2002). Empirical evidence suggesting the effectiveness of teams has begun to accrue (e.g., Majuka & Baldwin, 1991; Neuman, Edwards, & Raju, 1989; Neuman & Wright, 1999). A team is defined as a group of individuals working interdependently to achieve a common goal (Sundstrom, de Meuse, & Futrell, 1990). The complex nature of the modern work environment has made the use of teams a
necessity for organizations wishing to stay at the cusp of competitiveness. As organizations have increasingly reorganized around work teams, researchers have begun to look at a variety of variables that may affect team viability (Kirkman et al., 2001).

Recently, the focus of team research has fallen under one of three main areas of interest. The first area of focus includes the characteristics of the individual team members (Barrick, Stewart, Neubert, & Mount, 1998; Neuman & Wright, 1999). In this area, researchers are most interested in team member ability and personality constructs and their relationships to team effectiveness. The second area of focus has been understanding group processes. Research in this area has most recently shifted to understanding the influence that shared mental models have on team effectiveness (Marks et al., 2002; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Finally, the third area of focus has revolved around determining how to best measure team-level constructs (Kirkman et al., 2001; Klein, Conn, Smith, & Sorra, 2001). This study addressed each of these areas of interest. First, this article examined the influence of two variables on team effectiveness: personality and task type. Second, this article considered the utility gained from using team referent measures versus aggregate individual-level measures.

The effectiveness of the Big Five taxonomy of personality, especially the construct of conscientiousness, in predicting individual performance has been consistently demonstrated (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). The personality trait of conscientiousness is characterized by planning, thoroughness, hard work, and purposeful striving toward goals (Digman, 1990). In fact, Barrick et al. (1998) concluded that there “are two dispositional predictors in our field whose validity generalizes: general mental ability and conscientiousness.”

Due to the interdependent nature of teams and the interpersonal demands that interdependence poses, personality traits should be related to team performance (Moreland & Levine, 1992). Barrick et al. (1998) showed that teams possessing higher levels of emotional stability and extraversion were most likely to experience more positive interactions and increased social cohesion. However,
research relating personality and team effectiveness is equivocal (Neuman, Wagner, & Christiansen, 1999) largely because, within a team context, there is little consensus on how personality should be defined and measured. This study examined the relative influence of an individual measure of personality and a team referent measure of personality in predicting team performance. This study also examined the moderating effects of task type on those relationships.

Steiner (1972) categorized team tasks into several distinct types depending on the relative amount of team process interaction and interdependence. A team may divide its labor in several distinct ways; this article examined how this division of labor affected the magnitude of the relationship between personality factors and team performance.

Thus, the purpose of this article was twofold. The first goal of this article was to test the relative contribution of (a) an individual-level measure of conscientiousness and (b) a team referent measure of conscientiousness in predicting team performance. The second goal of this article was to assess the extent to which task type moderates the relationship between team-level conscientiousness and team-level performance. The study described in this article was conducted with a sample of airline cockpit crews made up of a captain, first officer, and second officer. The finding that 60% of all aviation accidents are due to human error (Lauber, 1987) underscores the need to explore how cockpit crews function as a team. The importance of personality characteristics for cockpit crew performance has only recently been recognized (Monfries & Moore, 1999). Further, cockpit crews perform several different types of tasks as a team and thus are an appropriate sample to highlight possible personality/task interactions and their effect on performance.

PERSONALITY AT THE INDIVIDUAL AND TEAM LEVEL

In the past several years, much research on personality and team performance has emerged (Barry & Stewart, 1997; LePine, Hollenbeck, Ilgen, & Hedlund, 1997) and reviews have concluded
that a relationship between personality and team performance does exist (Jackson, 1992; Moreland & Levine, 1992). For instance, the personality characteristics of conscientiousness and agreeableness have been shown to predict peer and supervisor ratings of work team performance (Neuman & Wright, 1999). Furthermore, Barrick et al. (1998) reported that work teams with higher mean levels of the personality characteristics of conscientiousness, agreeableness, extraversion, and emotional stability had higher supervisor ratings of team performance.

The construct of conscientiousness includes feelings of competence, achievement striving, self-discipline, and dutifulness (Costa, McCrae, & Dye, 1991). Teams made up of members who are conscientious are more deliberate, organized, and task focused, which should be related to effective team performance. Research has shown a relationship between achievement motivation (a component of conscientiousness) and group performance (Schneider & Delaney, 1972). Furthermore, research by Goeters (1995) indicated that for experienced pilots, nearly half of the problems in a sample of 200 pilots were related to interpersonal factors, not technical performance.

However, the studies that examined the effect of personality on team performance were generally based on an aggregate of individual-level personality as a proxy for team personality composition. This aggregation is often obtained by averaging individual team member scores on the variable of interest to derive a team-level score (Barrick et al., 1998) or by choosing one team member’s score to represent the team-level score (Neuman & Wright, 1999). The latter method is used only when there is an a priori theoretical reason to focus entirely on the highest or lowest scoring individual. Ostroff (1993) has shown that the relationships between predictors and criterion may differ when examined at the group and individual levels. She found that organizational- (team-) level correlations can be different from individual-level correlations depending on several factors. This is referred to as the fallacy of the wrong level. The fallacy occurs when correlations at a higher level of analysis (aggregations) are used to make inferences about individuals and vice versa. An aggregated correlation and an individual correlation are not
always equal (Ostroff, 1993). Failure to consider the context of the
correlations can result in erroneous interpretations about the
strength of the relationships among variables at differing levels of
analysis. Previous research on the personality composition of
teams has generally aggregated individual personality scores to
obtain a team-level score. It was asserted here that individual per-
sonality scores do not fully generalize to the team level of that spe-
cific trait. In other words, behavior may differ according to the con-
text (individual versus team). Individuals may behave and rate their
behavior differently when they are working in a team versus when
they are working alone. Individuals may also rate their individual
behavior differently from their team’s behavior. If the focal unit is
the team and the team’s performance, team referent measures
should be developed and used to predict team performance
(Tesluck, Mathieu, Zaccaro, & Marks, 1997). Thus, the purpose of
this study was to examine the extent to which team referent mea-
sures of personality predict team performance over and above the
effects of an individual-level measure of personality.

Conscientiousness was collected at both the individual and the
team referent level. The unique effect of these two predictors on a
team-based measure of performance was examined. Research has
continued to suggest that context-specific measurement will in-
crease predictive validity; error variances tend to be slightly lower
and predictive validity higher for context-specific instruments ver-
sus general instruments (Robie, Schmit, Ryan, & Zickar, 2000;
Schmit, Ryan, Stierwalt, & Powell, 1995). Thus, the following was
hypothesized.

Hypothesis 1: The team referent conscientiousness measure will
predict overall team performance over and above the aggregated
individual-level conscientiousness measure.

TASK TYPE

The extent to which task type moderated the predictor-criterion
relationships discussed was also examined. The critical features of
team tasks must be examined more closely to fully grasp the com-
plexity of personality constructs within a team and their relationship to overall team performance. Many theoretical frameworks have been developed that classify tasks on the basis of critical features. Hackman, Jones, and McGrath (1967) discussed production, discussion, and problem-solving tasks. Tasks completed by cooperative versus competitive groups have also been differentiated by Davis, Laughlin, and Komorita (1976). Finally, relevant for this study, Steiner (1972) developed a taxonomy distinguishing between divisible and unitary tasks, with further classification of unitary tasks into four additional distinct task types. Divisible tasks are those where it is feasible to further divide the labor, and unitary tasks are those where a division of labor is not feasible (Steiner, 1972).

Steiner (1972) reported that there are four types of unitary tasks. The first is the additive task, which requires the summing or averaging of individual resources for performance. An example of this task is brainstorming, where performance is based on the quantity of answers. Team performance will improve to the extent that each member adds to the collective pool of the overall team. Performance will increase proportionately to the ability of the team’s members, regardless of the distribution of ability among team members (Devine & Philips, 2001).

Steiner’s second task type is the conjunctive task, in which performance is based on the team’s lowest performer. An example of this task type is an assembly line where each assembler’s performance is contingent on the assembler previous in line. The line will only move as fast as the slowest assembler can keep up. The third type of task is the disjunctive task, in which performance is based on the highest performing member of the team. An example of this type of task is problem solving (Barrick et al., 1998). Problem solving is a disjunctive task if one best answer will represent the team’s performance. Steiner’s final task type is the discretionary task. In this task type, team members are able to combine their resources however they wish, usually weighting the contributions of high performers over low performers. The discretionary task is not representative of tasks that typical work teams perform because only self-managed work teams have the authority to autonomously de-
cide how to combine their resources. A management team may perform a variety of discretionary tasks when they are implementing organizational initiatives, as they decide how to allot their time, organize team structure, use resources, and evaluate their efforts. The majority of teams in today’s work force face tasks represented by the additive, conjunctive, and disjunctive classifications.

Previous research examined the interaction of conscientiousness and task type on the performance of different types of tasks. For instance, Waung and Brice (1998) conducted a study that examined the effects of conscientiousness on group performance on an additive task. It was hypothesized that the high-conscientious groups would perform at higher levels on a “generation of ideas” task than would the low-conscientious groups. They found that the high-conscientious groups generated higher quality ideas at the expense of quantity, whereas the low-conscientious groups generated a higher quantity of ideas at the expense of quality. Neuman and Wright (1999) examined the relationship between the personality traits of conscientiousness and agreeableness and conjunctive tasks. They found that these two personality traits predicted team performance over and above job-specific skills and general ability alone.

This study made several improvements on the previous research examining personality and task type. Previous studies dichotomized conscientiousness when examining its relationship with performance (e.g., Waung & Brice, 1998). By dichotomizing a continuous variable, such as conscientiousness, much variability is lost. This study examined conscientiousness as a continuous variable. Second, by focusing on three task types, we hoped to provide additional information on the personality construct of conscientiousness as it relates to team performance on varying types of tasks.

For this study, cockpit crews made up of a captain, first officer, and second officer were examined. The job of an airline pilot includes a combination of task types. It was believed that task type would moderate the relationship between team conscientiousness and task performance such that team conscientiousness would be related to the additive and disjunctive tasks the pilots perform, but a relationship would not be found between team conscientiousness
and the conjunctive task (see Figure 1). These predictions are discussed in the following paragraphs.

In additive tasks, each individual’s performance is summed and averaged to produce the team outcome. Social facilitation (the effect of increasing performance when working in the presence of others as opposed to when working alone) should occur during the additive task because each individual’s level of performance will contribute to the final team outcome, and there is social pressure to perform (Erez & Somech, 1996). Highly conscientious teams should reduce the problem of social loafing and support a climate that will encourage achievement of the groups’ objectives. Thus, the following was hypothesized.

**Hypothesis 2a:** There will be a significant relationship between team conscientiousness and team performance on additive tasks.

In the disjunctive task, the team outcome relies on the best performer. Highly conscientious team members have strong achievement motivation and are thus driven to perform their best at all times. Highly conscientious teams will push individual team members to perform at the highest levels, resulting in the high-conscientious teams’ producing high levels of performance on this
type of task. The low-conscientious teams will show little concern for who outperforms whom and therefore produce lower levels of performance on this task. Thus, the following was hypothesized.

**Hypothesis 2b:** There will be a significant relationship between team conscientiousness and team performance on disjunctive tasks.

In conjunctive tasks, the team outcome relies on the lowest performer. In this type of task, high-conscientious teams will not be forced to compete among themselves and thus overall decrements in performance should be observed. Equity theory would suggest that the high-conscientious individuals would exert the minimal amount of effort required to avoid being the lowest performer in the group. Under certain conditions, performance on conjunctive tasks may be predicted by conscientiousness, such that members low in the trait may elevate portions of their performance on high-conscientiousness teams. The level of performance related to conscientiousness may in fact increase but should stay close to the person’s true score level of conscientiousness over time. The low-scoring team member may be somewhat more conscientious but still not perform as high as his or her more conscientious team members. Steiner’s (1972) study found that teams rarely perform at the level of the highest team member. Previous research has found a significant correlation between the lowest team member’s score on conscientiousness and an overall measure of performance (Barrick et al., 1998; Neuman & Wright, 1999). However, our research is different in that it examined the relationship of conscientiousness and performance dimensions identified as conjunctive tasks. Conjunctive tasks are subject to a bottleneck, and a low-performing member can have a robust negative effect, such as the sucker effect (Kerr, 1983) and the free rider effect (Kerr & Bruun, 1983), on group dynamics. Prince, Brannick, Prince, and Salas (1997) suggested that if a low-performing team member makes a mistake and the team does not catch it, the error becomes a team error. Thus, the following was hypothesized.

**Hypothesis 2c:** There will not be a significant relationship between team conscientiousness and team performance on conjunctive tasks.
METHOD

PARTICIPANTS

Participants were 30 three-member cockpit crews from a large international airline carrier for a total of 90 individuals. Each crew was made up of a captain, a first officer, and a second officer. Eighty-eight of the 90 individuals sampled were males with the remaining 2% of the sample represented by 2 females. Team members ranged in age from 25 to 59 years, and tenure in the field ranged from less than 10 years to 40 years.

MEASURES

Two measures of conscientiousness, discussed below, were used in collecting the predictor data. The Team Conscientiousness Inventory (TCI) was used to assess conscientiousness using team referent items, and the Summated Conscientiousness Scale (SCS) was used to assess conscientiousness at the individual level.

TCI

To measure the level of conscientiousness within each team, a team referent conscientiousness measure was developed that assessed conscientiousness along the six dimensions of competence, order, dutifulness, achievement striving, self-discipline, and deliberation and was scored on a 7-point Likert-type scale. The TCI measure included four items each for the competence, order, dutifulness, self-discipline, and deliberation subscales and five items for the achievement striving subscale. The instructions for this scale asked the pilots to respond to the statements in the specific context of their team versus in general. A sample item read, “We try to do all jobs carefully, so they won’t have to be done again.” Individual scores on the TCI were aggregated and averaged to represent each team’s average team-level conscientiousness.
SCS

The SCS was used to obtain a general and global measure of conscientiousness at the individual level. This instrument was designed to measure conscientiousness in an applicant setting, has a previously demonstrated reliability of .84, and is correlated with the NEOPI-R conscientiousness scale at .75 (Griffith, Chmielowski, Snell, Frei, & McDaniel, 2000). The SCS is a 20-item instrument in which applicants are asked to respond to self-descriptive statements on a 7-point Likert-type scale. A sample item read, “I feel it is important to complete all of the tasks assigned to me by my supervisor.” Individual scores on the SCS were aggregated and averaged to represent each team’s average individual level of conscientiousness.

CREW PERFORMANCE APPRAISAL

A crew performance appraisal was developed to assess team performance. The performance appraisal was developed from a subject matter expert interview and used existing performance appraisal information from the organization. Whereas many team performance appraisals aggregate individual performance measures to produce an overall team performance measure, this crew performance appraisal was designed to assess overall crew performance at the team level. Congruent with the TCI and SCS measures, performance appraisal items were answered on a 7-point Likert-type scale (1 = strongly disagree to 7 = strongly agree).

The crew performance appraisal was developed to collect criterion data on each of the three distinct task types (additive, disjunctive, and conjunctive), as well as overall performance. Thus, an item measuring performance for the additive task was “Situational Awareness (SA) of the crew is always at a high level.” Maintaining situational awareness is consistent with Steiner’s (1972) taxonomy that suggested that additive tasks are structured so that the contribution of each member adds to a collective pool that can be used to help the team succeed. A disjunctive task pilots perform is identifying and documenting all aircraft mechanical discrepancies.
Because this is crucial to maintaining successful aircraft operation, it is the responsibility of each pilot to continually monitor, check, and report these mechanical discrepancies during flight. An unreported discrepancy may result in a future mechanical failure. Thus, an item measuring performance for the disjunctive task was “The crew recognizes and documents all aircraft mechanical discrepancies.” A conjunctive task that pilot crews perform is communicating with air traffic control. Typically, one pilot is assigned the duty of communicating directly with air traffic control. However, the message communicated is a product of the entire crew, and if there is a bottleneck in the information due to a low-performing member, disaster can occur. Thus, an item measuring performance for the conjunctive task was “The crew communicates clearly and effectively with air traffic control.” Finally, an item measuring overall crew performance was “The crew performs at high levels overall.” Crew captains, the crew supervisor, provided the criterion performance data.

PROCEDURE

Crew packets were developed to collect the data. Each packet contained three individual envelopes, one for each participating crewmember. Each individual envelope was labeled “captain,” “first officer,” or “second officer.” An informed consent form and the TCI and SCS scales were included inside each envelope. In addition to these items, the captain’s envelope also contained the crew performance appraisal.

Participants were asked to read and initial the informed consent form and then provide some brief biographical data. The biographical data collected four general pieces of information. Position in the crew, gender, age, and overall tenure in the field were collected from each participant. Age and tenure were collected in a distributed range manner so that anonymity could be guaranteed. Upon completing the biographical questions, crewmembers were asked to complete the SCS and the TCI, which took approximately 25 minutes to complete. Upon completion of the conscientiousness scales, the crew captain was asked to complete the crew perfor-
Performance appraisal. After all of the required measures were collected, crewmembers were asked to seal the individual envelopes and return them to the crew packet. The crew packet was self-addressed and stamped to the first author, so that a crewmember could immediately place the packet into a mailbox, further ensuring anonymity.

RESULTS

Descriptive statistics for all study variables may be found in Table 1. Internal consistency reliabilities for each conscientiousness scale were well within acceptable limits. Variable intercorrelations can also be found in Table 1. The individual (SCS) and team (TCI) measures of conscientiousness were moderately correlated at .50. A one-sample $t$ test was conducted that identified significant differences between these two measures ($p < .01$). The TCI mean at 150.47 ($SD = 10.30$) was significantly greater than the SCS mean at 117.78 ($SD = 8.84$). Furthermore, the TCI was significantly correlated with overall team performance ($r = .39$) but the SCS was not significantly correlated with overall team performance ($r = .26$, n.s.).

Hypothesis 1, that team referent conscientiousness (TCI) would predict team performance over and above aggregated individual-level conscientiousness (SCS), was tested using hierarchical regression analysis. The individual measure of conscientiousness

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SCS</td>
<td>117.8</td>
<td>8.8</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>TCI</td>
<td>150.5</td>
<td>10.3</td>
<td>.50**</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Overall performance</td>
<td>19.5</td>
<td>1.4</td>
<td>.26</td>
<td>.39*</td>
<td>.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Additive</td>
<td>19.2</td>
<td>1.0</td>
<td>.41*</td>
<td>.42*</td>
<td>.57**</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Disjunctive</td>
<td>19.2</td>
<td>1.8</td>
<td>.36</td>
<td>.50**</td>
<td>.52**</td>
<td>.58**</td>
<td>.89</td>
</tr>
<tr>
<td>6</td>
<td>Conjunctive</td>
<td>19.5</td>
<td>1.0</td>
<td>.23</td>
<td>.35</td>
<td>.32</td>
<td>.54</td>
<td>.48**</td>
</tr>
</tbody>
</table>

NOTE: Internal consistency reliabilities are found on the diagonal. SCS = Summated Conscientiousness Scale; TCI = Team Conscientiousness Inventory.

* $p < .05$. ** $p < .01$. 

English et al. / CONSCIENTIOUSNESS AND TASK TYPE 655
(SCS) was entered at Step 1 and the group measure of conscientiousness (TCI) was entered at Step 2. Results are presented in Table 2. For overall team performance, the TCI accounted for a nonsignificant $R^2$ change = .09 above that accounted for by the SCS. Thus, Hypothesis 1 was not supported. Furthermore, the team-level measure did not add a significant increment over and above the individual-level measure to the prediction of team performance for additive or conjunctive tasks but was significant for disjunctive tasks.

To test Hypothesis 2, the subgroup-based correlation coefficient (SCC) method was used. When the proposed moderator variable (here, task type) is not continuous, the SCC approach is the most appropriate method of analysis (Alexander & DeShon, 1994). The SCC approach entails computing correlation coefficients between two variables (in this case, between team conscientiousness and each of the task types) and then testing the equality of these correlations using Fisher $z$ score transformations. Using the SCC approach, differences in the strength or degree of a relationship will be reflected by differences in the magnitude of zero-order correlation coefficients across subgroups (Stone-Romero & Anderson, 1994). Three Fisher $z$ scores were computed between the team-

**TABLE 2: Results of Hierarchical Regression Analyses**

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>$B$</th>
<th>Significant Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCS</td>
<td>.067</td>
<td>.067</td>
<td>10.51</td>
<td>.167</td>
</tr>
<tr>
<td>TCI</td>
<td>.160</td>
<td>.092</td>
<td>3.30</td>
<td>.096</td>
</tr>
<tr>
<td>Additive task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCS</td>
<td>.168</td>
<td>.168</td>
<td>8.503</td>
<td>.024</td>
</tr>
<tr>
<td>TCI</td>
<td>.242</td>
<td>.074</td>
<td>3.678</td>
<td>.116</td>
</tr>
<tr>
<td>Disjunctive task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCS</td>
<td>.130</td>
<td>.130</td>
<td>3.155</td>
<td>.051</td>
</tr>
<tr>
<td>TCI</td>
<td>.281</td>
<td>.152</td>
<td>-8.672</td>
<td>.024*</td>
</tr>
<tr>
<td>Conjunctive task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCS</td>
<td>.051</td>
<td>.051</td>
<td>13.851</td>
<td>.228</td>
</tr>
<tr>
<td>TCI</td>
<td>.131</td>
<td>.080</td>
<td>9.053</td>
<td>.127</td>
</tr>
</tbody>
</table>

* Indicates a significant change at the .05 level.
level conscientiousness predictor and three task-type correlations to test for significant subgroup differences in the correlations.

Pearson product–moment correlation coefficients were computed between the TCI and each of the three distinct task types, and Fisher z scores were computed to test for significant differences between these correlations. Hypothesis 2a, that there would be a significant relationship between conscientiousness and crew performance for the additive task, was supported \((r = .42)\). Hypothesis 2b, that there would be a significant relationship between conscientiousness and crew performance for the disjunctive task measure, was also supported \((r = .50)\). Finally, Hypothesis 2c, that there would not be a significant relationship between conscientiousness and crew performance for the conjunctive task measure, was also supported \((r = .35, \text{n.s.})\). After Fisher z score transformations for the three correlations were analyzed, the results indicated that there were no significant differences between the correlations. Thus, full support for Hypothesis 2 was not found.

**DISCUSSION**

The focus of this study was to determine if a team referent measure of conscientiousness contributed unique variance to the prediction of team performance and if this relationship held across different task types. Researchers generally examine predictor–performance relationships at one level in isolation. Asking respondents to answer context-specific measures using their team as the frame of reference increased predictive validity in this sample. In this study, the TCI (team referent measure) was significantly correlated with overall team performance \((r = .39)\), but the SCS (general measure) was not correlated with overall team performance \((r = .26, \text{n.s.})\). This finding is supported by research that shows that context specificity may increase predictive validity when using personality measures (Robie et al., 2000; Schmit et al., 1995). However, the team referent measure did not significantly add to the prediction of overall team performance above the individual measure.
Limited support was found for Hypothesis 1. The TCI was more strongly correlated with overall team performance than was the aggregated individual-level conscientiousness measure. However, regression analysis indicated that the TCI did not account for significant variance in overall team performance above that of the SCS. Further analysis indicated that for the disjunctive task, the TCI accounted for significant variance above that of the SCS. Team conscientiousness predicted disjunctive tasks greater than the additive and conjunctive tasks.

Support for Hypothesis 2 was not found in that the type of task did not fully moderate the relationship between conscientiousness and overall crew performance. Although direct support was not found for the moderator relationship of task type, each of the proposed relationships in Hypotheses 2a, 2b, and 2c indirectly supported the notion of a task moderator. Although it could not be concluded that the task type moderated the conscientiousness/crew-performance relationship, conscientiousness was significantly correlated with both the additive and disjunctive tasks as hypothesized in Hypotheses 2a and 2b and not significantly correlated with the conjunctive tasks as Hypothesis 2c predicted.

McClelland and Judd (1993) discussed the difficulties in detecting moderator effects in field studies and cited four explanations for this difficulty. First, overall model error is more frequent in field studies. Second, considering this potential greater risk for measurement error in field studies, error is grossly increased as variables are multiplied together to form interaction terms. Third, theoretical constraints on the nature of the interaction in field studies restricts the magnitude of the moderator regression coefficient. Finally, the nonlinearity of the variables and the functional form of the interaction are known to cause difficulties in detecting moderator effects.

Another reason that strong support for the moderated relationship was not found may be additional moderating factors. For example, the size of the teams has been found to moderate the predictive validity of individual-level and team referent measures (Devine & Philips, 2001). Individual scores have a larger effect in predicting performance for small teams than do team-level scores.
In this study, flight teams were made up of three individuals. In a larger team, team referent measures may have demonstrated significant incremental prediction over individual measures. In addition, large-sample research should be conducted that examines these task-type relationships across all Big Five constructs. This study examined conscientiousness, and the predicted pattern of relationships received limited support. However, that pattern may not hold for other constructs. For example, although conscientiousness did not predict performance for conjunctive tasks, neuroticism might.

Although the results of this study did not demonstrate unequivocal support for team referent measures, research should continue to not only focus on personality traits at different levels of analysis but also take into account the respondent’s frame of reference. General measures cannot strongly predict specific behaviors (Ajzen & Fishbein, 1977). According to Ajzen and Fishbein (1980), when trying to predict a specific criterion, a specific predictor that corresponds to that criterion should be used. Therefore, to predict team performance, a team referent predictor is the appropriate choice.

As indicated by the significant difference found between the TCI and the SCS scales, individuals reported different levels of conscientiousness when asked to answer in the differing context of individual level versus team level (only 25% of the variance in team and individual measures of conscientiousness was common). Furthermore, the TCI scale was significantly correlated with overall performance and the SCS was not significantly related to overall performance. When team-level criteria were used, predictors at the team level of analysis were more appropriate and demonstrated slightly higher predictive validities.

Limitations to this study include the number of teams and the low number of items making up the crew performance appraisal. A larger sample size may increase the opportunity to detect moderator effects as well as provide more stable estimates of the relationships between conscientiousness and team performance. A performance appraisal made up of a greater number of items would provide more variability in the criterion data.

Another limitation to this study is that the captain of each crew provided the performance appraisal ratings. This may introduce
unwanted bias in the ratings. To safeguard against this, no form of identification was collected and all participants were assured that all data collected would remain anonymous. In addition, we conducted a post hoc analysis that examined the relationship between the captain’s TCI scores and the crew performance appraisal, also completed by the captain. The correlation between these two measures was \( r = .47, p < .01 \). A large magnitude correlation between these two measures would confirm the suspicion of bias in the data; however, the observed correlation suggests that the measures only share 22% variance. This correlation provides evidence that the two measures are sufficiently distinct. However, future research should collect performance appraisal ratings from an outside source versus a participant in the study.

Finally, the issue of within-group agreement must be addressed. In multilevel organizational research, rater agreement is assessed by examining rater response consistency or reliability. Two forms of the intraclass correlation coefficient (ICC) are used most often (James, 1982). The ICC (1) is interpreted as the proportion of total variance that can be explained by group membership. The ICC (2) reflects the reliability of the group mean. Both forms of the ICC are calculated from a one-way ANOVA. When the ICC (1) is large, there is high agreement among raters. In this case, a single individual rating will provide a reliable estimate of the group mean. If the ICC (1) is small, multiple ratings are necessary to provide reliable estimates of the group mean (James, 1982; Kirkman et al., 2001).

In this study, the between-team effect based on the one-way ANOVA was nonsignificant for both the TCI measure \( (F = .963, p = .532) \) and the SCS measure \( (F = .823, p = .713) \). In other words, a significant amount of variance in individual responses to both scales cannot be accounted for by team membership in this sample. The low ICCs (1) obtained from this sample suggest that averaging the multiple ratings for each team provides a more reliable group estimate than taking a sole member’s responses alone. ICCs (1) were computed to estimate between-group variability and within-group variability. An ICC (1) of −.06 was found for the SCS measure and an ICC (1) of −.01 was found for the TCI measure. The closer an ICC (1) comes to −1, the less between-group variability in
the data. Although there was little difference between the ICCs (1) for the TCI and SCS measures, the differences were in a direction that would be expected. The ICC (1) for the TCI measure, as compared with the SCS measure, produced greater between-group variance. Klein et al. (2001) found that group referent items produced greater between-group variability than individual referent items.

The very nature of a pilot’s job may have had a major effect on these findings with regard to group agreement. These findings suggest that very little variability exists between groups. A pilot operating an aircraft may have very little opportunity to exhibit non-conscientious behaviors, thus resulting in all pilots’ displaying high levels of conscientiousness. This naturally occurring “restriction of range” would make it very difficult to identify significant between-group variability.

This study has important implications not only to teams in general but specifically to cockpit crews and aviation research. Helmreich (1996) suggested that effective aviators need strong interpersonal skills in addition to technical expertise, aptitude, and training. The need for interpersonal skills may extend well beyond the cockpit to include the cabin crew, ramp crew, air traffic control, and maintenance. Whereas the increased use of automated equipment continues to rise in the aviation industry, interpersonal skills become highly relevant because the pilot’s level of responsibility for human lives is extremely high. In fact, aviation research has identified a common thread of human error among cockpit failures. A study by Goeters (1995) showed that nearly half the problem cases in a 200-pilot sample were related to interpersonal factors and not technical performance.

Personality measures have not been shown to be stable predictors of crew performance in previous aviation research (Martinussen, 1996). A study by Martinussen and Torjussen (1998) found that the best predictors of pilot performance were instrument comprehension, mechanical principles, and aviation information. In fact, only one of the two personality measures was found to predict performance and only on a Scandinavian sample (Martinussen & Torjussen, 1998). Aviation research should continue to examine the relationships between personality traits, specifically conscientiousness.
Conscientiousness and crew performance, as this study provided support for this relationship with a pilot sample.

This research also has important implications for organizations. In an organization using teamwork, results suggest that the type of task may indicate how teams should be created. If the task is additive or disjunctive, it may be important to test for conscientiousness and create a team of highly conscientious individuals. Likewise, if it is a conjunctive task, the team personality composition may not be such a crucial issue. Second, as demonstrated with airline pilots, many jobs will include various combinations of each task type. In this case, it is not feasible to match employees to task type, thereby reflecting a training issue. If conscientiousness predicts performance of a particular task type, training may need to focus on increasing conscientiousness-like behaviors for individuals low on this particular trait.

Finally, this research contributes to the existing literature concerning using context-specific measures to increase predictive validity (Robie et al., 2000; Schmit et al., 1995) and between-group variability, thereby decreasing within-group variability (Klein et al., 2001). Campion, Medsker, and Higgs (1993) suggested that team-level phenomena be measured by consensus ratings. Consensus ratings require all team members as a group to rate their team on the variable of interest. These team consensus ratings have recently been found to explain significantly greater variance in team effectiveness than aggregation methods (Kirkman et al., 2001). Obtaining team consensus ratings may be very time consuming and difficult to obtain in a practical setting. When consensus ratings are not an option, it is suggested here that the greatest utility will be found using a team referent measure versus aggregating general measures.

REFERENCES


Andrew English, Ph.D., graduated from Florida Institute of Technology and currently works for ThoughtLink, Inc. His research interests are personality measurement, team processes, and the use of simulation and games for training and education.

Richard L. Griffith, Ph.D., is the director of the Industrial Organizational Psychology program at Florida Institute of Technology. His research interests are in the areas of personality, team processes, and chaos theory in organizations.

Lisa A. Steelman, Ph.D., is an assistant professor of psychology at Florida Institute of Technology. Her research interests include feedback, performance appraisal, work-related attitudes, and group dynamics.