Fall 2002

Predicting Team Performance Over Time: A Secondary Data Analysis

Hope S. Hanner
Old Dominion University

PREDICTING TEAM PERFORMANCE OVER TIME: A SECONDARY DATA ANALYSIS

by

Hope S. Hanner
B.A.. May 1996, Queens College

A Thesis submitted to the Faculty of Old Dominion University
in Partial Fulfillment of the Requirement for the Degree of

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2002

Approved by:

Robert M. McIntyre (Director)

Glynn D. Coates (Member)

Bryan E. Porter (Member)

ABSTRACT

PREDICTING TEAM PERFORMANCE OVER TIME: A SECONDARY DATA ANALYSIS

Hope S. Hanner
Old Dominion University
Director: Dr. Robert M. McIntyre

Teams occupy a strong presence in the modern workplace. However, few studies have tracked the evolution of team behavior over time. The use of cross-sectional team research fails to appreciate the stages that teams pass through as they evolve and mature. The present study had three goals in mind. The first goal was to arrive at a causal model of team performance via a secondary statistical analysis on the research conducted by Cunningham (2001) and Strobel (2001). Strobel and Cunningham tracked the performance of eleven student-learning teams over the course of a semester. A team training intervention was administered after seven weeks. Contrary to their hypotheses, team performance declined post-intervention. Therefore, the second goal of the present study was to account for this decrement in team performance. The third goal was to examine the effect of the previously unresearched variable of task workload on the performance of these eleven teams. Results indicated that the same predictors did not come into play with all teams and thus, inconsistency existed across teams. Discussion focused on the implications of the results as well as the limitations and strengths of conducting time-based research.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The overall goal of the present proposed study was to arrive at a causal model of team performance. This was accomplished through several different ways. First, a secondary analysis of the data collected by Strobel (2001) and Cunningham (2001) was performed. Second, multiple-predictor models for team performance were examined rather than the univariate models used in Strobel (2001) and Cunningham (2001). Third, the present research examined data previously not analyzed and their effect on team performance. The new variables that were considered were: perceived task workload, intellectual composition, socio-demographic diversity, and diversity of college majors of team members. Predictors in combination were examined so that more sophisticated causal (time series) models could be developed.

The definition of team adopted for this research is as follows: "A distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common goal, objective, or mission, who have each been assigned specific roles or functions to perform and who have a limited-life span of membership" (Salas, Dickinson, Converse, & Tannenbaum, 1992, p. 4). The concept of team performance has become increasingly popular over the past decade. In addition, teams have become a marked trend in the workplace itself. In fact, it is becoming more and more common for employees and managers alike to work within the context of teams. According to Muchinsky (2003), teams are often viewed as "the organizing principle through which work is accomplished" (p. 258). As the workplace increases in complexity, the ability of one individual to manage his or her tasks becomes much more of a challenge. Teams in work settings allow us to deal with the complexity of modern organizations and the challenges that they pose (Bowen & Jackson, 1986). The increased use of teams has generated a great amount of research in the field of team dynamics, which examines a variety of facets of teams, including theories of team development, the interrelations of teams with others at the individual, group, and organizational levels, as well as the very nature of teams themselves (Wildmeyer, Brawley & Carron, 1986).

The present research attempted to contribute to the team dynamics literature through a secondary analysis of the Strobel (2001) and Cunningham (2001) studies. Strobel and Cunningham conducted longitudinal analyses using student-learning teams in a university classroom setting. Student participants, who were enrolled in an Industrial/Organizational psychology course, formed eleven teams of three-to-five members. All teams completed three intellectual tasks every week, for a total of 16 weeks. No other assignments were completed during these team meetings. Since the course instructor did not provide lectures, the teams were required to rely on one another and a psychology textbook for the completion of their team tasks. During the seventh week, teams received a three-hour training on teamwork processes that was based on the Dickinson-McIntyre model (1997). This model identifies dimensions or components of successful team performance. The seven core team processes are:

1. Communication - Communication is the active exchange of information between team members to clarify information.
2. Team orientation – Team orientation refers to the attitudes of team members toward one another and team tasks. It reflects the self-awareness of each member as a team member, the understanding of the importance of their efforts, and the commitment of the success of the team.

3. Team leadership –Team leadership is not necessarily reserved for a single individual with formal authority but can be possessed by several team members. It refers to the organization, guidance, and direction provided by team members.
4. Monitoring – Monitoring refers to the observation and awareness of the activities and performance of team members. Monitoring suggests that team members are able to provide feedback and backup behavior.
5. Feedback - Feedback occurs when team members provide one another with information about their performance. It also can refer to the seeking and receiving of information among group members regarding performance.
6. Backup behavior - Backup refers to the support that team members give one another in the performance of their tasks. It connotes that members have an understanding of other members' tasks.
7. Coordination – Coordination is the result of the preceding six team processes. Therefore, successful coordination is evidence that other components of teamwork have been achieved.

The authors used interrupted time series analyses (ITSA) to analyze the effects of the team training intervention on their dependent variables. Strobel specifically examined the effects of the training on team performance and team cohesion. Cohesion has been defined as "the total field of forces which act on members to remain in the group" (Festinger, Schachter, & Back, 1950). It can also be thought of as the attraction that members have to their team (Evans & Dion, 1991). Strobel noted that it is common to distinguish between "social cohesion" and "task cohesion." Social cohesion is defined as an interpersonal attraction to the team (Lott & Lott, 1965). Task cohesion, instead, refers to the group affiliation necessary for achieving task-related results (Craig & Kelly, 1999). In order to tap into both aspects of cohesion, Strobel defined team cohesion as an attraction to a team in pursuit of either social affiliation or task-related goals. The same definition will be assumed in the current study.

Strobel used the System for the Multiple Level Observation of Group (SYMLOG) Adjective Rating Form (Bales & Cohen, 1980) to assess cohesion within teams. The SYMLOG is a 26-item self-report, likert-type (0 = never, 1= rarely, 2 = sometimes, 3 = often, and 4 = always) measure. It contains the following three dimensions of team cohesion: Friendly-Unfriendly (P-N; P = positive or friendly, N = negative or unfriendly), Task-Oriented-Emotionally Expressive (F-B; F = forward or instrumentally controlled, B = backward or emotionally expressive), and Dominant-Submissive (U-D; D = dominant or downward, U = upward or submissive) (Strobel, 2001). The reported reliability coefficient for the P-N dimension is .95, for the F-B dimension .80, and for the U-D dimension is .77 (Bales & Cohen, 1979). Results demonstrated that the training intervention significantly and positively affected team cohesion. In fact, seven teams experienced an increase in Task-Oriented-Emotionally Expressive cohesion levels (F-B) during the post-intervention phase. For the Friendly-Unfriendly cohesion levels (P-N), six teams displayed significant increases after the training intervention. For the final dimension, Dominant-Submissive (U-D), five teams demonstrated increased levels of cohesion following the intervention. Clearly, the team training had a positive influence on team cohesion. However, the training did not affect team performance as hypothesized. All teams experienced a decline in performance following the training. This decline was addressed in the present study.

Strobel was interested in studying the forces that attract members to their team. Cunningham (2001), on the other hand, examined social loafing, an obstacle to team growth and performance. Social loafing is defined as a decrease in individual effort due to the social presence of other persons (Latané, Williams, & Harkin, 1979). It is thought to be a very strong phenomenon that may threaten a team, regardless of the personalities of its members (Cohen, 1988). Cunningham also studied the effect of team members' affective state, both positive and negative affect, on team performance. Positive affect is defined as the extent to which one feels enthusiastic, alert, and active. Negative affect, on the other hand, refers to subjective distress or unpleasant engagement. Cunningham used the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988) to assess affective state. This measure consists of two 10-item scales that are rated on a 5-point Likert scale. Participants are given words that describe various emotions and feelings (e.g., enthusiastic, hostile, jittery) and are instructed to choose the appropriate Likert response (1 = very slightly, 2 = a little or not at all, 3 = moderately, 4 = quite a bit, 5 = extremely) that represents the extent to which they have experienced the emotion during the team session. One of the scales represents the positive dimension of affect while the other represents the negative. Internal consistency reliability coefficients for the Positive Affect and Negative Affect scales range from .80 to .90 and .84 to .87, respectively (Watson, Clark, & Tellegen, 1988). Cunningham also used the Workload Sharing (WLS) to assess general feelings of team participation (Campion, Medsker, & Higgs, 1993). This scale contains three Likert-type statements with which team members express their level of agreement (e.g., "Everyone on my team does their fair share of the work.") The data from both measures were used in the present study.

The results of both Strobel and Cunningham demonstrated that team performance declined following the team training intervention. Unlike Strobel's findings, Cunningham's results (of social loafing and affect) were not definitive. Although it was hypothesized that positive affect would be associated with higher team performance and negative affect would be associated with lower performance, four teams demonstrated a negative relationship between affect and performance. Thus, this hypothesis was not supported. Moreover, Cunningham hypothesized that social loafing would decline following the training intervention. Again, this finding was not supported. It was further hypothesized that the training intervention would positively affect the affective state of team members. However, the results provided only partial support for this hypothesis. In other words, half of the teams experienced a positive relationship between the training and team affect while the other half experienced a negative relationship. Finally, Cunningham hypothesized that team performance would improve after the training intervention. However, team performance decreased significantly across all teams.

The present study can be considered "novel" in that it will examine predictors in combination. These predictors include, but are not limited to, task workload, affect, and team cohesion. In addition to these variables, the previously unresearched effects of perceived task workload data will be examined. Although the present author had collected these data, they have not yet been analyzed. Through the use of a time series design, this research attempted to demonstrate the importance of examining teams' behavior over time. It emphasized the dynamic nature of team behavior, which is frequently neglected in research that examines teams statically in a cross-sectional design. The temporal tracking of team performance is an area of study that has received little attention by researchers. In that sense, the present research is a unique contribution to the team literature.

Workload and Team Performance

Task workload, which is defined as the perceived complexity of a task, has been found to be a compelling influence on team performance. Previous research has shown task workload to be more consistently influential than any other variable, including task organization and team training (Naylor & Briggs, 1965). In the case of the present study, task workload is operationalized as the joint completion of multiple college-level psychology in-class assignments. Many researchers have found a link between high task workload and declining team performance (Bowers, Thornton, Braun, & Salas, 1998; Bray, Kerr, Norbert, & Atkin, 1978; Gallwey, & Drury, 1986; Xiao, Hunter, Mackenzie & Jefferies, 1996). In addition, increasing task complexity is often associated with a greater risk of coordination breakdown.

The National Aeronautics and Space Administration Task Load Index (TLX) was used to assess the level of perceived difficulty of the team tasks. The TLX is a subjective workload measure developed by Hart and Staveland (1988), consisting of the following six dimensions to assess workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. This measure will be described in greater detail in the following section.

Team members completed the measures mentioned above during the last five minutes of the team sessions. The majority of the time in each session was spent on the team-building exercises. The exercises were based on readings in the course textbook that included a comprehensive array of topics of Industrial/Organizational Psychology (e.g. job analysis, selection, performance appraisal, workplace diversity, training and development, and so on). Exercises were integrally related to the learning objectives of the course, as indicated by the course syllabus. In other words, all team assignments were intended to be challenging enough so as to require team member participation. Each assignment consisted of a true-false section (in which teams were to correct the false), a three-question short answer section, a long-answer essay section, and an optional extra credit section. A panel of subject matter experts (SMEs) trained on the material assessed the level of difficulty of the team assignments, the "teamness" of the assignments, and the degree to which the assignments were relevant to the course content.

As mentioned earlier, team performance decreased in the period following the training intervention. One of the objectives of the present study is to account for this decline. Although much time and effort were devoted to the development of the assignments, the level of difficulty of the topical areas may have contaminated team performance. By controlling for the difficulty of the assignments and thus, partialing out its effects, the decline in performance may become better understood. It is also possible that team performance may take on an altogether different pattern.

Unanswered Questions

Although the previously mentioned studies offer insight into the growth and maturation of teams over time, some important questions remain unanswered. For instance, it is presumed that the members of a team will vary in their levels of intelligence. However, it is unclear whether this intellectual diversity helps or hampers team performance. Also, it is uncertain whether a difference exists between low versus high intelligence teams. In other words, do teams that have members that are all highly intelligent outperform teams that have members with lower intelligence? The present study aims to investigate this question through the use of GPA scores

While ethnic and gender diversity are extremely popular topical areas in the team literature, the issue of intellectual diversity in teams has not yet been addressed. Nonetheless, the investigation of this type of diversity research may have very interesting implications for both student learning teams and teams in the workplace. Due to the great number of research studies pertaining to ethnic and gender diversity in work teams, supervisors have been advised about the effects of diversity and thus, have become better equipped to deal with heterogeneous teams. It is presumed that knowledge about varying levels of intelligence within a single team will also enable supervisors to better manage team members.

According to Shaw (1976), socio-demographic variables including ethnicity, age, and gender have an important impact of the performance of teams. Therefore, another question that has yet to be answered is: Did demographic diversity help or hinder team performance? Although appropriate demographic data were collected, Cunningham and Strobel did not determine whether degree of ethnic diversity within teams affected the performance of the teams. Likewise, they did not examine whether gender diversity within teams affected performance.

Research on diversity in teams has yielded mixed results. Some studies have indicated that diversity among teams can have beneficial effects including, enhanced creativity (Northcraft et al., 1995) and the ability to produce solutions of higher quality (Watson, Kumar, & Michaelsen, 1993). However, the majority of research indicated that teams with demographically diverse members have detrimental results including, higher turnover (O'Reilly, Caldwell, & Barnett, 1989), lower effectiveness (Fenelon & Megargee, 1971; Tsui & O'Reilly, 1989), lower psychological attraction (Tsui, Egan, O'Reilly, 1992), and lower satisfaction (Jehn, Northcraft, & Neale, 1999). Since there is no definitive prediction for the effect of heterogeneity of team performance, it is more appropriate to treat the final two issues not as hypotheses but as questions to be explored. Therefore, the present study seeks to determine if a difference exists between the performance of heterogeneous (i.e., gender, race, college majors) versus homogeneous teams. Among the variables under investigation in this study (affect, team cohesion and workload), this study examined which predictors, if any, distinguish heterogeneous teams from homogeneous ones.

A final topic that requires attention is task workload. As previously mentioned, the effect of perceived workload on the teams in Strobel and Cunningham has not yet been addressed. The present researcher examined the data that have been gathered within the same teams to determine which workload indices would be significantly related to team performance.

Statistical Analyses

The present study used the autoregressive integrated moving average (ARIMA) model of time series analysis, as described in the following section. Within a time series framework, the following analyses were performed.

1. The transfer function of perceived task workload on team performance. (The concept of transfer functions is generally described in the Methods with more detail in Appendix).
2. A multivariate transfer function analysis of team performance with the following predictors: (a) training, (b) intellectual composition of team members, (c) affect of team members, (d) work load of members, (e) demographic diversity of team members, (f) college majors, and (g) team cohesion.

The following hypotheses were tested:

1. Team performance improves after the difficulty of the topical areas has been partialed out.
2. The five dimensions of workload significantly predict team performance. More specifically, effort and mental demand are positively correlated with team performance. Conversely, frustration, performance and temporal demand are negatively correlated with team performance.
3. A multivariate transfer function analysis provides a better prediction of performance than the univariate models used in Strobel (2001) and Cunningham (2001). More specifically, workload, affective state and team cohesion significantly predict team performance. Positive affect is positively

related to team performance, while negative affect is negatively related. Furthermore, all dimensions of team cohesion are positively related to team performance.

4. Effort, mental demand, and positive affect are positively and significantly related to task cohesion (Forward-Backward dimension). Frustration and negative affect are negatively related to task cohesion.
5. Effort, mental demand, and positive affect are positively and significantly related to social cohesion (Positive-Negative dimension). Frustration and negative affect are negatively related to social cohesion (Positive-Negative dimension).
6. Effort, mental demand, and positive affect are positively and significantly related to social cohesion (Dominant-Submissive dimension). Frustration and negative affect are negatively related to social cohesion (Dominant-Submissive dimension).

Exploratory Questions

7. Do homogeneous teams (i.e., gender, race, college majors) perform better than heterogeneous teams?
8. What significant predictors, if any, distinguish heterogeneous teams from homogeneous teams?

METHOD

Participants in the Cunningham-Strobel Studies

The current study involves several different secondary statistical analyses on the data collected for studies by Cunningham and Strobel. Therefore, it seems useful to describe the sample used in this previous research. The participants in the original research were 45 undergraduate students, enrolled in an introductory Industrial/Organizational psychology course. The participants formed eleven teams to which they were randomly assigned. Teams consisted of three to six members, with no restriction on age or gender. The sample consisted of 36 women (77%) and 9 men (19%). The majority of participants were Caucasian (51%), followed by African American (34%), Asian (8%), and Pacific Islander (7%). Participants had a mean age of 22.27 years (SD = 3.83). Various college majors were represented in the sample including psychology, education, counseling, engineering, and liberal arts. In addition, 67% of the participants had previous experience working as a member of a team in either a class or work setting. The grades on each team assignment, as well as the overall course grade, served as the main incentive for team performance.

Measures Used in the Cunningham-Strobel Studies

Team Assignments: Teams of students met three times a week to complete their team assignments. The assignments were based on the aforementioned I/O psychology textbook. All assignments consisted of five true-false questions, three short essay questions, one long essay question, and one extra credit short essay question (see Appendix C. Teams had 45 minutes in which to work on each assignment. In both the former studies and the present one, grades on these assignments defined team performance. Graduate research assistants graded all assignments according to scoring rules that were established a priori by the course instructor. After research assistants completed the grading, the course instructor reviewed the scoring in an attempt to ensure accuracy of scoring across all teams. No scoring reliabilities were available. However, it was assumed that because scoring rules were applied and checked by the course instructor, the scoring was reasonably reliable (Strobel, 2001).

Other measures used in the previous research. As discussed in the Introduction, the System for the Multiple Level Observation of Group (SYMLOG) Adjective Rating Form was used to measure team cohesion (Bales & Cohen, 1980) in Strobel's research. In accordance with the instructions of the measure, each team member was assigned by a teaching assistant an individual to rate, whom they continued to rate for the duration of the study. As previously stated, team member behavior was evaluated along the dimensions of: Friendly-Unfriendly, Task-Oriented-Emotionally Expressive, and Dominant-Submissive (See Appendix D)

Also described in the previous section, the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1998) was used to assess affective state in Cunningham's research. It is important to note that unlike the SYMLOG, the PANAS is self-report measure designed to assess various positive and negative mood factors that a member feels on a given day. The measure consists of both a positive and negative scale, with ten mood adjectives listed per scale (See Appendix E). The final measure used in the previous research was Workload Sharing (WLS) (Campion, Medsker, & Higgs, 1993). The WLS was given to the participants in order to assess overall feelings of team member participation. More specifically, it was intended to measure how individual team members felt about the division of labor within their own team (See Appendix F).

Procedure in Difficulty of Exercises vs. Team Performance

Recall that the first goal of this study was to determine if the varying level of difficulty of the team assignments in the Cunningham-Strobel research led to the erroneous predictability of performance. Therefore, 44 undergraduate students who had just completed a course in an Industrial/Organizational Psychology course were asked to evaluate the level of difficulty of the content of each of the team assignments used in Cunningham-Strobel research. The evaluations of these students were presumed to be representative of the perceptions of the students who participated in the Cunningham and Strobel studies. The following information and directions were given to the participating students. First, they were informed that the set of assignments represented assignments in a previous course and that students in that previous course worked in teams to complete the assignments. Second, they were informed that prior students (in the Cunningham-Strobel studies) were allowed to use a textbook in order to complete each assignment in approximately 45 minutes. Third, students were asked to evaluate the difficulty of the assignments by reading each and referring to the textbook (*Psychology Applied to Work; Muchinsky, 2000)* that participants had used in Cunningham-Strobel research. The goal here was to provide students with sufficient material to make an educated estimate of the level of difficulty.

Because there were too many assignments for all students to evaluate, a sampling plan was developed whereby each student evaluated the difficulty of nine assignments in three sessions. In this plan, five rows of eight students were formed. This seating position was maintained across all three days of evaluation. Packets were distributed in numerical order, beginning with chapter one. On the second day, the student in the first seat of the first row was given chapter two. Distribution of team assignment packets continued with the next person's receiving chapter three, and so on. With this plan, it was ensured that students would rate different chapter on all three days. To register their reactions to each assignment, students completed a nine-item questionnaire that accompanied each assignment.

Measures

Students completed questionnaires that consisted of nine likert-type (1 = strongly agree, 3 = moderately agree, and 5 = strongly agree) items that were designed to determine the level of difficulty of the assignments (e.g., "It would be fair of the instructor to expect completion of the questions on this assignment by a team in 45 minutes," "The questions on this assignment are appropriate to the material presented in this chapter," "The questions on the assignment are clear for a team to answer") (see Appendix G). Students were given 50 minutes in which to evaluate a packet of three separate assignments. Each packet contained assignments that assessed understanding of material from a single chapter. As was pointed out above, a total of nine assignments were rated over the course of three days.

Statistical Analyses

It should be recalled here that a primary goal of the study was to explain the results of the Cunningham-Strobel studies by testing the hypothesis that the varying level of difficulty of the team tasks may have masked the effect of training and the effect of cohesiveness and social loafing on team performance. In particular, it was of interest to understand the lack of training effect on performance of teams. Therefore, we examined the effect of task difficulty on team performance. By controlling for difficulty, the effect of team training on team performance was re-assessed.

ARIMA was used to analyze the effects of rated level of difficulty on the team performance scores and to determine whether a multivariate transfer function model representing the combined effect of training and rated difficulty level on team performance was viable. In effect, the latter approach provided the basis for partialing out the effect of level of difficulty on team performance, which in effect can be considered a nuisance variable.

Procedure in Workload vs. Performance

As described in the Introduction, the procedure used by Strobel and Cunningham was also used in the current study. In short, eleven teams consisting of undergraduate students who voluntarily participated in a team-based course completed team assignments in I/O psychology over the course of an entire semester. After completing each assignment, all team members completed the NASA-TLX (described in detail below) to assess the level of challenge they believed to exist within the current team assignment. These were data not analyzed in the Strobel and Cunningham research.

Measures

NASA-TLX: The National Aeronautics and Space Administration Task Load Index (TLX) is a subjective workload measure developed by Hart and Staveland (1988). The TLX consists of the following six dimensions in which to assess workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Mental demand assesses the mental and perceptual activity required of a task (e.g., "Was the task easy or demanding?"). As the name implies, physical demand measures the amount of physical activity involved in a task (e.g., "Was the task restful or laborious?"). This dimension was the only one to be omitted in the current study since the task at hand was intellectual and thus, did not require any physical activity. The temporal dimension evaluates the time pressure team members may experience as they work on a task (e.g., "Was the pace slow and leisurely or rapid and frantic?"). The performance dimension measures the level of satisfaction that members feel regarding the accomplishment of goals (e.g., "How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?"). The effort dimension assesses the mental and physical energy exerted to accomplish team goals (e.g., "How hard did you have to work (mentally and physically) to accomplish your level of performance?"). The final dimension, frustration, measures the emotional responses of members as they work through a task (e.g., "How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?") (See Appendix H). For each dimension respondents indicate their perceptions on a one hundred-point, bipolar scale. TLX dimensions are labeled from either *low* to *high* or *good to poor.*

Although the TLX is a widely used measure, there has been little empirical focus on its psychometric properties. Subjective workload measures in general, including the TLX, are not commonly evaluated in terms of their reliability and validity (Gopher & Donchin, 1986). In contrast, the development of most workload measures has been guided by pragmatism, operator acceptance ratings, and face validity. When reliability is addressed in research, test-retest reliability is most commonly reported. The test-retest reliability for the TLX is rather high, with correlations ranging between .83 -.88 (Hart & Staveland, 1988; Scerbo, 2001).

Statistical Analyses

The present study used ARIMA, which is described in some detail in the following section. For each team, the correlation (expressed as a transfer function) between team members' perceived task workload and team performance was examined. Transfer functions represent the time series analysis (TSA)-equivalent of regression models in which the autodependence in the data is controlled. In effect, these correlations are computed by the application of transfer function analysis. They represent the relationships between the workload dimensions and the social loafing, cohesion, and team performance investigated by Strobel and Cunningham. In addition, once again through the application of transfer function analysis, the effect of team training on perceived task workload was assessed.

Participants and Procedure in Multivariate Transfer Function Analysis

The third study is a secondary analysis of Strobel and Cunningham studies. However, in this case, the following changes were made in the analyses: Within a time series paradigm, a multivariate transfer function analysis was performed using the following predictors: (a) training, (b) intellectual composition of team members, (c) affect of team members, (d) workload of members, (e) college majors, and (f) team cohesion.

Statistical software

The use of automated TSA software. AUTOBOX 5.0 from Automatic Forecasting Systems (AFS) will be used as the tool for analyzing all data. AUTOBOX provides an intelligent system for carrying out much of what is described below in the general overview of time series analysis. In other words, it is a tool that examines and re-examines the effects of the auto-dependence among the residuals to refine the parameters in the univariate TSA models and the multivariate transfer function models.

Overview of TSA

As mentioned above, I used ARIMA to test the hypotheses. For a detailed description of the technical details of the analyses used in this study, see Appendix A. In this part of the Method section, a conceptual overview of ARIMA is presented. For each team, the relationship between the key dependent variable apropos of the hypothesis and the hypothesis-relevant predictor(s) were assessed through transfer function analysis. A transfer function is the time-series-analysis (TSA) equivalent of a regression model. Simply put, it is a linear-regression equation in which the autodependence in the dependent and predictor variables is controlled for. In linear regression, the regression coefficients are examined to determine whether a predictor significantly predicts a dependent variable. In transfer function analysis, there are two types of prediction coefficients, similar to regression coefficients. An omega coefficient applies to different lagged values of the predictors. As such, an omega can be interpreted as the partialed change in the dependent variable per unit change in the predictor variable at different time lags. A delta coefficient applies to different lagged values of the **dependent** variable. There is no straightforward linear-regression analogue to the delta coefficient. In point of fact, a delta coefficient also expresses the relationship between the predictor(s) and the dependent variable—but indirectly through different lagged values of the dependent variable. By examining the standard errors of the omega and delta coefficients, and computing a Z statistic (sometimes referred to as T), one can test hypotheses concerning the relationship between the focal dependent variable and the predictor(s). Mathematically, the delta coefficients serve as a more elegant and parsimonious way of describing relationships (Wei, 1990).

Recall that the goal is to identify a reasonable causal model that represents the causal effects of all exogenous variables on the endogenous variables. To this end, a multivariate transfer function analysis was performed. The following exogenous variables were examined with regard to their effect on performance and cohesion: (a) team training, (b) intellectual composition of team members, (c) affect of team members, and (d) workload of members. Finally, the study examined the effect of diversity of demographic characteristics, intellectual achievement, and college major within teams on the predictability of the endogenous variable.

RESULTS

Summary of data can be found in Tables 2-30, which are located in Appendix A.

Effect of Difficulty Level on Team Performance

The objective in the first set of analyses was to determine the effect of the difficulty level of the team assignments on team performance. ARIMA was used to assess this relationship. Tables 2 through 4 contain the results of the ARIMA analyses (referred to as transfer function analyses). Transfer functions can be thought of as regression equations that take into account the time dependence of the response variable and the input variables. Of interest in transfer functions are two parameters: (1) the omega estimates which in effect are regression coefficients assessing the coincident and lagged direct effects of the independent variable(s) on the outcome variable; and (2) the delta estimates which are regression coefficients assessing the indirect effects of the independent variable(s) on the outcome variable. Appendix A provides a slightly more detailed description of the transfer function concept.

In two teams, statistically significant relationships existed between difficulty and team performance: Team 4, (ω (lag 0) = 4.09, $p < .05$) (see Table 2) and Team 9, (δ (lag 1) = 0.88, $p < .01$) (see Table 3). This means that there was a direct positive relationship between difficulty and performance for Team 4. However, for Team 9, the same relationship was indirect and expressed in terms of effects of current values of Y on later values of Y. Although the remaining nine teams did not demonstrate significant results, their findings can be viewed in Table 4. These results were not considered compelling enough to use difficulty as a covariate of team performance. Therefore, the difficulty variable is dropped from further analyses.

Effect of Workload on Team Performance

The results of the statistical analyses of the relationship between the mean of each team's workload scores and team performance are summarized in Tables 5 through 15, one table for each team. There are five indices comprising the TLX: effort, frustration, mental demand, performance, and temporal demand. It should be noted that the relationships between three of the team's five TLX indices and the team performance measures were hypothesized to be positive based on the scaling of the dimensions. These dimensions are effort, mental demand, and performance. Conversely, frustration and temporal demand dimensions were hypothesized to be negatively related to team performance. In other words, high mean levels of frustration within a team were expected to be negatively related to team performance. Similarly, the relationship between mean levels of temporal demand and team performance were hypothesized to be negative. In order to report the results clearly, concisely, and correctly, only relationships in the hypothesized direction are reported in the text. "Relationships" in the "opposite direction" are indicated in the tables only.

For Team 1 (Table 5), there was no support for the hypothesis of a statistically significant relationship between the mean performance index ($\omega = -1.05, p < .05$) and team performance. Similar results were discovered for Team 2 (Table 6) and Team 3 (Table 7). For Team 4 (Table 8), transfer function analyses indicated that the mean mental demand TLX index was related to performance as expected ($\omega = 2.26, p < .05$). In addition, the mean frustration TLX index ($\omega = -1.00, p < .05$) was related to team performance in the hypothesized direction. For Team 5 (Table 9), transfer function analysis indicated that the mean frustration TLX dimension was significantly related to performance ($\omega = 1.91, p < .05$). As hypothesized, frustration was negatively related to team performance. Transfer function analysis did not support the existence of hypothesized relationships between the TLX dimensions and performance in teams 6 through 9 (see Tables 10 – 13). For Team 10 (Table 14), transfer function analysis indicated that the relationship between mean mental demand index and team performance was supported ($\omega = 2.08, p < .05$). Finally, for Team 11 (Table 15), the TLX mean temporal dimension was significantly related to team performance ($\omega = -1.46, p < .05$). The latter finding supported the hypothesis and demonstrated that diminished team performance was associated with feelings of increased time pressures.

Multivariate Transfer Function Analyses

Eleven teams were examined with regard to the effect of team cohesion, task workload, mean level of positive and negative affective state, and range of positive and negative affect within teams on performance. Team cohesion scores were derived from the SYMLOG Adjective Rating Form, which is a 26 item self-report measure. The SYMLOG Adjective Rating Form was used to measure the evaluations that team members make of each other's behaviors following their 50-minute interaction period. Individual evaluation scores were aggregated to the team level. Due to either improper completion of measures or more than two changes made to team composition, Cunningham dropped Teams 7, 9, and 11 from her analysis. Consequently, the affective measures for these three teams were unavailable in the current research. The results of the following analyses can be found in Tables 16 through 26.

For Team 1 (Table 16), transfer function analysis provided no support for the hypothesized relationships. For Team 2 (Table 17), the analyses identified three significant predictors of team performance: the two workload measures of effort (ω (lag 0) = 3.26, $p < .01$) and frustration (ω (lag 0) = -0.86, $p < .01$), and the mean negative affect level (ω (lag 0) = -4.67, $p < .05$). Therefore, as hypothesized. a positive relationship existed between the amount of effort exerted and team performance, while a negative relationship existed between the level of experienced frustration and team performance. In addition, as hypothesized, mean level of negative affect was **negatively** related to team performance. For Team 3 (Table 18), only one variable was predictive of performance as hypothesized-- the **range** of positive affect significantly affected team performance as hypothesized (ω (lag 0) = 0.84, $p < .01$).

For Team 4 (Table 19) and 5 (Table 20), no hypothesized relationships were supported. For Team 6 (Table 21), mean negative affect (ω (lag 0) = -8.49, $p < .05$) was predictive of team performance. The hypothesized predictive effects of one affective measure—mean positive affect—was supported for Team 8 (Table 23) with no other hypothesized relationships supported. Mean positive affect (ω (lag 0) = 1.04, $p < .05$) was significantly predictive of team performance. For the final team, Team 10 (Table 25), the workload measure of mental demand (ω (lag 0) = 2.72, $p < .05$) was significantly related to team performance as hypothesized. In addition, the Dominant-Submissive cohesion dimension was a significant correlate of team performance (ω (lag 0) = 1.38, $p < .05$). As predicted, there was a positive relationship between cohesion and team performance.

Recall that Cunningham excluded Teams 7, 9, and 11 from her original analyses. Therefore, the multivariate transfer function analyses for these teams did not include affective measures. For Teams 7, 9 and 11 (Tables 22, 24 and 26), there were no statistically significant correlates of team performance.

Task and Social Cohesion

The following thirty-three analyses were conducted to determine the significant correlates of both task and social cohesion. The predictors that were used in the multipredictor transfer functions were the workload measures of effort, frustration, and mental demand and four affective measures: mean positive affect, range of positive affect in a team, mean negative affect, and range of negative affect in a team. Only three TLX dimensions were selected because they were the most frequently occurring predictors in the study of the relationship between workload and team performance. Therefore, they had the greatest impact on performance across all teams.

Task Cohesion (F-B)

The results of the task cohesion portion of the current study are summarized in Table 27. For Team 1, the range of positive affect (ω (lag 0) = 0.15, $p < .05$) was the single predictor of task cohesion. As hypothesized, higher levels of positive affect in a team were positively related to task cohesion. For Team 2, no significant relationships were observed. For Team 3, positive affective state (ω (lag 0) = 0.20, $p < .01$) was related to task cohesion in the hypothesized direction. In addition, the TLX frustration measure was related to task cohesion in the hypothesized direction (ω (lag 0) = -0.42, $p < .01$). Analysis of Team 4 data revealed no significant correlates of task cohesion. For Team 5, the TLX frustration index was significantly related to task cohesion in the expected positive direction (ω (lag 0) = -0.29, $p < .01$). For Teams 6 through 8, no significant correlates were discovered. For Team 9, frustration (ω (lag 0) = -0.24, $p <$

.05) was negatively related to team cohesion, thus supporting the hypothesis. Analyses of Teams 10 and 11 data yielded no evidence of significant predictors of task cohesion.

Social Cohesion (P-N)

The P-N or Positive-Negative dimension is one of the two markers of social cohesion within the SYMLOG measurement system. Results of this dimension are summarized in Table 28. For Team 1, effort was significantly related to social cohesion in the hypothesized direction (ω (lag 0) = 0.53, $p < .05$). For Team 2, effort was the sole predictor of social cohesion (ω (lag 0) = 0.42, $p < .05$). The positive relationship between effort and social cohesion lends support to the hypothesis. For Team 3, there were no significant correlates of social cohesion. For Team 4, negative affective state (ω (lag 0) = -0.32, $p < .05$) was significantly related to social cohesion in the hypothesized direction. For Team 5, the TLX mental demand measure was significantly and positively related to social cohesion (ω (lag 0) = 0.66, $p < .05$). In addition, mean negative affect was significantly related to social cohesion (ω (lag 0) = -1.55, $p < .05$). For Team 6, mean positive affect (ω (lag 0) = 0.27, $p < .05$) was significantly related to social cohesion, thus supporting the hypothesis. For Team 7, analyses did not yield evidence of significant predictors of social cohesion. For Team 8, mean positive affect was significantly related to social cohesion (ω (lag 0) = 0.51, $p < .01$) in the hypothesized direction. For the final three teams, Team 9-11, there were no significant correlates of social cohesion.

Social Cohesion U-D

The U-D or Dominant-Submissive SYMLOG dimension is the other marker of social cohesion. Results of this dimension are summarized in Table 29. In Team 1, range of negative affect within a team significantly predicted social cohesion (ω (lag 0) = -0.58, $p < .05$). Range of affect was negatively related to social cohesion, thus confirming the hypothesized relationship. For Team 2, mental demand was positively and significantly related to social cohesion (ω (lag 0) = 0.92, $p < .01$). Thus, the hypothesis was supported. For Team 3, results failed to support the existence of correlates of the dependent variable. For Team 4, range of positive affect in a team predicted social cohesion (ω (lag 0) = 0.12, $p < .05$). For Team 5, mental demand positively predicted social cohesion (ω (lag 0) = 0.88, $p < .05$). Thus, the hypothesis was confirmed. There were no significant correlates for Team 6. For Team 7, effort positively and significantly predicted social cohesion (ω (lag 0) = 0.75, $p < .05$). For Team 8, results did not support the existence of any hypothesized predictor. For Teams 9, 10, and 11, no significant predictors were found.

There were many anomalous findings among the analyses that were considered. A recurring anomaly is the existence of apparently significant relationships "in the wrong direction." Technically, these cannot be used as evidence of some unexpected phenomenon. Yet, it seems useful to incorporate all of the "in-the-wrong-direction effects" into one table to help identify any hidden patterns within the results. For this reason, Table 30 is presented to indicate for each team, which predictors were found to have apparently significant relationships in the direction other than hypothesized.

Demographics

Table 1 is comprised of the demographic information from the eleven teams. For gender, a dummy variable was created (males=1 and females =0) and variance was computed. Similarly for race, a dummy variable was created (white=1, non-white=0) and variance was computed. A dummy variable was also created for college major (psychology=1, non-psychology=0) and the percentage of psychology majors was then computed. In the last column of Table 1, the standard deviation of GPA scores was computed across team members. As evident from the table, very little variability exists with regard to any of the demographic variables. Based on low variability of demographic make-up, it seemed inappropriate to examine these data as possible moderators of team performance.

Table 1
Demographics: Teams 1-11

Team	Gender	Race	Major	GPA
1	0.33	0.33	0.33	0.60
2	0.33	0.33	0.33	0.45
3	0.25	0.25	0.00	0.42
4	0.00	0.25	0.25	0.42
5	0.00	0.25	0.25	0.40
6	0.25	0.33	0.50	1.04
7	0.20	0.00	0.40	0.61
8	0.20	0.30	0.00	0.44
9	0.20	0.00	0.60	0.67
10	0.00	0.25	0.25	0.36
11	0.30	0.20	0.20	0.41

DISCUSSION

The Discussion is organized according to the major sets of hypotheses.

Difficulty of Exercises vs. Team Performance

One of the focal questions in this study pertained to why the performance of all eleven teams declined after training. The first hypothesis was formed in an effort to determine whether the varying level of difficulty of the team exercises may have masked the effect of training. The results of the analysis of the relationship between difficulty and team performance provided no support for the hypothesis that performance would improve after the difficulty of the topical areas has been partialed out. Results indicated that for two teams, there was a significant relationship between difficulty and performance. Even with these two teams, there appeared to be no training-driven improvement in the partialed performance scores.

It was somewhat surprising that a relationship between performance and difficulty was found for only two of the teams. There are several possible reasons for this. The evaluations of the difficulty of the exercises made by a second sample of students may not have been consistent with the participants' perceptions. Recall that the sample of evaluators of difficulty were asked to carry out their task to inform the instructor with regard to the value of the team assignments as exercises for future classes. This instruction was intended to increase the evaluators' involvement and commitment in their evaluation. However, their motivation may not have been equivalent to the motivation of those students who had completed the exercises for a grade. In addition, it may have been difficult for these judges to estimate the difficulty level of the assignments without having actually worked through them.

Workload vs. Performance

The results of the analysis of workload and performance lend weak support for the hypothesis that effort and mental demand indices are positively correlated with team performance while frustration, performance, and temporal demand indices are negatively correlated with the same variable. Data on four of the eleven teams demonstrated support for these hypothesized relationships. It should be pointed out that there does not seem to be a reason why the relationship is supported for only certain teams. Because there is a pattern of inconsistency across the teams in several hypotheses, this inconsistency of results is treated separately at the end of the Discussion.

Multivariate Transfer Function Analysis

For the multivariate transfer function analysis, nearly half of the teams displayed significant predictors of team performance. Specifically, five teams demonstrated significant predictors, while six teams did not. Again, there seems to be no apparent reason for the nonsupport.

Prediction of Task Cohesion

The results lend support to the hypothesis that effort, mental demand, and positive affect are positively related to task cohesion, while frustration and negative affect are negatively related to it. There does not appear to be a reason that approximately half of the teams supported this relationship while the remaining half did not.

Prediction of Social Cohesion

In the study of the positive-negative social cohesion dimension, significant predictors were identified for six teams. This is considered support for the hypothesized relationship between P-N social cohesion and effort, mental demand, and positive affect. Once again, there is no evident reason for this partial support.

Prediction of Dominant-Submissive Social Cohesion

Data analysis in five of the eleven teams lends some support for the hypothesis regarding the dominant-submissive (U-D) social cohesion dimension. Recall that this hypothesis stated that effort, mental demand, and positive affect are positively related to U-D social cohesion, while frustration and negative affect are negatively related to it. Five teams displayed significant predictors of U-D social cohesion while the remaining six teams yielded no significant predictors. It cannot be explained why only half of the teams supported this relationship.

Unpredicted and Unusual Results

The most perplexing finding in the present study is that the coefficients for many of the predictors of team performance and cohesion had algebraic signs opposite from those that had been predicted. In fact, a total of twenty-seven predictors were found to be "statistically significant" (had a two-tailed test been used) but whose signs were reversed. There are several reasons for this. First, the signs themselves may be spurious with little interpretive value. This reason coincides with the rigorous procedures that research should follow after positing a directional hypothesis. Specifically, in standard null hypothesis testing, the rejection region falls only on one side of the test statistic distribution. To suggest that an opposite effect might be significant is to abuse the power-increasing one-tailed test. Following this orthodox line of thinking, all apparently significant effects in "the opposite direction" would not even be acknowledged.

A second reason for the apparent opposite effects that were found pertains to the possibility that the hypothesized arose from the fact that the hypothesized were erroneously conceived. Consider the analysis of the relationship between workload and team performance. According to the Yerkes and Dodson Law (Wickens & Hollands, 2000), an intermediate degree of stimulation is more favorable than extreme stimulation in either direction. In other words, stimulation that is either too high or too low may hinder, rather than enhance performance. Moderate levels of arousal will improve performance by allowing the individual to focus on relevant cues, whereas higher levels may be detrimental because relevant cues may also be excluded. Therefore, it might have been a flaw in logic to hypothesize that the TLX indices of effort and mental demand are positively related to performance. The relationship may be more akin to an inverted-U. Optimum levels of mental demand and effort may not be at extreme points. Rather, optimum levels may be intermediate.

The multivariate transfer function analyses also yielded several "in-the-wrong-direction" effects. This was particularly true for negative and positive affective states. Once again, an orthodox treatment of this would be to conclude that the null hypothesis is not rejected. Another way of treating it is that once again, the scientific hypotheses may have been erroneously conceived. Teams that scored high in positive affect, for example, may have not been serious enough about their performance. Conversely, it is conceivable that teams that scored very high in negative affect may have performed well if the negative affectivity served to bond members to their teams. The PANAS only evaluates the various feelings and emotions experienced by individual team members. Therefore, it is unclear whether these emotions were directed at the instructor, the course, the workload, or the other team members.

Several "in-the-wrong-direction" effects were also found with the two social cohesion predictors. Similar to the aforementioned analyses, the hypotheses regarding social cohesion may have been unrealistically stated. To reiterate, it was hypothesized that effort, mental demand, and positive affect would be positively and significantly related to both markers of social cohesion, while frustration and negative affect would be negatively related to social cohesion. It is possible that those teams that scored particularly high in social cohesion may have exerted more effort into the friendships among team members than they did to the team tasks. Therefore, it is not implausible that the performance of highly cohesive teams may be poorer than the performance of teams that receive *moderate* scores of social cohesion. Moreover, it was suggested above that teams might require moderate levels of workload, positive and negative affect, and team cohesion in order to become high-performing units. It was presumed initially that teams high in cohesion and positive affect would outperform teams with lower scores on these measures. However, these findings now suggest that extreme cohesion and affect scores, in either direction, may actually hinder team performance.

Another perspective on the "in-the-wrong-direction" findings concerns the complex problem of suppressor variables. Cohen and Cohen (1983) provide an explanation of suppressor variables as predictors whose presence in the regression model accounts for variance in the dependent variable because of their relationships with the other predictors. Although no source on the topic has been found, it is logical to assert that since transfer function analysis is analogous to multiple regression analysis, the same phenomena may occur. The point is that the algebraic sign of the predictors may be purely related to the nature of the other predictors in the model and not have much importance in and of themselves. Statisticians who accept this point of view place little credence in the interpretation of individual transfer function or regression coefficients. In a sense, these statisticians are "nihilistic" with regard to the "meaning" of prediction systems (David Reilly, personal communication, 2002).

The perspective that I take in this study is that none of the three approaches is completely correct. Specifically, prediction coefficients whose algebraic sign is opposite to that predicted must be treated with care. They should not be over- or underinterpreted. For example, let us assume that the time-based paradigm makes it at least somewhat unique. Under this assumption, it seems foolish to adopt a statistically orthodox view and completely discount as nonsignificant prediction coefficients whose algebraic signs oppose the prediction. On the other hand, it also seems equally foolish to begin to interpret as meaningful prediction coefficients as though a two-tailed null hypothesis had been in effect. Finally, it seems as though ignoring the *possible interpretability* of statistically significant prediction coefficients, in spite of the risk of suppressor variable effects may be a missed opportunity in this exploratory study. Therefore, I have taken the approach to cautiously examine possible future research that might be carried out in the event that the opposite-signed findings are NOT spurious.

In line with this thinking, the following can be asserted. The frequency of the "in-the-wrong-direction" findings may suggest limitations in the study. Because the course instructor did not hold a traditional role in the class, some of the participants may have felt that their role as a student was somewhat ambiguous. Instead of holding the conventional role of a college student, they had become members of a self-managing learning team. Although they had been warned several times about the new roles they were expected to espouse, it may not have made a difference.

In addition, students may have harbored some resentment regarding their simultaneous participation in a college course and psychology experiment. On several occasions, the teaching assistants witnessed both verbal and nonverbal expressions of boredom and even hostility. Several students were frustrated with particular team members that consistently arrived to their session late and/or left prematurely. It is not surprising, therefore, that many students said that they were unhappy with the nature of the team-based course in the instructor's teaching evaluations.

Another possible limitation of the study is that the frequency of the team exercises may have overwhelmed the student-learning teams. Unlike most college courses, the teams in this study in effect received a "test" every time they attended class (i.e., three times per week). This much testing may have overwhelmed the students, resulting in general resentment toward the course in addition to student apathy and carelessness. This point is particularly poignant given the previously unknown fact that the majority of students taking the course were taking it not as an elective within their major course of study but as a means of meeting certain course cluster requirements. Many students throughout the course were not well prepared for the course material and felt little enthusiasm for it.

The fact that the original performance measures themselves may have had several measurement-related problems presents another possible limitation. Due to the large number of team exercises to be graded, grading inconsistency may have occurred. Recall that each exercise was scored first by one of the teaching assistants and second by the course instructor. Both followed a pre-specified key for grading. In spite of these attempts to control for grading inconsistency, it may not have been eliminated completely. In addition to the reliability-related issues, there may have been concerns with the validity of the measures. Although the measures appeared to have face validity, no formal test validation was performed. In spite of a lack of formal validation, it seems reasonable to accept that the exercises were reasonably content valid. That is, the material came from a careful reading of the text, the instructor's guide, and was examined by a second subject matter expert. The real psychometric concern seems to revolve around the fact that the content domain changed from week to week and even class to class. In other words, the subject matter's difficulty and complexity varied. This was something that had not been anticipated at the outset of the project. It was thought that assessing the relative difficulty of each exercise by a different group of students who took the same course may serve as a way of partialing out differential difficulty. This approach not only did not seem to work but also was never anticipated to deal with the variability of the constructs being measured.

A further drawback of the present study pertains to the way in which many of the teams appeared to carry out their team tasks. Several teams seemed to form a pattern of breaking down their exercises into smaller, disjunctive units. Since the exercises were designed to be completed *as a team*, an aggregation of individual efforts would be unlikely to result in high performance. If members became complacent in their "disjunctive teamwork," it is possible that the training intervention may not have been potent enough to change this behavior.

A final limitation is that problems may have existed with the team training intervention itself. For instance, a three-hour team training may have been insufficient to create a significant effect on team performance. Rather, the participants may have benefited from a longer or more spaced-out intervention. In addition, the participants were not able to receive the training with their own team members due to numerous scheduling conflicts. Therefore, participants often did not receive the training with their fellow team members. Training was conducted in settings that were different from those in which they met during their team sessions. Therefore, there may have been a lack of transfer climate for training.

An Overall Issue—The Inconsistency Across Teams

In addition to these limitations, there were many instances of inconsistency among the positive findings across teams. More specifically, the same predictors did not always come into play with all teams. Because traditional science depends on consistency and replication, critics may claim that the present inconsistencies invalidate the results, pointing to a lack of credibility. These critics may be correct. However, on the other hand, it may be that certain entities such as teams may behave differently from others *over time*. In other words, there may be some individual team characteristics that may account for the differences. This perspective indicates that the findings may not be generalizable; but there may be moderators that account for the differences. It is not clear what the moderators are. Our attempts at identifying possible demographic characteristics of team members that serve as moderators seemed to be unsuccessful.

Strengths of the Study

Despite these potential limitations, the study does possess several strengths. This is the first study of its kind to look at team outcome as a performance measure over time within the context of a university setting. Although the use of team projects within university courses has gained immense popularity over the past twenty years (Feichtner & Davis, 1985), the dynamics of teams have yet to be followed over an extended period of time, such as a four-month semester. Team research at the college level is typically cross-sectional in nature. The present study, on the other hand, is unique in that it tracked changes in behavior and perceptions that occur in student teams over several months. Further, it provided insight into the complexities that university students experience when working in teams over time. This study points to a paradigm that might be used, perhaps with modification, to study teams over time. If the psychometric problems can be addressed, the time series paradigm might be a powerful one for examining team performance. The findings therefore can be used to inform college instructors who use teams in teaching that there may be variability in reactions over time. There may also be a need to carefully and regularly monitor the conditions that exists within teams. For instance, it is important to address levels of workload and team cohesion. Is the workload either too high or too low? Are teams becoming too cohesive or are members not bonding enough?

A further strength of the current research is that it addressed some of the questions that Cunningham and Strobel's analyses left unanswered. For instance, it was not clear whether the difficulty level of certain topical areas of I/O psychology played a role in the declining team performance. Because difficulty was not related to performance for most of the teams, it was determined that the varying level of difficulty was *not* a factor in explaining the declining performance scores after the training intervention. In other words, the effect of difficulty on performance was ruled out. Ruling out difficulty might lead to a conclusion that negative motivation levels and lack of personal control on the part of the students (both of which were suggested by student comments) contributed greatly to the nonsuccess of the training.

The use of the ARIMA model of time series analysis may be viewed as another strength of the study. ARIMA is a powerful statistical technique that has not been frequently used in team research. However, the need for flexibility and adaptability in teams requires a developmental approach to team research. ARIMA is considered a very useful tool in that it is able to capture this dynamic nature of teams. According to Kozlowski, Gully, Nason, and Smith (1999), there exists a noticeable absence of theories that incorporate team development and performance. These authors feel strongly that team theory should be developed "with a more dynamic conceptualization of team performance and its compilation" (p. 241). A true understanding of team performance must be capable of viewing performance at various points in time. The ARIMA model allowed this study to examine these dynamics first hand.

An additional strength of the study is the use of multiple-predictor models for the dependent measures. The studies conducted by Cunningham (2001) and Strobel (2001) examined only univariate models. In contrast, the present research examined predictors in combination so that more comprehensive prediction models could be developed.

Furthermore, the use of psychometrically sound instruments including, the SYMLOG, PANAS, and TLX contributed to the strengths of this study. The moderate reliabilities coefficients of the SYMLOG dimensions make the measure an appropriate tool for tracking changes in levels of task and social cohesion (Bales & Cohen, 1979). The high internal consistency reliability coefficients for the positive and negative affect scales of the PANAS indicate that the scales sufficiently capture both mood factors. TLX also has moderate test-retest reliabilities and high ratings of operator acceptance, making it useful in the study of task workload.

Future Team Research with Student Teams

There were three purposes of this study: 1) to conduct a secondary analysis on the Cunningham and Strobel data, 2) to examine the previously unresearched effects of workload, and 3) to arrive at a multivariate transfer functions models of team performance. Future research may benefit from revising the methodology when using student-learning teams. As mentioned previously, taking performance measures as often as three times per week may be overwhelming for team members. Therefore, team members may function more successfully when they are expected to perform either weekly or biweekly. The format of team exercises may need to be altered as well. For example, team members may be more satisfied with their team experience if their assignments are varied. Some assignments may consist of essays and multiple-choice questions, while others could require members to create group presentations. Clearly, the development of assignments would require a great deal of effort and imagination on the part of the instructor. Finally, additional contact with the course instructor may also prove to be beneficial to team performance and team satisfaction. Some of the participants in the present study were discontented with the limited presence of the instructor. In the future, he or she should be readily available to address concerns relating to either the assignments or team functioning

During the course of the study, it appeared that many of the teams created a pattern of dividing the team assignments into individualized, smaller tasks. As a result, assignments became an exercise in disjunctive group work. In future studies, the course instructor may choose to encourage students to work as a team. In addition, the team process training intervention may require revamping. A three-hour session may have been inadequate to impart sufficient knowledge on how to successfully function as a team. Rather than lecture about the seven core processes through an hour-long PowerPoint presentation, the instructor may need to focus on each component for an entire session. In other words, the team training most likely should be modularized. The trainer should be prepared with several role-play exercises for each of the seven core processes. Mere memorization of definitions will prove insufficient for a true understanding of the team components. The goal, rather, should be for an experiential understanding of teamwork in a controlled setting. Future team training can be made more successful through the use of analogies, error-based training, and learner control (Kozlowski, Gully, Nason, & Smith, 1999). It is also likely that the training needs to be longer in duration (e.g., several days long). While the present training occurred for one day during a three-hour session, it is highly recommended that future training take place over several days. Three hours appeared to be an insufficient amount of time to acquire the necessary team knowledge. It would seem that team process training could easily be two to three times as long in duration.

In conclusion, this study is of value to the team research literature due its novel approach to tracking team performance. In a sense, it is a proof of concept that such a study can be performed in a university setting. It provides the basis of a paradigm for team research. The merits and weaknesses of the research were identified and recommendations for future research were suggested. Future researchers also may choose to adopt the ARIMA model, or other methods of time-based research, to investigate team dynamics. The continuation of the approach should elucidate the results of the present study, as well as other questions that remained unanswered in the field of team research.

References

Bales, R. F. & Cohen, S. P. (1979). *Symlog: A System for the Multiple Level Observation of Groups*. New York: Free Press.

Bowen, D. D., & Jackson, C. N. (1986). Curing those "omigod-not-another-group-class." *Organizational Behavior Teaching Review, X(4),* 21-31.

Bowers, C., Thornton, C., Braun, C., Morgan, Jr., B. B., & Salas, E. (1998). Automation, task difficulty, and aircrew performance. *Military Psychology, 10,* 259-274.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology, 36,* 1224-1240.

Campion, M.A., Medsker, G.J. & Higgs, A.C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology, 46*, 823-850.

Cohen, C.J. (1988, October). Social loafing and personality: The effects of individual differences on collective performance. (Doctoral Dissertation, Fordham University). *Dissertation Abstracts International, 49*, 1430B-1431B.

Craig, T. Y., & Kelly, J. R. (1999). Group cohesiveness and creative performance. *Group Dynamics, 3(4),* 243-256.

Cunningham, A. (2001). *The influence of affective state and social loafing on teams over time.* Unpublished master's thesis, Old Dominion University, Virginia.

Dickinson, T. L. & McIntyre, R. M. (1997). *A Conceptual Framework for Teamwork Measurement.* In M. T. Brannick, E. Salas, & C. Prince (Eds.), Team Performance Assessment and Measurement: Theory, Methods, and Applications (pp. 19-43). Mahwah, NJ: Lawrence Erlbaum Associates.

Eggemeier, F.T., & Wilson, G.F. (1991). Subjective and performance-based assessment of workload in multi-task environments. In D.L. Damos (Ed.), *Multiple task performance* (pp. 217-278). London: Taylor and Francis.

Evans, C. R. & Dion, K. L. (1991). Group cohesion and performance a meta-analysis. *Small Group Research*, 22, 175-186.

Fenelon, J. R., & Megargee, E. I. (1971). Influence of race on the manifestation of leadership, *Journal of Applied Psychology, 55,* 353-358.

Feichtner, S. B., & Davis, E. A. (1985). Why some groups fail: A survey of students' experiences with learning groups. *Organizational Behavior Teaching Review, 10,* 58-73.

Festinger, L., Schacter, S., & Back, K. (1950). *Social pressures in informed groups: A study of human factors in housing.* Stanford, CA: Stanford University Press.

Gallwey, T.J., & Drury, C. G. (1986). Task complexity in visual inspection. *Human Factors*, *28,* 595-606.

Glass, G.V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time-series experiments.* Boulder, CO: Colorado Associated University Press.

Gopher, D., Donchin, E. (1986). Workload – An examination of the concept. In K. Boff, L. Kauffman, and J.P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 41.1-41.9. New York: Wiley

Gottman, J.M. (1981). *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge: Cambridge University Press

Hart, S. G., & Staveland, L. (1988). Development of the NASA task load index (TLX): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.) *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Hill, S. G., Iavecchia, H. P., Byers, J. C., & Bittner, A. C. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34(4), 429-439.*

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make us different: A field study of diversity, conflict, and performance. *Administrative Science Quarterly, 44,* 741-763.

Kozlowski, S. W., Gully, S. M., Nason, E. R, & Smith, E. M. Developing Adaptive Teams: A Theory of Compilation and Performance Across Levels and Time. In D. R. Ilgen & E. D. Pulakos (Eds.), *The Changing Nature of Performance: Implications for Staffing, Motivation, and Development* (pp. 240-292). San Francisco, CA: Jossey-Bass.

Latané, B., Williams, K. & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37,* 822-832.

Lott, A. J., & Lott, B. E. (1965). Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological Bulletin, 64,* 259-309.

McCleary, R. & Hay, R. A. (1980). *Applied time series analysis for the social sciences.* Beverly Hills, CA: Sage Publications.

McDowall, D., McCleary, R., Meidinger, E.E. & Hay, R.A. (1980). *Interrupted time series analysis*. Beverly Hills, CA: Sage Publications.

Muchinsky, P. M. (2000). *Psychology applied to work.* (3rd ed). Pacific grove, CA: Brooks/Cole Publishing Company.

Naylor, J. C., & Briggs, G. E. (1965). Team-training effectiveness under various conditions. *Journal of Applied Psychology, 49,* 223-229.

Northcraft, G. B., Polzer, J. T., Neale, M. A., & Kramer, R. M. (1995). Diversity, social identity, and performance: Emergent social dynamics in cross-functional teams. In S. E. Jackson & M. N. Ruderman (Eds.) *Diversity in Work Teams* (pp.1-13). Washington, DC: American Psychological Association.

O'Reilly, C. A., Caldwell, D. F., & Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly, 34,* 21-37.

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

Sawyer, J. E., Latham, W. R., Pritchard, R. D., & Bennett, W. R. (1999). Analysis of workgroup productivity in an applied setting: Application of a time series panel design. *Personnel Psychology, 52,* 927-956.

Scerbo, M.W. (2001). Stress, workload, and boredom in vigilance: A problem and an answer. In P.A. Hancock & P.A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 267-278). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shaw, M. E. (1976). *Group Dynamics: The psychology of small group behavior.* New York: McGraw-Hill.

Strobel, K. (2001). *Longitudinal analysis of cohesion in student work teams: The efficacy* *of teamwork skills training.* Unpublished master's thesis, Old Dominion University, Virginia.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics.* Boston, MA: Allyn and Bacon.

Tsui, A. S., Eagan, T. D., & O'Reilly, C. A. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly, 37,* 549-579.

Tsui, A. S., & O'Reilly, C. A. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal, 32,* 402-423.

Vandaele, W. (1983). *Applied Time Series and Box-Jenkins Models.* New York, NY: Academic Press, Inc.

Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063-1070.

Watson, W. E. Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal, 36,* 590-602.

Wei, W.W. (1990). *Time series analysis: Univariate and multivariate methods.* Redwood City, CA: Addison-Wesley Publishing Company, Inc.

Wickens, C.D. & Hollands, J.G. (2001). *Engineering psychology and human performance.* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Wildmeyer W. N., Brawley, L. R., & Carron, A. V. (1986). *The measurement of cohesion in sport teams: The group environment questionnaire.* London, Ontario: Sports Dynamics.

Xiao, Y., Hunter, W.A., Mackenzie, C. F., & Jefferies, N.J. (1996). Task complexity in emergency medical care and its implications for team coordination. *Human Factors, 38,* 636-645.

APPENDIX A

Data Summary Tables 2-30

Table 2

Team 4: Difficulty Level and Team Performance

Model Component	Lag	Coefficient	Std Error	P Value
Constant		65.0231	8.9266	0.0000
INPUT SERIES X1 Difficulty				
Omega Factor # 1	2	4.0861	1.6216	0.0156

Significance level is $p < .05$

$$Y_t = 65.0230 + [X1_t][(+\ 4.0861B^{*}\ 2)] + a_t$$

Table 3
Team 9: Difficulty Level and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		-0.0123	9.8566	0.9990
AR factor #1	1	0.3926	0.1624	0.0207
	2	-0.3335	0.1362	0.0192
	3	0.1049	0.1551	0.5031
INPUT SERIES X1	Difficulty			
Delta Factor #2	1	0.8818	0.0817	0.0000
Omega Factor # 1	0	2.2686	1.4015	0.1143
	1	0.2314	1.4691	0.8757

Significance level is $p < .05$

Y_t= -.0147+[X1$_t$][(1- .8820B1)]-1[(+ 2.2686- .2310B)]+[(1- .3930B+.3340B*2- .1050B3)]-1a$_t$

Table 4

Teams 1-3, 5-8, 10-11 Difficulty Level and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Team 1:				
Constant		-0.0123	9.8560	0.9990
AR factor #1		10.3926	0.1624	0.0207
INPUT SERIES X1	Difficulty			
Omega Factor # 1		02.2686	1.4015	0.1143
Team 2:				
Constant		-0.0123	9.8566	0.9990
INPUT SERIES X1	Difficulty			
Omega Factor # 1		02.2686	1.4015	0.1143
Team 3:				
Constant		-0.0123	9.8566	0.9990
AR factor #1	1	0.3926	0.1624	0.0207
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.2686	1.4015	0.1143
Team 5:				
Constant		-0.0123	9.8566	0.9990
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.2686	1.4015	0.1143
Team 6:				
Constant		-0.0123	19.8566	0.9990
AR factor #1	1	0.3926	0.1624	0.0207
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.2686	1.4015	0.1143
Team 7:				
Constant		-0.0123	9.85660	0.9990
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.2686	1.4015	0.1143
Team 8:				
Constant		-0.0123	9.8566	0.9990
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.2686	1.4015	0.1143
Team 10:				
Constant		-0.0123	9.8566	0.9990
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.26861	0.4015	0.1143
Team 11:				
Constant		-0.0123	9.8566	0.9990
INPUT SERIES X1	Difficulty			
Omega Factor # 1	0	2.2686	1.4015	0.1143

Significance level is $p < .05$

Table 5
Team 1: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		80.6854	3.3715	0.0000
X1: Team1 Performance				
Omega		-1.0464	0.5633	0.0701

Significance level is p < .05

$Y_t = 80.6854 + a_t$

Table 6

Team 2: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		105.4283	3.7837	0.0000
X1: Team2 Performance				
Omega	0	-5.9161	1.0743	0.0000

Significance level is $p < .05$

$$Y_t = 105.4300 + [X1_t][(-5.9161)] + a_t$$

Table 7
Team 3: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		55.1446	13.2694	0.0002
X1: Team3 Effort				
Omega	0	-0.7117	0.8617	0.4135

Significance level is $p < .05$

$$Y_t = 55.1450 + [X1_t][(-.7120)] + [(1 - .3860B)]-1\ (a_t)$$

Table 8
Team 4: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		60.4555	13.6335	0.0001
X1: Team4 Frustration				
Omega	0	-0.9995	0.4322	0.0257
X2: Team4 Mental Demand				
Omega	0	2.2581	1.005	0.0300

Significance level is $\underline{p} < .05$

$$Y_t = 60.4550 + [X1_t][(-1.0000)] + [X2_t][(+2.2581)] + a_t$$

Table 9
Team 5: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		24.9290	13.1798	0.0658
X1: Team5 Frustration				
Omega	0	-1.6316	0.6332	0.0138
X2: Team5 Temporal Demand				
Omega	0	1.9108	0.7436	0.0140

Significance level is $p < .05$

$$Y_t = 24.9290 + [X1_t][(-1.6316)] + [X2_t][(+1.9108 + 1.4593\,B)] + [(1 - .3530B)] - a_t$$

Table 10
Team 6: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		36.9094	12.1677	0.0041
X1: Team6 Performance				
Omega	0	-0.9072	0.6238	0.1533

Significance level is $p < .05$

$Y_t = 36.9090 + [X1_t][(-.9070)] + [(1 - .6130B)]-1\ a_t$

Table 11
Team 7: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		99.4841	9.1854	0.0000
X1: Team7 Performance				
Omega	0	-2.2603	1.3301	0.0972

Significance level is $p < .05$

$Y_t = 99.4840 + [X1_t][(-2.2603)] + a_t$

Table 12
Team 8: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		90.6582	6.0934	0.0000
X1: Team8 Frustration				
Omega	0	-0.8012	0.6778	0.2437

Significance level is $p < .05$

$$Y_t = 90.6580+[X1_t][(-.8010)] + a_t$$

Table 13
Team 9: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		86.9571	4.0262	0.0000
X1: Team9 Frustration				
Omega	0	0.5237	0.4402	0.2413

Significance level is $p < .05$

$$Y_t = 86.9570 + [X1_t][(+.5240)] + a_t$$

Table 14
Team 10: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		124.9754	9.5119	0.0000
X1: Team10 Effort				
Omega	0	-3.3331	1.0935	0.0044
X2: Team10 Frustration				
Omega	0	1.4073	0.5022	0.0082
X3: Team10 Mental Demand				
Omega	0	2.0845	1.0314	0.0510
X4: Team10 Performance				
Omega	0	-2.2738	0.6904	0.0023

Significance level is $p < .05$

$$Y_t = 124.9800 + [X1_t][(-3.3331)] + [X2_t][(+1.4073)] + [X3_t][(+2.0845 - 1.8079B)] + [X4_t][(-2.2738)] + a_t$$

Table 15

Team 11: Workload and Team Performance

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		103.2761	6.6521	0.0000
X1: Team11 Temporal Demand				
Omega	0	-1.4573	0.5419	0.0106

Significance level is $p < .05$

$$Y_t = 103.2761 + [X1_t][(-1.4573)] + a_t$$

Table 16
Team 1: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		23.4184	34.6122	0.5034
1 Effort				
Omega Factor 1	0	-1.5939	1.1139	0.1619
2 Frustration				
Omega Factor 2	0	0.9456	0.8324	0.2641
3 Mental				
Omega Factor 3	0	-0.1999	1.5807	0.9002
4 Forward-Backward				
Omega Factor 4	0	0.19871	0.0271	0.8478
5 Friendly-Unfriendly				
Omega Factor 5	0	0.0248	0.6451	0.9696
6 Dominant-Submissive				
Omega Factor 6	0	0.5761	0.6296	0.3668
7 Positive Affect				
Omega Factor 7	0	-0.5300	1.2707	0.6793
8 Positive Range				
Omega Factor 80	0	0.5298	0.2930	0.0797
9 Negative Range				
Omega Factor 9	0	-0.4748	0.9192	0.6089
10 Negative Affect				
Omega Factor 10	0	4.2896	1.3370	0.0030

Significance level is $p < .05$

$$Y_t = 23.4180 + [X1_t][(+.5300)] + [X2_t][(-1.5939)] + [X3_t][(+.9460)] + [X4_t][(-.2000)] + [X5_t][(+.1990)] + [X6_t][(+.02500)] + [X7_t][(+.5760)] + [X8_t] + [(.5300)] + [X9_t][(-.4750)] + [(X10_t)][(+4.2896)] + a_t$$

Table 17

Team 2: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant	1	24.8210	30.9530	0.0003
Moving Avg-Factor	1	0.2033	0.1981	0.3127
1 Effort				
Omega Factor 1	0	3.2610	0.7532	0.0001
2 Frustration				
Omega Factor 2	0	-0.8649	0.3721	0.0268
3 Mental				
Omega Factor 3	0	-0.1900	0.7372	0.7983
4 Forward-Backward				
Omega Factor 4	0	-0.2570	0.6530	0.6966
5 Friendly-Unfriendly				
Omega Factor 5	0	-2.8561	0.8101	0.0013
6 Dominant-Submissive				
Omega Factor 6	0	-0.2493	0.3279	0.4528
7 Positive Affect				
Omega Factor 7	0	-0.2474	0.0800	0.0042
8 Positive Range				
Omega Factor 8	0	0.0038	0.2101	0.9856
9 Negative Range				
Omega Factor 9	0	0.7231	1.135	0.5288
10 Negative Affect				
Omega Factor 10	0	-4.6651	2.1913	0.0413

Significance level is $p < .05$

$$Y_t = 25.3780 + [X1_t][(+3.2610)] + [X2_t][(.8650)] + [X3_t]\ 190)] + [X4_t]\ (.2570)]\ [X5_t]\ (2.8561) + [X6_t][(.2490)] + [X7_t][(-.2470)] + [X8_t][(+.0040)] + [X9_t][(+.7230)] + [X10_t[(-4.6651)]\ 1 - .203B - .5930B*2)]a_t$$

Table 18
Team 3: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		82.4889	45.1929	0.0733
Autoregressive Factor	3	-0.1413	0.1538	0.3651
1 Effort				
Omega Factor 1	0	0.7127	1.2906	0.5846
2 Frustration				
Omega Factor 2	0	-0.0141	0.8134	0.9863
3 Mental				
Omega Factor 3	0	-0.2410	1.2766	0.8514
4 Forward-Backward				
Omega Factor 4	0	-0.7612	1.0310	0.4657
5 Friendly-Unfriendly				
Omega Factor 5	0	-0.6231	0.6715	0.3604
6 Dominant-Submissive				
Omega Factor 6	0	-1.6355	0.4997	0.0026
7 Positive Affect				
Omega Factor 7	0	-1.2180	0.5842	0.0452
8 Positive Range				
Omega Factor 8	0	0.8445	0.2861	0.0059
9 Negative Range				
Omega Factor 9	0	-1.0191	1.0887	0.3562
10 Negative Affect				
Omega Factor 10	0	4.3423	3.8737	0.2706

Significance level is $p < .05$

$$Y_t = 72.2760 +[X1_t][(+.7130)]+[X2_t][(-.0140)]+[X3_t][(-.2410)]+[X4_t][(-.7610)]+[X5_t][(-.6230)]+[X6_t][(-1.6355)]+[X7_t][(-1.2178)]+[X8_t][(+.8450)]+[X9_t][(-1.0191)]+[X10_t][(+4.3423)]+[(1+.1410B3)]-1a_t$$

Table 19
Team 4: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		109.5198	21.1681	0.0000
1 Effort				
Omega Factor 1	0	-0.2797	0.9651	0.7738
2 Frustration				
Omega Factor 2	0	-0.4639	0.4542	0.3145
3 Mental				
Omega Factor 3	0	2.1705	1.4349	0.1399
4 Forward-Backward				
Omega Factor 4	0	-1.3082	0.7153	0.0765
5 Friendly-Unfriendly				
Omega Factor 5	0	-0.5153	0.7993	0.5235
6 Dominant-Submissive				
Omega Factor 6	0	-0.3836	0.4218	0.3697
7 Positive Affect				
Omega Factor 7	0	-0.4646	0.2172	0.0399
8 Positive Range				
Omega Factor 8	0	0.0345	0.1836	0.8521
9 Negative Range				
Omega Factor 9	0	0.0129	0.3077	0.9667
10 Negative Affect				
Omega Factor 10	0	-1.2737	0.7342	0.0921

Significance level is $p < .05$

$$Y_t = 109.5200 +[X1_t][(-.2800)]+[X2_t][(-.4640)]+[X3_t][(+2.1705)]+[X4_t][(-1.3082)]+[X5_t][(-.5150)]+[X6_t][(-.3840)]+[X7_t][(-.4650)]+[X8_t][(+.0340)]+[X9_t][(+.01300)]+[X10_t[(-1.2737)]+ a_t$$

Table 20
Team 5: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		12.9424	52.2996	0.8861
1 Effort				
Omega Factor 1	0	2.2270	1.3772	0.1154
2 Frustration				
Omega Factor 2	0	-1.0025	0.7621	0.1974
3 Mental				
Omega Factor 3	0	0.5888	1.2347	0.6366
4 Forward-Backward				
Omega Factor 4	0	1.7301	1.066	0.1142
5 Friendly-Unfriendly				
Omega Factor 5	0	-0.2994	0.7790	0.7031
6 Dominant-Submissive				
Omega Factor 6	0	0.3687	0.4237	0.3905
7 Positive Affect				
Omega Factor 7	0	0.1760	0.5003	0.7272
8 Positive Range				
Omega Factor 8	0	-0.4158	0.2233	0.0715
9 Negative Range				
Omega Factor 9	0	-0.1197	1.0819	0.9126
10 Negative Affect				
Omega Factor 10	0	1.1772	3.8368	0.7609

Significance level is $p < .05$

$$Y_t = 12.9420 + [X1_t][(+2.2270)] + [X2_t][(-1.0025)] + [X3_t][(+.5890)] + [X4_t][(+1.7301)] + [X5_t][(-.2990)] + [X6_t][(+.3690)] + [X7_t][(+.1760)] + [X8_t][(-.4160)] + [X9_t)][(-.1200)] + [(X10_t)[(+1.1772)] + a_t$$

Table 21
Team 6: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		250.1729	36.3371	0.0000
1 Effort				
Omega Factor 1	0	-0.3242	0.6855	0.6393
2 Frustration				
Omega Factor 2	0	0.1244	0.3338	0.7117
3 Mental				
Omega Factor 3	0	-1.2247	1.4441	0.4025
4 Forward-Backward				
Omega Factor 4	0	-0.3181	0.7677	0.6813
5 Friendly-Unfriendly				
Omega Factor 5	0	-0.2163	0.5531	0.6982
6 Dominant-Submissive				
Omega Factor 6	0	-1.6660	0.5627	0.0057
7 Positive Affect				
Omega Factor 7	0	-0.1416	0.4731	0.7666
8 Positive Range				
Omega Factor 8	0	-0.1315	0.2450	0.5950
9 Negative Range				
Omega Factor 9	0	1.3270	1.2760	0.3058
10 Negative Affect				
Omega Factor 10	0	-8.4943	3.5517	0.0226

Significance level is $p < .05$

$$Y_t = 250.1700 + [X1_t][(-.3240)] + [X2_t][(+.1240)] + [X3_t][(-1.2247)] + [X4_t][(-.3180)] + [X5_t][(.216)] + [X6_t][(-1.6659)] + [X7_t][(-.1420)] + [X8_t][(-.1320)] + [X9_t][(+1.3270)] + [X10_t[(-8.4943)] + a_t$$

Table 22
Team 7: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		98.59104	18.3708	0.0000
1 Effort				
Omega Factor 1	0	-0.4513	1.2122	0.7118
2 Frustration				
Omega Factor 2	0	0.4131	0.7848	0.6017
3 Mental				
Omega Factor 3	0	-0.8040	0.9598	0.4076
4 Forward-Backward				
Omega Factor 4	0	-0.2011	0.9079	0.8259
5 Friendly-Unfriendly				
Omega Factor 5	0	-0.2264	0.7389	0.7610
6 Dominant-Submissive				
Omega Factor 6	0	0.1388	0.7340	0.8510

Significance level is $p < .05$

$Y_t =$ 98.5910 +[X1$_t$][(-.4510)]+[X2$_t$)][(+.4130)]+[X3$_t$][(-.8040)]+[X4t][(-.2010)]+[X5$_t$)][(-.2260)]+[X6$_t$][(+.1390)]+ a_t

Table 23
Team 8: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		34.5534	16.1706	0.0404
AR Factor 1	3	0.4351	0.1458	0.0054
1 Effort				
Omega Factor 1	0	0.9548	0.7513	0.2130
2 Frustration				
Omega Factor 2	0	-1.3061	0.7597	0.0952
3 Mental				
Omega Factor 3	0	-0.0870	1.0145	0.9322
4 Forward-Backward				
Omega Factor 4	0	-1.5561	0.9312	0.1045
5 Friendly-Unfriendly				
Omega Factor 5	0	-0.0232	0.4167	0.9560
6 Dominant-Submissive				
Omega Factor 6	0	0.4318	0.6144	0.4873
7 Positive Affect				
Omega Factor 7	0	1.0378	0.4793	0.0380
8 Positive Range				
Omega Factor 8	0	-0.0660	0.1848	0.7236
9 Negative Range				
Omega Factor 9	0	0.7992	0.3339	0.0227
10 Negative Affect				
Omega Factor 10	0	-0.5071	0.9464	0.5958

Significance level is $\underline{p} < .05$

$Y_t =$ 61.1660 +[X1$_t$][(+.9550)]+[X2$_t$][(- 1.3061)]+[X3$_t$][(-.0870)]+[X4$_t$][(- 1.5561)]+[X5t][(-.0230)]+[X6$_t$][(+.4320)]+[X7$_t$][(+ 1.0377)]+[X8$_t$][(-.0660)]+[X9$_t$][(+.7990)]+[X10$_t$)[(-.5070)]+[(1.4350B3)]-1a$_t$

Table 24
Team 9: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		36.5567	17.9075	0.0488
AR Factor 1	1	0.5078	0.1537	0.0022
	2	–0.1555	0.1586	0.3336
	3	0.2618	0.1346	0.0599
1 Effort				
Omega Factor 2	0	-0.2877	0.9742	0.7695
2 Frustration				
Omega Factor 3	0	-0.1058	0.4561	0.8179
3 Forward-Backward				
Omega Factor 4	0	-0.2969	0.5731	0.6076
5 Friendly-Unfriendly				
Omega Factor 5	0	0.1904	0.5413	0.7272
6 Dominant-Submissive				
Omega Factor 6	0	0.2238	0.4497	0.6218

Significance level is $p < .05$

$$Y_t = 94.7530 + [X1_t][(-.2880)] + [X2_t][(-.1060)] + [X3_t][(-.2970)] + [X4_t][(+.1900)] + [X5_t][(+.2240)] + [(1-.5080B1+.1550B2-.2620B3)]-1a_t$$

Table 25

Team 10: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		25.4272	36.0075	0.4859
1 Effort				
Omega Factor 1	0	-2.9849	1.2887	0.0269
2 Frustration				
Omega Factor 2	0	0.3045	0.5690	0.5962
3 Mental				
Omega Factor 3	0	2.720	1.1432	0.0233
4 Forward-Backward				
Omega Factor 4	0	-0.8657	0.9917	0.3890
5 Friendly-Unfriendly				
Omega Factor 5	0	0.3050	0.4536	0.5060
6 Dominant-Submissive				
Omega Factor 6	0	1.3766	0.4157	0.0022
7 Positive Affect				
Omega Factor 7	0	0.0319	0.4445	0.9433
8 Positive Range				
Omega Factor 8	0	0.3343	0.1711	0.0592
9 Negative Range				
Omega Factor 9	0	-1.7887	1.0936	0.1114
10 Negative Affect				
Omega Factor 10	0	2.4485	2.1805	0.2696

Significance level is $p < .05$

$$Y_t = 25.4270 + [X1_t][(-2.9849)] + [X2_t][(+.3050)] + [X3_t][(+2.7201)] + [X4_t][(-.8660)] + [X5(T)][(+.3050)] + [X6_t][(+1.3766)] + [X7_t][(+.0320)] + [X8_t][(+.3340)] + [X9(T)][(-1.7887)] + [X10_t[(+2.4485)] + a_t$$

Table 26
Team 11: Multi-transfer Function Analysis

Model Component	Lag	Coefficient	Std. Error	P Value
Constant		97.2352	13.4368	0.0000
1 Effort				
Omega Factor 1	0	0.9578	1.3918	0.4956
2 Frustration				
Omega Factor 2	0	-0.5671	0.7954	0.4803
3 Mental				
Omega Factor 3	0	-1.7830	1.3571	0.1970
4 Forward-Backward				
Omega Factor 4	0	0.6742	0.7995	0.4045
5 Friendly-Unfriendly				
Omega Factor 5	0	-1.4716	0.9799	0.1416
6 Dominant-Submissive				
Omega Factor 6	0	0.1942	0.4803	0.6884

Significance level is $\underline{p} < .05$

$$Y_t = 97.2350 + [X1_t][(+.9580)] + [X2_t][(-.5670)] + [X3_t][(-1.7830)] + [X4_t][(+.6740)] + [X5_t][(-1.4716)] + [X6_t][(+.1940)] + a_t$$

Table 27
Teams 1-11: Task Cohesion

Model Component	Lag	Coefficient	Std. Error	P Value
Team 1:				
X4: Positive Range				
Omega Factor # 4	0	0.1453	0.0552	0.0125
Team 2:				
X5: Positive Affect				
Omega Factor # 5	0	-0.0406	0.0191	0.0412
Team 3:				
X2: Frustration				
Omega Factor # 2	0	-0.4207	0.1168	0.0009
X5: Positive Affect				
Omega Factor # 5	0	0.2029	0.0717	0.0076
Team 4:				
No significant predictors				
Team 5:				
X1: Effort				
Omega Factor # 1	0	-0.5841	0.1737	0.0018
X2: Frustration				
Omega Factor # 2	0	-0.2873	0.0978	0.0057
X3: Mental Demand				
Omega Factor # 3	0	-0.4703	0.1470	0.0029
Team 6:				
No significant predictors				
Team 7:				
No significant predictors				
Team 8:				
No significant predictors				
Team 9:				
X2: Frustration				
Omega Factor # 2	0	-0.2360	.0875	0.0102
Team 10:				
No significant predictors				
Team 11:				
No significant predictors				

Significance level is $p < .05$

Table 28
Teams 1-11: Social Cohesion (P-N)

Model Component	Lag	Coefficient	Std. Error	P Value
Team 1:				
X1: Effort				
Omega Factor # 1	0	0.5269	0.2475	0.0404
X2: Frustration				
Omega Factor # 2	0	0.4408	0.1679	0.0126
Team 2:				
X1: Effort				
Omega Factor # 1	0	0.4190	0.1861	0.0305
Team 3:				
No significant predictors				
Team 4:				
X6: Negative Range				
Omega Factor # 6	0	0.1205	0.0566	0.0401
X7: Negative Affect				
Omega Factor # 7	0	-0.3232	0.1332	0.0204
Team 5:				
X3: Mental Demand				
Omega Factor # 3	0	0.6596	0.3118	0.0416
X6: Negative Range				
Omega Factor # 6	0	0.4613	0.2177	0.0412
X7: Negative Affect				
Omega Factor # 7	0	-1.5543	0.7530	0.0465
Team 6:				
X5: Positive Affect				
Omega Factor # 5	0	0.2658	0.1030	0.0141
Team 7:				
No significant predictors				
Team 8:				
X5: Positive Affect				
Omega Factor # 5	0	0.5058	0.1400	0.0009
Team 9:				
No significant predictors				
Team 10:				
No significant predictors				
Team 11:				
No significant predictors				

Significance level is $\underline{p} < .05$

Table 29
Teams 1-11: Social Cohesion (U-D)

Model Component	Lag	Coefficient	Std. Error	P Value
Team 1:				
X6: Negative Range				
Omega Factor # 6	0	-0.5762	0.2308	0.0172
X7: Negative Affect				
Omega Factor # 7	0	0.8765	0.4137	0.0411
Team 2:				
X1: Effort				
Omega Factor # 1	0	-0.6342	0.2295	0.0094
X3: Mental Demand				
Omega Factor # 3	0	0.9196	0.2368	0.0005
X5: Positive Affect				
Omega Factor # 5	0	-0.0597	0.0239	0.0178
Team 3:				
No significant predictors				
Team 4:				
X4: Positive Range				
Omega Factor # 4	0	0.1244	0.0539	0.0274
X5: Positive Affect				
Omega Factor # 5	0	-0.2403	0.0535	0.0001
Team 5:				
X3: Mental Demand				
Omega Factor # 3	0	0.8772	0.3691	0.0229
Team 6:				
No significant predictors				
Team 7:				
X1: Effort				
Omega Factor # 1	0	0.7489	0.3681	0.0493
Team 8:				
X5: Positive Affect				
Omega Factor # 5	0	-0.7118	0.1398	0.0000
Team 9:				
No significant predictors				
Team 10:				
No significant predictors				
Team 11:				
No significant predictors				

Significance level is $p < .05$

Table 30
Teams 1-11: "In-the-Wrong-Direction" Effects

Study	Dependent Variable	Predictor	Team
Task Workload	Team Performance	TLX Performance	1
		TLX Performance	2
		TLX Frustration	4
		TLX Temporal	5
		TLX Frustration	10
		TLX Effort	10
		TLX Performance	10
Multivariate	Team Performance	Negative Affect	1
		P-N Cohesion	2
		Positive Affect	2
		U-D Cohesion	3
		Positive Affect	3
		Positive Affect	4
		U-D Cohesion	6
		Range of Negative Affect	8
		TLX Mental Demand	10
Task Cohesion (F-B)	Task Cohesion	Positive Affect	2
		TLX Effort	5
		TLX Mental Demand	5
Social Cohesion (P-N)	Social Cohesion	TLX Frustration	1
		Range of Negative Affect	5
Social Cohesion (U-D)	Social Cohesion	Negative Affect	1
		TLX Effort	2
		Positive Affect	2
		Positive Affect	4
		Positive Affect	8

APPENDIX B

Time Series Overview

A time series is defined as a set of N time-ordered observations of a process. A process is understood to be a mathematically defined function that generates realizations of the process. A realization is one sample generated from a process. The concept of process is roughly analogous to a population distribution while realization is roughly analogous to a sample from the population in traditional cross-sectional research designs (McCleary & Hay, 1980).

Many think of time series designs as substitutes for traditional randomized experimental designs when such designs are not feasible. Glass, Wilson, and Gottman (1975) correct this erroneous position. They explain that the time series designs offer a unique perspective on the evaluation of intervention (or "treatment") effects. They go on to say that the traditional "Fisherian" designs fail to address the fact that interventions (such as training) affect social systems (such as teams) in time over time. The effects of interventions may occur immediately after the intervention is implemented or they may affect the team after some period of time has passed. Further, the effect may take a variety of forms. It may be abrupt and temporary, abrupt and permanent, gradual and temporary, or gradual and permanent. It may show decay in time that cannot in general be captured in the traditional research design. (Glass, Wilson, and Gottman (1975) discuss ten different types of effects that may follow an intervention.) The Interrupted Time Series Experiment (ITSE) is therefore not just a weak fallback position for investigators of teams. It offers a feasible solution to the constraints of group research in field settings (Sawyer, Latham, Pritchard, & Bennett, 1999). In fact, it may be the preferred approach to addressing the dynamic and interdependent nature of team performance.

An ITSE requires the collection of time series data over time. At some point in time, the time series data are "interrupted" by the intervention. Prior to the intervention, data are treated as baseline data. After the intervention, data are treated as the experimental data of interest. To test the hypothesis that an intervention has an impact on the data, an interrupted time series analysis (ITSA) is conducted (Gottman, 1981; Glass, Wilson, & Gottman, 1975; McDowall, McCleary, Meidinger, & Hay, 1980). This analysis allows for the evaluation of an intervention within an ITSE by controlling for the autocorrelation in the data. Autocorrelation implies that there is time dependency within the data—that there is some predictability from the past of a series of data to the current values. The existence of autocorrelation makes it difficult to determine whether an intervention has an impact on the data. That is, when a change in trend appears, autocorrelation is an obstacle to determining whether change following an intervention is the result of the intervention or simply the normal behavior of the (interdependent) series of data (Gottman, 1981).

In the current study, ARIMA model of TSA was followed (Glass, Wilson, & Gottman, 1975; Gottman, 1981; McCleary & Hay, 1980; McDowall, McCleary, Meidenger, & Hay, 1980; Vandaele, 1983; Wei, 1990).

The ARIMA model follows the theory that any time series observation X_t consists of a random error component e_t (or a_t) plus some deterministic component. In this case, e_t is referred to as white noise and is ordinarily assumed to be normally distributed with a mean of zero and variance, σ_e^2. The deterministic component refers to two phenomena in training and intervention evaluation research: The first is the effect or impact of the intervention. The second is the mathematical process that generates the data.

Note that there are several processes considered in the ARIMA model. The first is trend or drift. In point of fact, it is should be noted that trend is technically a deterministic type of behavior while drift is considered random. Unfortunately, it is difficult in social sciences to distinguish between trend and drift. Therefore, here, the phenomenon is treated as deterministic. In addition to trend-drift, two mathematical processes generate a particular time series: auto-regressive and moving average. In an auto-regressive process, prior observations in a time series affect the current observation. In a moving-average process, prior random shocks (that is, random error components) are assumed to affect the current value.

The final deterministic element of a time series observation is attributable to the intervention. It was stated implied above that interventions can take on a variety of dynamic forms (e.g., abrupt permanent change, gradual permanent change, etc.) in time which simpler pre-test—post-test designs may not pick up. Therefore, the exact value of the intervention part of the deterministic component depends on the nature of the intervention effect. ARIMA procedures provide a means of isolating autoregressive (AR), trend-drift, and moving average (MA) aspects of an observation in an ITSE so that the phenomenon of primary interest, the intervention effect, can be examined. That is, ARIMA accounts for the existence of AR, trend-drift, and MA processes, allowing the investigator to analyze the size of the effect attributable to the intervention.

In general, ARIMA models can take on a variety of forms described by three parameters. For this reason, one regularly finds the expression ARIMA (p,d,q), where p is the order of the autoregressive component of the model, d is the order of the trend-drift component of the model, and q is the order of the moving average component. The "orders" can take on values equal to or greater than zero. An ARIMA (1,0,0) means that the model is a pure first order autoregressive model. The ARIMA (0,1,0) means that the model accounts for a first order trend-drift with no autoregressive or moving average tendencies.

Model Identification

Stationary models and trend and drift. It is important to understand trend and drift a bit more precisely. As was pointed out above, unless there is a firm foundation in the literature to guide a researcher's thinking, trend and drift are not easily distinguishable from one another. Another way of expressing the existence of trend-drift is through the concept of a stationary model. A time series is considered stationary in its mean if it neither trends nor drifts. A stationary model in the mean is one for which the parameter d = 0. However, if a series of data appear to trend or drift in the mean, then the data are usually transformed by a process called differencing. Differencing refers to subtracting from a current observation X_t a previous observation X_{t-1}. In other words,

$$z_t = X_t - X_{t-1}$$

Trend or drift can sometimes be discovered by examining the plot of time series data. However, a much better way of carrying out the process is by examining the autocorrelations that underlie the time series data. An autocorrelation is defined as the correlation between pairs of data in the time series separated by k time points (or k seasonal points). This means the a correlation can be computed for pairs of observations

(X_1, X_2), (X_2,X_3), (X_3,X_4), etc. It also means that the separation of k, sometimes called lag, can increase. For example, after computing the correlation for k =1, a correlation would then be computed for k=2 involving the ordered pairs (X_1,X_3), (X_2,X_4), ...,(X_{n-2},X_n). The autocorrelation indicates the degree to which there is dependency within a time series data set. The autocorrelation function (ACF) refers to the series of autocorrelations up to, say 20 lags, for a given time series. The ACF can be plotted and examined to determine whether trend or drift are operative in a given time series. The plot is referred to as a correlogram. When the values in the correlogram "neither damp out or truncate for a given level of d, but instead remain large, then nonstationarity [in the mean] at the level of differencing is indicated" (Glass et al., 1975, page 97).

Identifying p and q. Identifying the level of differencing required in modeling a given time series is the first step in identifying the model. Thereafter, efforts are invested in identifying the degree to which the model is an autoregressive, a moving average, or a mixed model. In addition to using the ACF, another function is examined called the PACF, the partial autocorrelation function. McCleary and Hay (1980) explain the PACF in the following way: "The PACF has an interpretation not unlike that of any other measure of partial correlation. The lag-k PACF, PACF (k), is a measure of correlation between time series observations k units apart after the correlation at intermediate lags has been controlled or 'partialed out'" (p. 75). The computation of the PACF(k) is not as straightforward as that of the ACF (k). It is a complex function of ACF. Fortunately, time series computer programs compute the function values as a matter of course.

All theoretical time series processes have a known pattern of ACF and PACF. Therefore, theoretically, if one examines the ACF and PACF, one should be able to identify the proper values of p and q in an ARIMA (p,d,q)(P,D,Q). (Note that d and D have already been addressed above.) McCleary and Hay (1980) as well as other authors provide detailed guidelines for identifying the values of p and q (and P and Q) on the basis of examining the ACF and PACF. (The different patterns of ACF and PACF for identification of the models will not be described here. The interested reader would do well to review McCleary and Hay (1980).)

Parameter Estimation

An ARIMA model is nonlinear in its parameters which means that ordinary least squares (OLS) procedures, so commonly used in traditional experimental designs, are usually not recommended (McCleary & Hay, 1980; Wei, 1990). Instead, two procedures are recommended. One is referred to as the Exact Likelihood Function. The other more commonly recommended is the nonlinear least squares estimate procedure. This procedure "involves an iterative search technique" (Wei, 1990, p. 144). Wei describes this procedure in the following way.

> "The nonlinear least squares routine starts with initial guess values of the parameters. It monitors these values in the direction of the smaller sum of squares and updates the initial guess values. The iterations continue until some specified convergence criteria area reached" (Wei, 1990, p. 145).

McCleary and Hay (1980) point out that after the parameters are estimated, two concerns arise. First, the estimated autoregressive and moving average parameters should be statistically significant. If a parameter estimate is not statistically significantly different from zero, it is dropped from the model and the model is estimated again. Second, the estimates of the autoregressive and moving-average parameters must lie within the bounds of stationarity (for autoregressive parameters) and invertibility (for moving-average parameters). Stationarity of autoregressive parameters is a mathematical requirement that must be met to retain the autoregressive nature of the model. It states that given a p of some level, the values of p must take on values so that the nature of the autoregressive model is retained.

Invertibility for the moving-average model is similarly defined. It refers to values of q that keep intact the nature of the moving-average model. Recall that the moving-average model dictates that an observation at time t is influenced by previous random error values (random shocks) of previous observations in the time series. Further, the influence the random shocks decreases as the time lag between the present value of t and the previous value of t increases. Fortunately, statisticians have worked out exact values for the range of parameters to satisfy the stationarity and invertibility requirements. McCleary and Hay (1980) point out that for social science data, the order of autoregressive and moving-average models rarely exceeds 2. Therefore, the "rules" for stationarity in autoregressive parameters and invertibility in moving-average parameters are readily available in most texts on time series analysis.

After estimation has taken place and after the requirements of statistical significance and stationarity-invertibility are met, a tentative model has been computed. At this point, the third stage of the time series analysis begins—the diagnosis.

Model Diagnosis

McCleary and Hay (1980) indicate that model diagnosis proceeds as follows. First, model residuals are calculated by computing for each observation the difference between the value of the model implied observation and the actual value. Second, the residuals of the tentative model must be statistically independent at a first and second lag. That is, the following must hold:

$$\mathrm{ACF}(1) = \mathrm{ACF}(2) = 0$$

Third, the residuals must be distributed as white noise. McCleary and Hay (1980) point out that for 20 or 30 lags of an ACF, given a significance level of .05, it would be expected that some of the ACF(k) values would be significant by chance. This third criterion requires that overall, the ACF(k) values are nonsignificant. To test this, the Q statistic can be used:

$$Q = N\sum_{i=1}^{k}[\mathrm{ACF(i)}]^2,$$

where df = k-p-q-P-Q. The Q statistic is distributed approximately as a chi-square with degrees of freedom as indicated. The null hypothesis that the model residuals are white noise is:

$$H_0\text{: } \mathrm{ACF}(1) = \mathrm{ACF}(2) = \ldots = \mathrm{ACF}(k) = 0$$

If the Q statistic takes on a value greater than chance, then the model residuals are presumed to be different from white noise and the model is to be rejected. McCleary and Hay (1980) recommend a value of the number of lags (k) would be 25 due to the influence of k on the power of the Q statistic.

Impact of the Intervention

After the three stages have been completed, usually within the pre-intervention data, the intervention must be examined for its impact. In the case of most social science interventions, researchers have chosen to examine the abrupt-permanent type of impact. This may in part be due to the computer software available to the researchers. It may also be due to the number of data points that have been collected. In the absence of any theoretical reasons, perhaps the abrupt and permanent type of impact is the most reasonable to assess. However, it seems fruitful to recognize that there alternative approaches to investigating impacts. McCleary and Hay (1980) discuss the process of examining these alternatives as "rival hypotheses." These would include the abrupt-permanent, the gradual-permanent, and the abrupt-temporary impact hypotheses.

In all cases, the researcher must identify a transfer function associated with each of these types of impacts. This transfer function in the simplest case—that is, the abrupt permanent impact—requires that parameters in the following function be solved for:

$$f(\mathrm{I}_t) = \omega_0 \mathrm{I}_t,$$

where $\mathrm{I}_t = 0$ prior to the intervention and 1 after the intervention and ω_0 is the level change attributed to the intervention. Here I_t is a dichotomous variable indicating whether the intervention is impacting or not.

For the gradual constant pattern, the parameters in the following function must be solved for:

$$f(\mathrm{I}_t) = \frac{\omega_0}{1 - \delta_1} \mathrm{I}_t,$$

where δ_1 is a parameter that is constrained to the interval,

$$-1 < \delta_1 < +1,$$

and I_t is again a dichotomously valued coding variable indicating whether the intervention has an effect. Here, δ_1 algebraically operates on ω_0 to effectively increase or reduce the magnitude of ω_0 over time depending on the sign of δ_1. The other parameters have already been defined.

For an abrupt temporary impact, the parameters of the following function must be solved for:

$$f(\mathbf{I}_t) = \frac{\omega_0}{\mathbf{1} - \delta_1}\mathbf{I}_t(\mathbf{1} - \mathbf{B})\mathbf{I}_t,$$

where all values have been defined except for B. B is the backward shift operator and is interpreted as "B operates on I by shifting I back one point in time." McCleary and Hay (1980) refer to the following as a pulse function:

$$(\mathbf{1} - \mathbf{B})\mathbf{I}_t,$$

Such that, for example,

$$(\mathbf{1} - \mathbf{B})\mathbf{I}_t = \text{ 0 prior to intervention}$$

$$(\mathbf{1} - \mathbf{B})\mathbf{I}_t = 1 \text{ at the onset of the intervention}$$

$$(\mathbf{1} - \mathbf{B})\mathbf{I}_t = 0 \text{ thereafter.}$$

McCleary and Hay (1980) refer to the abrupt temporary impact as a differenced step function. This implies that the difference transformation is applied to the I_t values in time which are in effect the dichotomous code indicating whether the intervention is operating. Consider the values of I_t

$$\ldots 0, 0, 0, 0, 1, 1, 1, 1, \ldots$$

By applying a first-order differencing transformation to these codes, one gets a pulse function, in effect:

...(0-0), (0-0), (0-0),(1-0),(1-1),(1-1),...

... 0 , 0 , 0, 1 ,0 , 0

Once again, these codes represent a dummy variable indicating whether the intervention is impacting the time series data or not. In this case, the impact is a single "pulse" which abruptly diminished after one impact.

There are other types of interventions that may be examined. The three presented above are perhaps the most commonly observed. The most important thing to understand is that the current statistical software allow for an investigation of the impact of interventions such as team training that provide a much richer understanding of its effects.

Transfer Function Model

In univariate time series analysis, one can look at the ARIMA parameters of a single time series and use these parameters to estimate future values. In effect, the memory structure represented by the ARIMA parameters is the memory structure of a univariate time series. One can also examine the effect of interventions (like training) on univariate time series data. In effect, this type of examination is a bivariate relationship which may become clear as transfer functions are discussed.

It is often useful to examine bivariate (and multivariate) time series variables through examining their relationship. To understand how relationships are examined between time series data, it is necessary to examine some terminology.

Typical cases of transfer function models are the consumption of alcoholic beverages in a given nation, aggregate consumption and advertising, and forecasting sales using advertising expenditures (Vandaele, 1983). Intervention models are special cases of transfer function models that most commonly assess the impact of a discrete intervention or event on a random or stochastic process. It is important to note that a transfer function model and the standard regression model are conceptually related. In fact, both models have a dependent variable and one or more explanatory or predictor variables (Vandaele, 1983). However, there are several reasons why transfer function analysis is considered to be more appropriate for data containing autocorrelation. First, the multiple regression model violates the assumption of independence of errors, thus increasing the Type I error rate. Second, with multiple regression, patterns may obscure or spuriously augment the effect of an intervention unless it is accounted for in the model (Tabachnick & Fidell, 2001).

In the econometric literature, there is the general distributed lag model that says that a current level of Y_t is a function of a number of past values of X_t:

$$Y_t = \nu_0 + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \ldots + e_t$$

where v are the impulse response weights (regression coefficients) and the subscripts of X_i indicate the point in time when the data point is collected. A subscript (t-1) indicates a lag of one time period. Another way of representing this distributed lag equation is as follows:

$$Y_t = \nu(\mathrm{B})X_t + e_t,$$

where

$$\nu(\mathrm{B}) = \nu_0 + \nu_1\mathrm{B} + \nu_2\mathrm{B}^2 + \ldots$$

and

B is the backward shift operator, defined, e.g., as

$$BX_t = X_{t-1}.$$

In transfer function investigation, it is assumed that $\nu(B)$ is approximated by a ratio of two finite rational polynomials in B:

$$\nu(B) = \frac{\omega(B)}{\delta(B)},$$

where

$$\omega(B) = \omega_0 - \omega_1 B - \ldots - \omega_l B^l$$

$$\delta(B) = \delta_0 - \delta_1 B - \ldots - \delta_r B^r$$

Ultimately, therefore, the final equation representing the transfer function of X on Y and the memory structure within X and Y is as follows:

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} a_t,$$

with

$$y_t = \nabla^{d'} Y_t$$

$$x_t = \nabla^{d} X_t,$$

where d and d' represent the order of consecutive differencing that may be necessary to make the series stationary in the mean; and x_{t-b} is the differenced value for the X series at time t with a lag b.

The goal in transfer function analysis is to estimate the impulse response weights expressed as the ratio of the two polynomials at lags b. These estimated coefficients are similar to regression coefficients and represent the relationship between an antecedent and consequent variable. If the set of ω and δ coefficients are statistically significant, then there is some relationship between X and Y at lag b. Therefore, it is sufficient to examine the values in the ratio, $\frac{\theta(B)}{\phi(B)}$, to assess the nature of the relationship between X and Y time series variables. Conceptually, this relationship can be broken down into two components, delta (δ) and omega (ω). Omega reflects a fairly straightforward relationship between X and Y. That is, the omega coefficient describes the impact of the variable X on the variable Y, at lag b. Delta is also a reflection of this dynamic relationship between X and Y. However, it is more indirect, describing the impact of current values of Y on later values of Y.

APPENDIX C

Team Assignment

True/False Questions (Correct the false)
For the following true/false questions: mark your choices to the left of the question in the spaces provided. For those answers that you believe are false, write the correct answer below the question.

1. ___ An industrial psychologist decides he wants to use two predictors to select grocery store clerks. Ideally, these predictors will have a low intercorrelation.

2. ___ Dr. Banner determines the percentage of the employees in her organization that are currently successful. This percentage is known as the criterion percentage.

3. ___ The research of Schmidt and Hunter demonstrates that validity is situationally specific.

4. ___ The human resources director at the Perpsi Cola Company just hired 10 individuals to be production managers. There had been 100 applicants. Based on this information, you know that the base rate was .10.

5. ___ The value of a valid predictor is greatest when the base rate is .50. (True)

Short Answers - Please use the space provided for your answer. Please be concise.

1. Explain carefully the premise of validity generalization. To what extent has research supported the existence of validity generalization?

2. Clearly explain how the validity of a predictor, the selection ratio, and the base rate relate to how useful a predictor can be to an organization making selection decisions

3. (a) Draw a Venn diagram that illustrates two uncorrelated predictors of a criterion. (b) Draw a Venn diagram that illustrates two predictors of a criterion that are correlated.

Long Essay – Using the space provided, please answer the following question to the best of your ability.
(a) Define true and false positive selection decisions. (b) Define true and false negative selection decisions. (c) Explain how setting a predictor cutoff score higher or lower influences the number of selection errors that are made.

Extra Credit: Please attach your own paper in answering the following question:

(a) ADA states that employers must provide disabled persons with *reasonable accommodations.* (b)What is meant by reasonable accommodation? (c) Name 2 forms of reasonable accommodation that are not listed in the textbook.

APPENDIX D

The SYMLOG Adjective Rating Form

Team # _____

Please record your alphabetic identifier _______

Please record the alphabetic identifier of the person being described _____

Circle the best choice for each item

U...active, dominant, talks a lot.................... never....rarely....sometimes....often....always
UP... extroverted, outgoing, positive............. never....rarely....sometimes....often....always
UPF...a purposeful democratic task leader.......never....rarely....sometimes....often....always
UF...an assertive business-like manager..........never....rarely....sometimes....often....always
UNF...authoritarian, controlling, disapproving...never....rarely....sometimes....often....always
UN ...domineering, tough-minded, powerful.....never....rarely....sometimes....often....always
UNB...provocative, egocentric, shows off........never....rarely....sometimes....often....always
UB... jokes around, expressive, dramatic.........never....rarely....sometimes....often....always
UPB...entertaining, sociable, smiling, warm.....never....rarely....sometimes....often....always
P... friendly, equalitarian..........................never....rarely....sometimes....often....always
PF ...works cooperatively with others...........never....rarely....sometimes....often....always
F... analytical, task-oriented, problem-solving...never....rarely....sometimes....often....always
NF ... legalistic, has to be right...................never....rarely....sometimes....often....always
N... unfriendly, negativistic.........................never....rarely....sometimes....often....always
NB...irritable, cynical, won't cooperate.........never....rarely.....sometimes....often....always
B... shows feelings and emotions................never....rarely....sometimes....often....always
PB...affectionate, likeable, fun to be with ......never....rarely....sometimes....often....always
DP... looks up to others, appreciative, trustful..never....rarely....sometimes....often....always
DPF... gentle, willing to accept responsibility..never....rarely....sometimes....often....always
DF...obedient, works submissively..............never....rarely....sometimes....often....always
DNF...self-punishing, works too hard...........never....rarely....sometimes....often....always
DN... depressed, sad, resentful.....................never....rarely....sometimes....often....always
DNB... alienated, quits, withdraws...............never.....rarely....sometimes....often....always
DB...afraid to try, doubts ability.................. never....rarely....sometimes....often....always
DPB... quietly happy just to be with others ......never....rarely....sometimes....often....always
D... passive, introverted, says little...............,never ....rarely.... sometimes...often ...always

APPENDIX E

The PANAS

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you have felt this way today. Use the following scale to record your answers.

1	2	3	4	5
very slightly	a little or not at all	moderately	quite a bit	extremely

________ interested	________ irritable
________ distressed	________ alert
________ excited	________ ashamed
________ upset	________ inspired
________ strong	________ nervous
________ guilty	________ determined
________ scared	________ attentive
________ hostile	________ jittery
________ enthusiastic	________ active
________ proud	________ afraid

APPENDIX F

Workload Sharing

This questionnaire consists of statements about your team and how your team functions as a group. Please indicate the extent to which each statement describes your team. Use the following scale:

1	2	3	4	5
strongly disagree	disagree	neither agree nor disagree	agree	strongly agree

____________ Everyone on my team does their fair share of the work.

____________ No one in my team depends on other team members to do the work for them.

____________ Nearly all the members on my team contribute equally to the work.

APPENDIX G

Difficulty Questionnaire

Please indicate the degree to which you agree with the statements by circling one of the five points.

1. This assignment would require more than 45 minutes to complete as a graded test.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

2. This assignment is appropriate as a regular test of future students knowledge of Industrial-Organizational Psychology.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

3. Assume that this assignment would be used as a regular way of assigning grades to teams of students in this course. Indicate the degree to which you agree or disagree that this assignment would be an acceptable **graded** team assignment to be completed during the first 45 minutes of a regular class.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

4. Working on this assignment for 45 minutes in a team of fellow students would be frustrating to you.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

5. The questions on this assignment are appropriate to the material presented in this chapter.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

6. It would be fair of the instructor to expect completion of the questions on this assignment by a team in 45 minutes.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

7. The questions on the assignment are clear for a team to answer.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

8. I would prefer completing this assignment by myself than with other members in class.

1----------------------|----------------------3----------------------|----------------------5

Strongly Agree Moderately Agree Strongly Agree

9. Overall, on a scale form 1 (very easy) to 10 (extremely difficult), please circle the difficulty level of this assignment as a team assignment.

1----------2----------3----------4----------5----------6----------7----------8----------9----------10

Very Easy Moderately Difficult Extremely Difficult

APPENDIX H

NASA-TLX

Rating Scale Definitions

Title	Descriptions
MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Place a mark at the desired point on each scale:

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

VITA

Contact Information:

Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267

EDUCATION

Masters of Science in Psychology
Old Dominion University 2002
Masters of Arts with distinction in Psychology
Fairleigh Dickinson University 1998
Bachelor of Arts with high honors in Psychology
Queens College 1996

ACADEMIC APPOINTMENTS

Research Assistant, 9/00 to 5/01, Dr. Mark Scerbo, Graduate Program Director and Dr. Barbara Winstead, Department Chair of Old Dominion University

Teaching Assistant, 9/01 to 5/02, Quantitative Methods Course, Department of Psychology, Old Dominion University

Course Instructor, 9/02 to present, Introduction to Industrial and Organizational Psychology, Old Dominion University

PAPER PRESENTATIONS

Hanner, H. S. (2002, May). *When looks are deceiving: The effects of psychological diversity in mixed gender teams.* Paper presented at the 12th Annual Conference on Feminist Scholarship at Old Dominion University.

Hanner, H. S., & McIntyre, R. M. (2002, March). *The effects of team process training and perceived task workload on team performance.* Paper presented at the 23rd Annual Industrial/Organizational Behavior Graduate Student Conference, Tampa, FL.

Hanner, H. S., & McIntyre, R. M. (2001, May). *Description of an interrupted time series experiment.* Paper presented at the Virginia Academy of Science Meeting, Harrisonburg, VA.