ARI Research Note 99-01

Linking Leadership Emergence to Leadership Effectiveness and Team Performance in a Military Population

Roseanne J. Foti, Neil M. A. Hauenstein, and Joseph A. Sgro
Virginia Polytechnic Institute

Research and Advanced Concepts Office
Michael Drillings, Chief

October 1998

U.S. Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Virginia Polytechnic Institute

Technical review by

Michael Drillings

NOTICES

DISTRIBUTION: This Research Note has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This Research Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this Research Note are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
October 1998	RN/Final Report	

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Linking Leadership Emergence to Leadership Effectiveness and Team Performance in a Military Poplulation.	DASW01-95-K-0008
	5b. PROGRAM ELEMENT NUMBER 0601102A
6. AUTHOR(S) Foti, Roseanne J., Hauenstein, Neil M.A., Sgro, Joseph A. (VA Tech)	5c. PROJECT NUMBER 20161102
	5d. TASK NUMBER 1902
	5e. WORK UNIT NUMBER C08

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Dept of Psychology Virginia Polytechnic Institute Blacksburg, VA 24061	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM
U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	ARI
	11. MONITOR REPORT NUMBER Research Note 99-01

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

Study of individual traits on te emergence of leaders in a team and the impact on subsequent performance.

15. SUBJECT TERMS

Leadership Effectiveness Team Performance Leadership Emergence Military ROTC

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	76	COR: Dr. George W. Lawton

Final Technical Report

Linking Leadership Emergence to Leadership Effectiveness and Team Performance in a Military Population

Submitted by the

Virginia Polytechnic Institute and State University

Office of Sponsored Programs

to the

Army Research Institute

Prepared by

Dr. Roseanne J. Foti, Dr. Neil M. A. Hauenstein, and
Dr. Joseph A. Sgro

of the

Department of Psychology
Virginia Polytechnic Institute
and State University
Blacksburg, Virginia 24061

February, 1997

Linking Leadership Emergence to Leadership Effectiveness and Team Performance in a Military Population

According to Katz and Kahn (1978), leadership implies an influence increment, that goes beyond mechanically complying with one's role in an organization and routinely applying rewards or coercive power. A key argument we will make is that the ability to go beyond one's formal role depends on how a person is perceived by others. Based on this logic, we define leadership as the process of being perceived by others as a leader. Thus, leadership is not solely in leaders or solely in followers. Instead, it involves behaviors, traits, characteristics, and outcomes produced by leaders and interpreted by followers. Traits, behaviors and events are critical distinguishing features of leaders. Though these features may be made salient by leaders, they also must be noticed by perceivers. Perceptions others hold of leaders are critical for understanding the nature of leader-subordinate interactions, the use of direct and indirect influence by leaders and the amount of discretion afforded to leaders.

Leadership as a determinant of performance has been the central focus of leadership research for several decades. As noted by Lord and Maher (1991), leaders can have both a direct and indirect influence on performance. Leaders can directly influence subordinates in ways that change subordinate task or social behaviors and have a substantial impact on performance. For example, lower-level leaders may set goals or provide feedback to subordinates as a means of increasing their motivation; alternatively, such leaders may instruct or train subordinates as a means of increasing their job skills. For direct means, the source of a leader's effects on subordinates can be localized in specific leader behaviors. Less direct means by which leaders affect

performance generally change the cognitive structures, needs, or values of subordinates. These elements take longer to change but should have more lasting and powerful effects on subordinate performance. Thus, when leaders are perceived by followers as leaders, we expect the result to be improved subordinate performance and organizational effectiveness. Specifically, according to Hunt, Boal and Sorenson (1990), two major responses should occur. First, human resource maintenance variables will be enhanced. Subordinates who perceive their leaders as such are more satisfied with and committed to leadership in the organization. Second, implementation of strategy will be more successful, because of increased subordinate commitment and effort, resulting in improved subordinate performance and organizational effectiveness.

In sum, defining leadership in terms of perceptions has several advantages. First, it allows a more comprehensive view of the leadership process, inCorporating leader traits and behaviors as well as subordinate responses. Second, it affords a way to link perceptions of leader emergence with leadership effectiveness.

Leadership Emergence

One of the earliest approaches for studying leadership potential was to identify the personality traits of individuals perceived by others to emerge as the leader of a group. However, traits theories have not been given serious attention since Mann (1959) and Stogdill (1948) reported that no traits consistently differentiated leaders from nonleaders across a variety of situations.

Over the past few years, however, there have been several research findings indicating that the trait approach to emergent leadership may have been abandoned prematurely. First, in work conducted on perceptions of leaders by followers,

researchers found a core set of characteristics (i.e., decisive, determined) related to leadership in diverse situations (Foti, Fraser, & Lord, 1982; Lord, Foti, & De Vader, 1984). This research demonstrates the importance of traits as perceiver constructs, helping them to understand and predict a leader behavior. Second, a recent meta-analysis reexamined the relationship between personality traits and perceptions of leadership emergence (Lord, De Vader, & Alliger, 1986). The authors presented evidence that several leadership traits, specifically, intelligence, dominance and masculinity, indeed were related to leadership emergence. Thus, consistent with much of the earliest thinking on leadership, there are traits that are generally associated with leadership perceptions.

Traits and leadership emergence. A number of theorists and researchers recently have discussed cognitive factors associated with leadership (Fiedler & Garcia, 1987; Kotter, 1988; Lord et al., 1986). For example, Lord et al., (1986) found intelligence to be the trait with the strongest relationship to leadership emergence ($r = .52$). In addition, Lord et al. concluded that masculinity was significantly and positively associated with leadership perceptions. However, Rueb (1994) found that femininity was positively correlated with perceptions of leadership in a team based military environment. Finally, dominance also appears to be associated with leadership perceptions (Lord et al., 1986; Mann 1959; Stogdill, 1948). In a more recent study, Hills (1985) found dominance to be related to leadership in a sample of 237 managers.

Self-monitoring is another personality characteristic that has been studied as a correlate of leadership emergence. Self-monitoring is the ability to monitor and control expressive behavior is known (Synder, 1974). The high self-monitoring individual is

particularly sensitive to situational and interpersonal cues regarding the appropriateness of his/her social behavior. Furthermore, the high self-monitor uses these cues as guidelines for regulating his or her expressive behavior and self-presentation. That is, the high self-monitoring individual is sensitive to interpersonal and task requirements and has the ability to control his/her actions to present a desired identity (i.e., impression management).

Garland and Beard (1977) found that for females, high self-monitors were more likely to be chosen as leaders on a brainstorming task requiring discussion, consensus, and only minimal feedback on performance, than were low self-monitors. The same effect did not occur for males, nor did it appear for either gender on an anagram task. Foti and Cohen (1986) examined self-monitoring and leadership perceptions by establishing three-person groups each composed of one high, one moderate, and one low self-monitor. The groups were informed that their task required either a highly structured leader or a considerate leader. The results indicated that high self-monitors were significantly more likely to emerge as leaders in both situations. Ellis (1988) and Ellis, Adamson, Deszca, and Cawsay (1988) also reported significant correlations between self-monitoring and leader perceptions in classroom groups. Dobbins, Long, Dedrick, & Clemons (1990) found that high self-monitors emerged as leaders of problem-solving groups more frequently that did low self-monitors and men emerged as leaders more frequently than did women. Finally, Zaccaro, Foti and Kenny (1991) found self-monitoring was correlated with emergent leadership across situations.

General self-efficacy is a global, relatively stable trait that is an accumulation of success and failure experiences (Shelton, 1990). Although there is little general self-

efficacy research in the leadership area, research suggests that, relative to low general self-efficacy, high general self-efficacy individuals expend more effort, and persist longer on tasks (e.g., Tipton & Worthington, 1984). In a recent study, Smith and Foti (1997) found general self-efficacy was an important trait (along with intelligence, dominance, and masculinity) in predicting emergent leadership.

Finally, several researchers have attempted to develop a personality profile of emergent leaders. Hogan, Raskin and Fazzini (1990) found emergent leaders in a sample of police applicants to be high in intelligence, ambition and likeability. In a series of studies, Gough (1990) using the California Psychological Inventory (CPI) found an emergent leadership criteria to be highly correlated with several traits including, capacity for status, dominance, empathy, and independence. Finally, Morrow and Stern (1990) reported that individuals who performed better on a management assessment center exercise known as the Leaderless Group Discussion scored higher on the personality traits of ascendancy (dominance), intelligence and sociability. Thus, based on the recent research concerned with the leadership personality, there is additional evidence to support the link between traits and perceptions of leadership, especially for intelligence and dominance.

In summary, the current study was concerned with the relationship between leadership emergence and individual difference measures. To this end, intelligence, dominance, general self-efficacy, and self-monitoring were included as individual difference measures. It was generally expected that intelligence, dominance and general self-efficacy generally would be related to emergence. However, as is

explained shortly, self-monitoring was not expected to predict emergence in all situations.

Cross-situation consistency. Assessment of emergence in a single exercise is easy, but raises questions of generalization to other situations. More sophisticated emergence research looks at emergence across situations (e.g., Zacarro, et. al. 1991). Such research uses rotation designs in which individuals rotate through multiple group exercises that require different leadership behaviors. The cross situation prediction of rotation designs is that leadership is a function of the personal qualities of the leader, thus the same persons will emerge as a leader when aspects of the situation are varied.

The earliest attempts at testing the trait hypothesis of leadership emergence using the rotation design, manipulated only group membership. Both Borgatta, Bales, and Couch (1954) and Bell and French (1950), found leadership emergence was stable across groups when group membership was varied. Furthermore, Carter and Nixon (1949) found partial support for the trait-based explanation of leadership emergence when tasks were manipulated but group membership remained constant. Barlund, (1962) was the first to manipulate both group membership and task requirements. He concluded that leadership emergence depended, not on individual traits, but rather on situation variables. Kenny & Zaccaro (1983) reanalyzed Barlund's data using a quantitative model of social relations (Kenny, 1988, Kenny & Hallmark, 1991) and found that between 49% and 82% of the leadership variance could be attributed to some stable characteristic. Although this study seems to indicate that leadership is a stable characteristic, Kenny and Zaccaro did not identify the trait(s) responsible for this stability. However, the authors speculated that persons who consistently emerge as

leaders, possess the ability to perceive the needs and goals of a group and to adjust their own behavior toward the group accordingly.

Zacarro et. al. (1991) extended the previous investigations by incorporating a complete rotation design experiment. Subjects in 12 separate rotation sets completed four tasks in newly composed groups. Thus, all subjects within a rotation set interacted with one other (but only once) in completing the four tasks. After each task session, subjects indicated their perceptions of leadership in the group and the amount of consideration, initiating structure, persuasion and production emphasis behaviors displayed by each group member. Results indicated 40% of the variance in leadership perceptions was stable across tasks and groups and could be attributed to some individual variable. In addition, self-monitoring was correlated $\underline{r} = .22$ with these stable leadership perceptions.

Purpose of Study

Using a rotation design, the current study examined leadership emergence in a military population in terms of both cross-situational leadership and situational leadership. We identified behaviors emitted by participants that lead to perceptions of emergent leadership. Furthermore, we assessed relationships between personality, cognitive abilities, leadership behaviors, and perceptions of emergent leaders. The individual difference data and the leadership emergence data also were examined in relation to long-term leadership criteria, both across and within situations. In this manner, we attempted connect leadership emergence with leadership effectiveness and team performance.

The ROTO program. In order to accomplish the above goals, data collected in the initial rotation design phase of the study were analyzed using the Social Relations Model (Kenny, 1988; Kenny & Hallmark, 1991) and its corresponding ROTO computer program (Kenny, 1989). This model partitions the variance of the leadership ratings collected in rotation designs into three separate parts: the rater effect, the ratee effect, and an interaction term.

The ratee effect is the true leadership score and represents the extent to which an individual tends to be seen by others as a leader. The rater effect is a rater bias term which refers to the tendency of individuals to differ in terms of their willingness to ascribe high leadership ratings to other group members (i.e., similar to severity versus leniency). Finally, the interaction term is an error term which refers to that variance which stems from the interaction of the ratee and rater.

In order to examine leadership stability across tasks, it is necessary to partition further the ratee and rater variance into their stable and unstable components. Stable ratee variance indicates that an individual is seen as a leader across tasks. Stable rater variance examines the tendency of a rater to see others a high on leadership across tasks. In other words, stable variance is predictive in nature, in so far as it indicates the degree to which performance or rating in one task are related to performance or ratings on another. In contrast, unstable variances reflect fluctuations in the behavior of the ratee and rater. For example, unstable variances are the extent to which performance or ratings in one task is not indicative of performance or ratings on another task. Unstable variance can be further partitioned so that true unstable variance, or that variance not related to random error, is isolated.

Of particular interest in this analysis is the statistic lambda squared which is computed by dividing the stable variance in a set of ratings by the sum of the stable variance and true unstable variance. Thus, it represents the extent to which leadership is stable across different tasks. This term can be tested using an F-test with number of rotations minus one as the degrees of freedom. The ROTO program also produces an individual-level variable (what Zacarro, et. al. labeled "leadership score") that reflects the extent to which a person emerges in his/her rotation. Once the significance of lambda squared is establish, it is typical to then examine relationships with other variables using the leadership scores.

A distinguishing characteristic of the current study is the use of a "double" rotation design in which subjects participated in two different rotations. The double rotation design was used for two reasons. First, we wanted to estimate the reliability of

the leadership scores derived from the rotation analyses, and second, we wanted to estimate both across-situation leadership emergence and within-situation leadership emergence.

In single rotation designs, estimation of the reliability of the leadership score is analogous to an intraclass correlation where the stable variance in the leadership ratings (i.e., variance due to the repeated emergence of the same subjects) is divided by the total variance in the leadership ratings. Unfortunately, when testing for cross-situation consistency in leadership (e.g., Zaccaro et. al., 1991), this reliability coefficient has little substantive meaning because the leadership exercises require different leader behaviors. Thus, true changes in leadership as a function of the situation attenuates the reliability of the leadership scores when using this intraclass correlation.

A major advantage of the current study is that the reliability of leadership scores can be accurately estimated when testing for cross-situation leadership emergence. This is possible because in the full within subjects replication, each subject participates twice in each leadership exercise. Therefore, the reliability of the cross-situation leadership scores was estimated through test-retest reliability. Establishment of the reliability of cross-situation leadership scores is important because it allows better understanding of the relationships between personal characteristics, leader behaviors, and the cross-situational leadership scores. Furthermore, reliability of the cross-situation leadership scores is a prerequisite for the application of such rotation-patterned exercises for the prediction of long-term leadership effectiveness.

The second reason for the double rotation was to estimate within-situation leadership scores, as well as the across-situation leadership scores. In previous cross-

situation emergence rotation studies, four tasks are used and each is designed to tap different leadership skills. For example, the four tasks used by Zaccaro et al. (1991) were designed to tap initiating structure, persuasion, consideration, and production emphasis. Instead of using four different exercises (each requiring different leadership qualities), we assessed only two leadership abilities (initiating structure and consensus/team building) in the four tasks. In the second set of rotations, subjects completed parallel forms of the exercises used in the first rotation. In this manner, each subject participated in four initiating structure exercises and four consensus/team building exercises. Designing the study in this manner allowed for the estimation of two "leader-in-situation scores", as well as two across-situation leadership scores.

Potential Application. Although the primary goal of this research is basic understanding of leadership emergence and effectiveness, there is clear potential for application in terms of the development of methods for the early identification of effective leaders. Rotation designs used in leader emergence studies share some similarities with assessment centers traditionally used for the early identification of effective mangers. Both use situational exercises in which participants know that they are essentially in competition with other.

In contrast, there are some clear differences in how rotation designs are conducted that may prove advantageous in comparison to traditional assessment centers. First, rotation designs ensure that participants always perform each exercise with a different cohort, which should produce less biased estimates of leadership emergence. Second, rotation designs use ratings from participants instead of observers (cf. Ilgen & Fujii, 1976), which, if valid, would most likely provide better utility

than traditional assessment center ratings. Finally, the construct validity of situational exercises historically has been a thorny problem is assessment center research (e.g., Brannick, Michaels, & Baker, 1989; Schneider & Schmitt, 1992; Shore, Thorton, & Shore, 1990). The use of rotation designs has the potential to provide some insight into this issue because the focus is on leadership emergence across different leadership situations.

Overview and Predictions

In the typical rotation design used to assess cross-situation emergent leadership, each rotation contains nine subjects who perform four different exercises in groups of three. In each exercise session, each participant is teamed with two members of the rotation with whom they have not performed any other exercise. At the end of each session, participants evaluate the leadership qualities of each subject. One important result of the rotation analysis of these leadership ratings is an estimation of the extent to which leadership ratings are due characterlogical properties of each subject. This is represented by the aforementioned lambda squared statistic. Once the significance of lambda squared is established, attention turns the interpretation of the relationships of the leadership scores with other variables.

In the current study, we made two significant modifications to the traditional rotation design. First, instead of using four different exercises (each requiring different leadership qualities), we included only two types (initiating structure and consensus building) of leadership exercises. As before, effective performance in each of these two exercises required different leadership behaviors. To complete the necessary condition of four task sessions for each rotation, alternate forms of each exercise were

developed. The second modification was that, after the first rotations were completed, we conducted a complete within subjects replication in that each participant was assigned to a new rotation containing subjects with whom they have not interacted.

In rotation designs, estimation of the reliability of the leadership score is analogous to an intraclass correlation where the stable variance in the leadership ratings that is due to the repeated emergence of the same subjects in a rotation is divided by the total variance in the leadership ratings. Unfortunately, when testing for cross-situation consistency in leadership (e.g., Zacarro et. al., 1991), this reliability coefficient has little substantive meaning because the leadership exercises require different leader behaviors. Thus, true changes in leadership as a function of the situation attenuates the reliability of the leadership scores when using this intraclass correlation.

A major advantage of the current study is that the reliability of leadership scores can be accurately estimated when testing for cross-situation leadership emergence. This is possible because with the full within subjects replication, each subject participates twice in each leadership exercise. Therefore, the reliability of the cross-situation leadership scores can be estimated through test-retest reliability. Establishment of the reliability of cross-situation leadership scores is important because it allows better understanding of the relationships between personal characteristics, leader behaviors, and the cross-situational leadership scores. Furthermore, reliability of the cross-situation leadership scores is a prerequisite for the application of such rotation-patterned exercises for the prediction of long-term leadership effectiveness.

Another major advantage of the replicated rotation design is that it allows for the estimation of a "leader-in-situation score" that is not available in the typical leadership rotation design. This is possible because of the use of alternate forms for each leadership exercise and because of the full within subjects replication, which together result in four sets of leadership ratings for each participant in a particular leadership situation. Therefore, beyond the two cross-situation leadership scores generated for each participant, there were also two leader-in-situation scores that reflect to what extent he/she emerges as a leader in the type of leadership situations simulated by our two exercises.

Finally, the exercises were videotaped and trained observers coded the frequency of leadership behaviors exhibited by the cadets. These behavioral codings provided many important comparisons. First, it allowed a standard by which to compare the convergence of the leadership scores produced from the rotation analyses. To estimate convergent validity, the behavioral codings were aggregated across situations and within situations, as was done with the leadership scores. Second, the behavioral codings were similar to assessment center type ratings (in fact the leadership coding scheme was borrowed from an assessment center scoring protocol). To this end, comparisons could be made between traditional assessment center scoring procedures and an alternative scoring procedure. The across situation behavioral codings represent a common assessment center scoring procedure. We also computed a within exercise score, which is another common assessment center scoring procedure. These two traditional scoring procedures were compared with our alternative, within situation scoring procedure.

Reliability and convergence predictions. The first substantive issue concerns the reliability of the emergence data. We expected the across situation leadership scores and across situation behavioral codings to exhibit the lowest reliabilities. Given the across situation scores measured leadership in exercises requiring different types of leader behavior, it was less likely that these scores would tap large percentages of systematic variance. We expected the within exercise behavioral codings to produce the highest reliabilities. This prediction was based on the common finding in assessment center research that dimension scores within exercise are highly intercorrelated. As to the within situation behavioral codings, we predicted the reliabilities to be higher than the across situation reliabilities, but we believed that the within situation scores were not likely to achieve the levels of reliability seen in the within exercise behavioral codings.

Addressing the convergence of the leadership scores and the behavioral codings, we expected that the two sources (i.e., the peer ratings and the trained observer ratings) to converge. However, we expected that, relative to the across situation measures, the convergence would be higher for the within situation scores/codings. The logic of this prediction was based on the reliability hypotheses. If across situation measures tap less systematic variance than within situation measures, then it was likely that the within situation measures would exhibit better convergence.

Emergence Predictions. We made several predictions regarding relationships between cadet individual differences and emergence. First, we expected cadets with higher cross-situation leadership scores to exhibit higher levels of dominance, intelligence, general self-efficacy, and self-monitoring. We treated masculinity-

femininity as an exploratory dimension because, as previously mentioned, prior research has found conflicting evidence for this dimension. Second, we expected cadets with high leader-in-situation scores to exhibit higher levels of dominance, intelligence, and general self-efficacy. However, we did not expect a relationship between self-monitoring and leader-in-situation scores because we believed that cadets that are effective in only one situation will not exhibit behavior flexibility tapped by self-monitoring.

Leader Effectiveness Predictions. The final set of predictions dealt with the connections between leader emergence and leader effectiveness. We expected the lowest predictive validities from the across situation leadership scores/codings, due to the relatively lower levels of reliability. However, we believed predicted the highest predictive validities from the within situation leadership scores/codings, even though we the within situation scores were likely to be less reliable than the within exercise codings. This prediction was based on the notion that the systematic variance of the within exercise codings likely was inflated due to potential halo effects and systematic biases in observers (in spite of training). That is, we expected the systematic variance of the within exercise codings to contain a much larger percentage of irrelevant systematic variance than the within situation scores/codings.

Method for Phase I, Assessment of Leadership Emergence

Subjects

Subjects were 99 male freshman members of the Virginia Tech Corps of Cadets (the Corps). Eighty-one of these subjects formed the focal group who participated in two rotations. The Corps is a militarily structured organization in which all Virginia Tech students are eligible to enroll and is supervised by the Commandant of Cadets, who establishes overall policies and methods of operations for the Corps. Although it is not a requirement that Cadets be enrolled in ROTC, it is a requirement that all students enrolled in ROTC be a member of the Corps. The current study used only rising Freshman cadets enrolled in the Fall of 1995. Each cadet who participated in two rotations was paid $20 for participation.

Design

The current study was a "double" rotation design. In the typical, single rotation design, each rotation contains nine subjects who perform four different exercises in groups of three. In each exercise, each participant is teamed with two members of the rotation with whom he/she has not performed any other exercise. At the end of each session, participants evaluate the leadership capabilities of each subject. We modified the traditional design by requiring subjects to participate in two rotations. After completion of the first rotation, subjects were assigned to a second rotation of nine cadets. As with the first rotation, for each exercise, each participant was teamed with cadets with whom he had not worked with before. As previously mentioned, the reasons for adding the second rotation were to estimate the test-retest reliability of across-situation leadership scores and to generate within situation leadership estimates (to compliment the traditional across-situation leadership scores).

Leadership Exercises

Single rotation designs typically require 4 performance sessions. That is, 4 sessions are needed so that all nine members of a rotation work with each other once and only once. If the focus of a particular study is on within-situation leadership emergence, then 4 parallel forms of one exercise are used (e.g., Kenny & Hallmark, 1992). In contrast, if the interest is across-situation leadership emergence, then the four different exercises designed to tap different aspects of leadership are utilized (e.g., Zaccaro et al., 1991). In the current double rotation study four exercises were used, and each exercise had a parallel form. Two exercises (manufacturing game and tower building) were designed to measure initiating structure behaviors and two exercises (admissions/placement committee and a "lost" exercise) were designed to measure consensus/team building. The use of the parallel exercises allowed us to use the ROTO program to generate two across-situation leadership scores for each cadet. That is, an across-situation leadership score was computed from the first rotation using the four original forms of the exercises (leadership across two situations), and a second across-situation leadership score was computed from the second rotation using the four parallel exercises.

To estimate within-situation leadership scores, the data were reconfigured so that leadership scores were estimated within situation. That is, an initiating structure leadership score was computed by using the ratings generated from the parallel form for each of the two initiating structure exercises. Similarly, a consensus/team building leadership score was computed by using the ratings generated from the parallel form for each of the two consensus building exercises. To summarize, use of the double rotation led to four leadership scores for each dependent variable, for each cadet, as

opposed to a single leadership score produced in the typical rotation design. Two across-situation leadership scores (i.e., emergence across initiating structure and consensus building exercises) and two within-situation leadership scores (i.e., emergence within initiating structure exercises and emergence within consensus building exercises).

Initiating structure exercises. A manufacturing game and a tower building exercise were used to assess initiating structure behaviors. The manufacturing game was similar to that used by Zaccaro et. al. (1991). The purpose of the simulation was to maximize profit from the sale of finished products. Each team purchased raw materials (Lego blocks), manufactured toy products (jeeps, robots, or boats in the first rotation), and sold the completed product back to a buyer (i.e., the research assistant) for a profit. The simulation was divided into 2 organization sessions and 2 manufacturing sessions. After reading the exercise instructions, the cadets were given 8 minutes to plan and organize the first production session. This was followed by an 8 minute production phase. Next, cadets were given 5 minutes to plan and organize for the second 8 minute production phase. In the beginning, cadets were given $10,000 credit, a price list for the purchase of Lego blocks, a sheet listing the prices of the finished products, and sheets diagramming the assembly of the toys. After the first production session, the price list for the purchase of Lego blocks was changed. In the second set of rotations, a parallel form of the simulation was used in which the price lists, selling prices were changed and toys produced were modified to trucks, barges, and lifeguard stands.

Tower building was the second initiating structure task. In the first rotation, cadets were given forty seconds to build a tower of tinker toys as high as possible. They were given 20 minutes to plan the building phase. During this planning phase, cadets were allowed to examine and move the materials, but they were instructed that there would be a 5 second reduction in building time for every connection they made between the pieces. The parallel form of the exercise was very similar, except cadets were instructed to build the tallest and widest structure in which only one tinker toy piece was touching the table.

Consensus/team building exercises. An admissions/placement committee task and a "lost on the moon" exercise were used as the consensus building exercise. In the first rotation, cadets participated in a admissions committee task in which cadets were asked to assume they were an admissions committee at a business school. Cadets were given profiles of 8 applicants to consider for admission. The profiles provided information about high school g.p.a, standardized test scores, work history, personal interests, demographic information, etc. Cadets were instructed to take 10 minutes to review the profiles, individually, and to rank each of the applicants. After the individual ranking were completed, the cadets were given 20 minutes to come up with a committee ranking of the 8 profiles. For the group rankings, cadets were instructed to "avoid changing your mind simply to reach an agreement or avoid conflict", and to "avoid conflict-reducing techniques such as majority vote, averaging, or trading in reaching your decision".

We used a job placement committee exercise as the parallel form of the admissions committee exercise. This exercise entailed assigning 10 new hires to 10

different job titles. A profile of each employee was provided. Each employee profile contained demographics, major personality characteristics, interests and hobbies, and a ranking of the 3 job titles to which the employee preferred to be assigned. Also included on the employee profile was a ranking (relative to the other 9 new hires) of each employee's predicted performance on each job title based on psychological test scores. As with the admissions committee, cadets were given 10 minutes to generate their personal rankings, and then 20 minutes to generate the committee's rankings. The same conflict resolution instructions were used in both the admissions committee and placement committee exercises. Development of this exercise required two pilot studies to ensure comparability with the admissions committee exercise.

Finally, two lost exercises were used as the second consensus building exercises. In the first rotation, cadets performed the "Lost on the Moon" exercise in which they were instructed to assume their spacecraft had crash landed on the moon. Their mothership was on the other side of the moon. Besides themselves, 15 items survived the crash and they need to decide which items to take. Each cadet was given 10 minutes to rank order the list of 15 items in terms of importance for survival. After that, each team of 3 cadets was given twenty minutes to generate the group's rank ordering of the 15 items.

The "Lost at Sea" simulation was used in the second rotations. In this scenario, it is assumed that your ship as sunk at sea and, besides the crew, only 15 items have survived. Because the life raft is small, the crew must decide what keep. As with the Lost on the Moon exercise, cadets had 10 minutes to rank order the items by

themselves before taking 20 minutes to generate the team rankings. Again, for both of the "lost" exercises, cadets were given the same conflict resolution instructions.

Procedures

Because we were conducting a leadership emergence study, it was necessary to minimize pre-existing leadership perceptions of participating cadets. For this reason, only entering Freshman cadets were used and we conducted the emergence phase in the Fall semester (i.e., first semester attended). To further facilitate the minimization of pre-existing leadership perceptions, as much as possible, for each nine cadet rotation, we to assigned one member from each of the nine different companies in the Corps. That is, the organizational structure of the Corps is based on nine companies and each company lives together in different areas. As such, there is little cross-company communication among Freshman cadets. Thus, assigning cadets from different companies to each rotation reduced the probability that cadets had formed leadership perceptions prior to participation.

Ninety-nine cadets participated in the first 11 rotations, and 81 of these cadets returned for nine more rotations. Cadets were recruited from the company rosters, reported to the laboratory in groups of nine, and signed informed consent sheets. When participating in the first rotation, each cadet performed four exercises in different 3 person groups. After completion of each exercise, cadets rated leadership capabilities of fellow group members and filled out an individual difference measure before rotating to the next exercise. After completion of the first 11 rotations, 81 cadets were contacted for scheduling of the second rotations. As with the first rotations, cadets reported in groups of nine and completed each of the four parallel exercises.

Once again, each cadet worked with two new group members in each exercise. After completion of each exercise, cadets rated the leadership perceptions of fellow group members. Upon completion of the fourth exercise, each cadet was debriefed and paid for participation. Each session was videotaped.

Emergence criteria

Peer ratings. Three sets of emergence criteria were collected. First, after each exercise, participants evaluated each other using the five-item General Leadership Impression scale (GLI; Lord, et. al.., 1984). A sample item asks, "What degree of influence did this member exert in determining the final outcome of the task?" Group members rated each other and themselves using 5-point Likert rating scales ranging from "extreme amount" to "nothing". The GLI has been used in numerous studies (Zaccaro, et. al., 1991; Lord et. al., 1984) and is noted for both its reliability and validity.

Participants also record their observations of co-participants on a behavioral checklist designed to measure three behavioral dimensions. The dimensions and behavioral items were similar to those used by Gatewood, Thorton, and Hennesy (1990). The dimensions were: 1. Clarifying the situation (2 items), 2. Developing ideas (5 items), and 3. Influencing Action (5 items). Each item was rated on a frequency scale from 1 (Behavior Never Occurred) to 5 (Behavior Always Occurred). These item ratings were summed within dimensions to form a composite behavioral leadership rating for each dimension in each exercise. The term peer ratings is used to refer to participant measures of GLI and leadership behaviors.

Computations of the leadership scores. The peer ratings were used as the input to the ROTO program. As previously mentioned, the ROTO program allows an

estimation of leadership emergence via the λ^2 statistic. These peer ratings are then converted to leadership scores in which the raw ratings are converted to deviation scores that adjust the peer ratings given to an individual by the mean ratings that the rater gives to the other participants (see Kenny & Hallmark, 1992, pp. 33-34). That is, in the typical nine person, four task rotation (as was used in this study), each participant rates the other eight participants in the rotation. When the leadership score is computed for one individual, the raw ratings given that individual are adjusted for the rater's mean ratings of the eight participants. To compute the across situation leadership scores, the scores were computed for each rotation because each rotation contained both leadership situations. Computations of the within situation leadership scores required the data to be reconfigured so that the four initiating structure ratings were group together and the four team-oriented ratings were grouped together.

Behavioral Coding. The third set of emergence criteria was trained observer ratings of leadership emergence. The "behavioral codings" were based on an extension of the behavioral checklist filled out by the participants. The behavioral coding dimensions included the three dimensions from the behavioral checklist (Clarifying the Situation, Developing Ideas, and Influencing Actions), plus the dimensions of Acknowledging Contributions and Facilitating group processes. The Acknowledging Contribution dimensions included eight behaviors (e.g., "makes procedural suggestions to move discussion along). The facilitating dimension included seven behaviors (e.g., "praises others' contributions"). Coders noted the number of times that participants emitted target behaviors, and these frequency counts on each

dimension were then converted to a rating on a scale of 1 (little leadership behavior) to 5 (high amount of leadership behaviors).

Eight graduate and undergraduate students completed the behavioral coding of the videotapes, although the two graduates did approximately 85% of the coding. Each coder went through at least 12 hours of training which included practice with the various tasks, lecture and discussion of the rationale of each exercise, the difference between coding leadership behaviors and making leadership judgments, presentation of videotaped examples of each behavior, coding of transcribed exercises, and individual coding of practice videotapes.

Rater training primarily focused on improvements in the observation process, through emphasis on observing carefully, watching for specific behaviors, using behavioral checklists, and an introduction to systematic errors of observation. Approximately four hours of training involved the coding and discussion of the transcribed exercises. Coders then spent another four to five hours coding practice tapes. Observers were considered adequately trained when they obtained three consecutive interrater reliability scores of .9 or better with the graduate student ratings of the practice tapes.

Reliability of behavioral codings. Given the large number of tapes and the hours of observation involved, it was not feasible to have each tape coded by two observers in order to assess interrater reliability. Instead, one tape from each coder was randomly selected to be coded twice. For the undergraduate coders, one of the graduate students coded the randomly selected tape. For the two graduate student coders, the other graduate student coded the randomly selected tape. For the codings of each

participant on the tape (usually nine different cadets), interrater reliability correlations were computed and averaged across the participants. One undergraduate coder proved to be unreliable (i.e., mean interrater reliability of less than .80), and her tapes were recoded by one of the graduate students. Collapsing across all coders, .90 was the average interrater reliability.

Aggregations of the behavioral codings. For analyzing data, three different aggregations were used with the behavioral coding data. First for across situation leadership emergence, the dimension ratings were averaged across exercises, within rotation. For example, in rotation one, the four ratings (one for each exercise) of acknowledging contributions were averaged to represent acknowledging contribution. The same aggregation was done for rotation two behavioral codings. This aggregation within dimensions, across exercises/situations is typical in assessment center scoring.

To measure within situations leadership emergence, the behavioral codings were averaged within leadership situations (i.e., initiating structure or team-oriented), across the two rotations. For example, the two ratings of acknowledging contributions from the two manufacturing games were averaged with the two acknowledging contribution ratings of tower building, producing the acknowledging contribution rating for initiating structure tasks. The same procedure was done for team-oriented tasks.

Finally, behavioral codings were aggregated across dimension within each exercise. For example, the behavioral codings of the five different emergence dimensions were averaged for the manufacturing game in rotation one. Eight "exercise scores" were computed, one for each exercise. Such exercise scores are also used in assessment center scoring.

Individual Difference Measures

Cognitive Variables. The quantitative and verbal subtests of the Scholastic Aptitude Test (SAT) were used as intelligence measures. These data were available in the cadets' applications to Virginia Tech.

Personality variables. We measured each cadet on dominance, masculinity-femininity, general self-efficacy, and self-monitoring. Dominance and masculinity-femininity was measured by using the corresponding scales from the California Psychological Inventory (CPI; Gough, 1990). The dominance scale of the CPI is purported to measure leadership ability, dominance, persistence, and social initiative. The Femininity scale of the CPI assesses masculinity or femininity of interests. Factor analyses have indicated that the Dominance and Femininity scales of the CPI are orthogonal (Gough, 1990).

The General Self-Efficacy scale (Sherer, Maddux, & Mercadante,1982). This scale is designed to measure a general set of expectations that individuals bring to new situations (Smith & Foti, 1997). Psychometric research supports that the general self-efficacy scale is reliable and valid (e.g., Sherer, et. al., 1982).

Self-monitoring was measured by the Lennox and Wolfe (1984) self-monitoring scale. As previously mentioned, self-monitoring is a person's sensitivity to environmental and social cues, and the ability to adjust behavior accordingly. The scale contains 13 items that subjects rate on a 6-point scale ranging from 1 - "certainly, always false" to 6 - "certainly, always true". An example item is "In social situations, I have the ability to alter my behavior if I feel something is called for".

Results for Phase I, Leadership Emergence

Rotation Analyses

In order to test our hypotheses, it was necessary to establish stability in leadership emergence using the peer ratings. That is, do peer perceptions of the ratee as a leader in one situation match peer perceptions of the ratee in other group situations, where both task and membership have been varied? For rotation one, analysis conducted on GLI ratings indicated a significant proportion of stable or trait based variance in leader emergence, $\lambda^2 = .44$, $t(10) = 1.80$, $p < .01$. A similar effect was found for behavioral ratings, $\lambda^2 = .86$, $t(8) = 7.38$, $p < .01$. These data indicate a significant tendency for a person to be seen as a leader across different group situations. For the second rotation, the proportion of stable leader based variance was not significant. For GLI ratings, $\lambda^2 = .12$, $t = 1.11$, ns. For the behavioral ratings, $\lambda^2 = .16$, $t = 1.19$, ns. Thus, leadership scores were stable for the first rotation set, but not for the second rotation set.

Reliability of Leadership Scores

Cross-situation reliability Consistent with the rotation results findings, the test-retest reliability for the leadership scores across the two trials was low. For the leadership scores estimated from the GLI peer ratings, the correlation between rotation one and rotation two leadership scores was $r = .21$ ($p < .07$). For the leadership scores estimated using the behavioral ratings, the correlation between rotation one scores and rotation two scores was $r = .15$ (ns). Given the lack of stable across-situation leadership variance in the second rotation, greater faith was accorded to the rotation one leadership scores. However, it must be recognized that there is no direct evidence of the reliability of the rotation one leadership scores.

The interrater reliability of the behavioral codings already has been established. However, we were also interested in the reliability of the codings across exercises. Test-retest reliabilities were also examined for the across situation behavioral codings: Acknowledging Contributions ($r = .18$, $p > .10$), Clarifying Situations ($r = .33$, $p < .05$), Developing Ideas ($r = .34$ $p < .05$), Facilitating ($r = .42$, $p < .01$), and Influencing ($r = .30$, $p < .05$). As seen from these results the reliability of the across situation behavioral codings were modest, producing at most sixteen percent systematic variance.

Within Situation Reliability.

We also examined the reliability of the behavioral codings in terms of within situation codings. Each behavioral coding dimension represents a homogenous construct, especially within a particular leadership situation. Therefore, we estimated the internal consistency of the behavioral codings within each leadership situation. Table 1 presents the coefficient alpha for each dimension within each situation. These internal consistencies were generally low, as low as .18 with .50 as the highest internal consistency estimate. These results indicated that, although interrater reliability for each dimension was good, the codings within dimensions across exercises (even within a leadership situation) were not highly correlated.

Table 1.

Internal Consistency of Behavioral Codings for Initiating Structure Tasks and Team Tasks.

Behavioral Coding Dimensions	Coefficient Alpha	
	Initiating Structure	Team
Acknowledging Contributions	.18	.44

Clarifying Situations	.39	.30
Developing Ideas	.16	.50
Facilitating	.46	.44
Influencing	.47	.31

Table 2.

Internal Consistency of Behavioral Codings of Exercise Performance Within Rotation

Exercises	Coefficient Alpha	
	Rotation One	Rotation Two
Manufacturing Game	.68	.62
Tower Building	.61	.64
Admissions/Placement	.72	.77
Lost on the Moon	.77	.58

Within Exercise Reliability

Finally, we examined the reliability of the behavioral codings within exercise. First we examined the internal consistency of the behavioral codings within each exercise. As seen in Table 2, these estimates were reasonably high indicating that dimensions intercorrelations within each exercise were relatively high. We also computed test-retest correlations for the parallel forms of the same exercise (see Table 3). These test-retest results were very low, with only four of the sixteen correlations reaching significance.

Convergence of Leadership Scores and Behavioral Codings

The final measurement issue addressed was the convergent validity of the rotation analysis generated leadership scores and the behavioral codings. Table 4 presents the convergent validity coefficients within rotation. Convergence was moderate in rotation one, with somewhat higher correlations between the GLI leadership scores and the behavioral codings. Convergence was weaker in rotation 2, with the same trend for greater convergence on the GLI leadership scores, relative to the behavior leadership scores. Table 5 presents the convergence of the leadership scores and the behavioral codings within the two leadership situations. Convergence again was not particularly high. For both initiating structure and team-oriented exercises, the behavioral codings converged better with the GLI leadership scores that with the behavioral leadership scores.

Leadership Emergence Correlates

Across situation emergence. We expected across situation leadership scores and across situation behavioral codings to correlate positively with dominance, general

Table 3.

Test-Retest Reliabilities for Rotation One and Rotation Two Behavioral Codings by Exercise

Behavioral Coding Dimensions	Test-Retest Correlations			
	Manufacturing Game	Tower Building	Admissiona/ Placement	Lost on the Moon
Acknowledging Contributions	-.02	-.04	.15	.17
Clarifying Situations	.27*	.13	.09	.14
Developing Ideas	.29*	-.19	.13	.08
Facilitating	.15	.11	.03	.32**
Influencing	.33**	.18	.16	.04

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

Table 4.

Convergence of Peer Leadership Ratings with Behavioral Coding of Leadership, Within Rotation

Behavioral Coding Dimensions	Convergent Validity			
	Rotation One[a]		Rotation Two[b]	
	Behaviors	General Leadership	Behaviors	General Leadership
Acknowledging Contributions	.17	.12	.29*	.33**
Clarifying Situations	.49**	.55**	-.15	.27*
Developing Ideas	.39**	.45**	.21	.34*
Facilitating	.36**	.55**	.06	.12
Influencing	.49**	.56**	.15	.33**

a peer ratings correlated with rotation one behavioral codings
b peer ratings correlated with rotation two behavioral codings
* $p < .05$
** $p < .01$

Table 5.

Convergence of Peer Leadership Ratings
with Behavioral Coding of Leadership, Within Leadership Situation

Behavioral Coding Dimensions	Convergent Validity			
	Initiating Structure		Team	
	Behaviors	General Leadership	Behaviors	General Leadership
Acknowledging Contributions	.31*	.41**	.20	.32*
Clarifying Situations	.29*	.44**	.38**	.50**
Developing Ideas	.06	.30*	.30*	.34**
Facilitating	.43**	.59**	.40**	.50**
Influencing	.39**	.55**	.51**	.64**

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

Table 6.

Correlations of Rotation One Across Situation Leadership Measures with Personality Measures and Cognitive Measures

Individual Difference Measures	Peer Ratings		Behavioral Codings				
	Behaviors	General Leadership	Acknowledging Contributions	Clarifying Situations	Developing Ideas	Facilitating	Influencing
Personality[a]	r	r	r	r	r	r	r
Dominance	.09	.23*	-.15	-.01	.07	.14	.30*
Femininity	-.05	-.02	.10	.08	.13	.21*	.04
Self-Efficacy	-.08	-.08	.04	-.18	-.04	-.09	.12
Self-Monitoring	-.01	.07	-.20	-.08	-.16	-.22	.01
Cognitive							
S.A.T.-Quantitative	.00	.17	.02	.21	.36**	.30*	.04
S.A.T.-Verbal	.14	.25**	-.03	.31*	.36**	.45**	.29*

[a] Personality scores were deviated from rotation one means.

* $p < .05$

** $p < .01$

Table 7.

Correlations of Rotation Two Across Situation Leadership Measures with Personality Measures and Cognitive Measures

Individual Difference Measures	Peer Ratings		Behavioral Codings				
	Behaviors	General Leadership	Acknowledging Contributions	Clarifying Situations	Developing Ideas	Facilitating	Influencing
Personality[a]	r	r	r	r	r	r	r
Dominance	.18	.25*	.14	.15	.30*	.06	.09
Femininity	-.03	-.06	.19	-.08	.01	-.06	.06
Self-Efficacy	.11	.08	.10	-.11	.04	.02	-.12
Self-Monitoring	-.17	-.06	.00	-.09	.16	.08	-.03
Cognitive							
S.A.T.-Quantitative	.16	.18	-.04	.35**	.16	.24	.34*
S.A.T.-Verbal	.34**	.29**	.17	.15	.04	.03	.24

[a] Personality scores were deviated from rotation two means.

* $p < .05$

** $p < .01$

self-efficacy, self-monitoring, and intelligence. We made no directional predictions for femininity. Tables 6 and 7 present the correlations between the rotation one and two (across situation) emergence measures and the individual difference measures. Across the leadership scores and the behavioral coding, intelligence was clearly the best predictor of emergence. Focusing on the rotation one results (remembering that the rotation two scores are suspect), verbal SAT scores predicted five of the seven emergence measures. Personality correlates were weak. In rotation one, dominance predicted GLI leadership scores and influencing behaviors. Femininity predicted facilitating behaviors, but all other correlations were not significant.

Within situation emergence. Turning to within situations, it was expected that leadership scores and behavioral codings would correlate with dominance, self-efficacy and intelligence, but not self-monitoring. Table 8 presents the correlations for the initiating structure tasks. Again, intelligence was the best predictor of emergence with both SAT scores predicting GLI leadership score, developing ideas, facilitating, and influencing behaviors. There was no support for the prediction that dominance and general self-efficacy would correlate with emergence. In fact, the only significant personality correlate was self-monitoring's correlation with developing ideas.

Turning to the team-oriented exercises presented in Table 9, intelligence was once again the best predictor of emergence. Verbal SAT scores predicted six of seven emergence scores and quantitative verbal scores predicted three out of seven. Dominance again emerged as a moderate predictor of emergence in that both behavior and GLI leadership scores were related to dominance, along with influencing. Femininity and general self-efficacy did not correlate with emergence, but self-

Table 8.

Correlations of Initiating Structure Leadership Measures with Personality Measures and Cognitive Measures

Individual Difference Measures	Initiating Structure Peer Ratings: Behaviors	Initiating Structure Peer Ratings: General Leadership	Initiating Structure Behavioral Codings: Acknowledging Contributions	Initiating Structure Behavioral Codings: Clarifying Situations	Initiating Structure Behavioral Codings: Developing Ideas	Initiating Structure Behavioral Codings: Facilitating	Initiating Structure Behavioral Codings: Influencing
Personality[a]	r	r	r	r	r	r	r
Dominance	.05	.13	-.04	-.03	.18	.05	.08
Femininity	-.07	.05	.18	.11	-.03	.17	-.02
Self-Efficacy	-.02	.03	.05	-.15	.13	-.07	-.04
Self-Monitoring	-.04	.01	.04	-.01	.33*	-.05	-.01
Cognitive							
S.A.T.-Quantitative	.14	.38**	-.04	.20	.34*	.26*	.35**
S.A.T.-Verbal	.23*	.44**	.13	.10	.27*	.25*	.32*

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

Table 9.

Correlations of Team Leadership Measures with Personality Measures and Cognitive Measures

Individual Difference Measures	Team Peer Ratings: Behaviors	Team Peer Ratings: General Leadership	Team Behavioral Codings: Acknowledging Contributions	Team Behavioral Codings: Clarifying Situations	Team Behavioral Codings: Developing Ideas	Team Behavioral Codings: Facilitating	Team Behavioral Codings: Influencing
Personality[a]	r	r	r	r	r	r	r
Dominance	.23*	.37**	-.04	.08	.15	.13	.26*
Femininity	.02	.11	.11	-.05	.14	.06	.02
Self-Efficacy	-.12	.16	-.05	-.12	-.17	-.06	-.04
Self-Monitoring	-.09	.05	-.26	-.19	-.28	-.14	-.08
Cognitive							
S.A.T.-Quantitative	.08	.14	-.03	.28*	.27*	.33*	.08
S.A.T.-Verbal	.20*	.24*	.06	.31*	.25*	.30*	.34*

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

monitoring showed an interesting trend. Self-monitoring was not predicted to correlate with emergence within situations. However, self-monitoring was negatively related (two-tailed test) to acknowledging contributions and developing ideas (both at $p < .05$) and clarifying situations ($p < .10$). That is, higher self monitors tended to emerge less in the team-oriented exercises.

Within exercise. Finally, we aggregated the behavioral coding scores within each exercise, as often done with assessment center ratings. These exercise scores were also correlated with the individual difference measures (See Table 10). Results for the exercise scores were generally weaker than for the other operationalizations of emergence. Verbal SAT, the best predictor of emergence in prior results, was only significant in two of eight relationships. Only two personality correlates reached significance. In rotation two, dominance predicted emergence for the lost in the moon and self-monitoring predicted emergence in tower building. Interestingly, self-monitoring was negatively related to emergence in the tower building task in rotation one.

Discussion of Phase I

The lack of stable, person-based leader variance in rotation two was most disappointing. Although we expected relatively lower reliabilities for the across situation peer ratings, we thought that there would still be a significant amount of variance in peer ratings would be do to individual differences in the participants. There are several potential explanations of the poor ROTO results for rotation two. Because the second rotations were run the later in the semester, perhaps the cadets

Table 10.

Correlations of Rotation One and Rotation Two Behavioral Coding (Within Exercise) with Personality Measures and Cognitive Measures

Individual Difference Measures	Rotation One				Rotation Two			
	Manufacturing Game	Tower Building	Admissions/ Placement	Lost on the Moon	Manufacturing Game	Tower Building	Admissions/ Placement	Lost on the Moon
Personality[a]	r	r	r	r	r	r	r	r
Dominance	.15	-.17	.13	.06	-.02	.18	.16	.29*
Femininity	.10	.06	.17	-.01	.04	-.02	-.07	.08
Self-Efficacy	.01	-.20	.02	.00	-.04	.06	-.13	-.03
Self-Monitoring	.06	-.24	-.07	-.20	.04	.33**	-.07	-.08
Cognitive								
S.A.T.-Quantitative	.28*	.07	.17	.15	.30*	.02	.30*	.14
S.A.T.-Verbal	.19	.17	.26*	.26*	.20	.19	.17	.12

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

had begun to form leadership perceptions that influenced ratings in the second rotation. Assuming some accuracy in these leadership perceptions, then this explanation is unlikely because these pre-existing leadership perceptions would likely artificially inflate the amount of stable leadership variance. Practice effects are also a potential explanation. Familiarity with the exercises may have allowed poor leaders from the first rotation to acquire/model skills and abilities of effective leaders in the second rotation. This explanation suggests that the behavioral codings in rotation two would have higher means and lower variances than the codings from rotation one (i.e. reflecting that cadets became better, more consistent leaders at time two). However, paired t-tests of the rotation means produced no significant mean differences between rotations, and variability was actually greater on each coding dimension in rotation two.

These findings of no mean differences between the codings of the two rotations, and the greater variability in rotation two, both suggest a motivational explanation. Given the cadets had no internal incentives for doing well, once the novelty of the exercises waned, there may have been greater variability in the desire to do well and less conscientiousness in terms of rating other cadets. Deci's (1975) cognitive evaluation theory could also be relevant to this motivation problem. Our initial intention was to pay cadets for participation in the study, but the Commandant of the Corps. requested that we not remunerate cadets. However, to encourage greater willingness to participate in the second rotation, we convinced the Commandant to allow us to pay the subjects who came back for the second rotation. Cadets coming to the second rotation knew they would be paid. Deci's theory would predict that the awareness of this external reward would reduce the intrinsic motivation to do well in the exercises.

In terms of reliability, the poor test-retest reliabilities for the across situation leadership scores is not surprising given the problems with rotation two. Although some of the problem may have been due to conscientiousness to rate others, the generally modest test-retest correlations for the behavioral codings indicate it the problem was also due to real changes in who exhibited leadership behaviors in the second rotations. Our reliability hypothesis was essentially an ordering hypothesis that predicted the within exercise reliabilities would be highest, followed by within situation reliabilities, and that the across-situation reliabilities would be the lowest. This ordering pattern was not necessarily the case. Although the test-retest reliabilities of the across situation leadership scores were low, the reliabilities of the across situation behavioral codings were similar in magnitude to the reliabilities for the within situation behavioral codings. Also, the internal consistencies of the within exercise codings were much higher than the reliabilities for the across situation scores/codings, but the test-retest reliabilities for the within exercise codings were poorer than reliabilities of the across situation/scores/codings.

In conclusion, reliability suffered as soon as the operational definition of emergence used ratings (peer or codings) that went across exercises (whether across situation or within situations). That is, the highest reliabilities were the interrater reliabilities for the behavioral codings (i.e. when two raters coded the leader behaviors of a cadet in one exercise) and the internal consistencies of the within exercise codings (i.e., looking at the correlations among dimensions within each exercise).

The convergence of the peer ratings with the behavioral codings were reasonable, except for the cross situation scores/codings from rotation two. This

suggests that the reliability problems discussed above were not caused by differences between the sources. Peers and observers appeared to see similar evidence in the cadets. We expected the within situation codings to show better convergence than the across situation codings. Discounting the problematic across-situation scores from rotation two, there is slight evidence that the within situation scores converge better than the rotation one across situation scores. However, the trend is not strong.

The stronger trend was that GLI leadership scores converged better with the behavioral codings than the behavioral scores. This is interesting given the three behavioral dimensions rated by the cadets were identical to the three of the five dimensions coded by the observers. This suggests the expectation of peers to monitor and keep track of the behaviors of other participants is too demanding. It appears that peers ratings are likely to be accurate more at general, categorical levels (i.e., like the dimensions on the GLI), instead of a more behavioral level (cf. Lord, 1985).

Turning to emergence, the results were strongest for the intelligence prediction. Discounting the across situations scores/codings from rotation two, SAT scores consistently predicted emergence regardless of how emergence was operationalized. Personality dimensions were less successful. Dominance predicted emergence best in the team-oriented exercises and the across situation scores/codings, but dominance did not predict emergence for the initiating structure emergence measures. Femininity and general self-efficacy did not predict emergence. Self-monitoring was expected to predict emergence in the across-situation emergence scores, but not the within situation emergence scores. Instead, the most consistent finding for self-monitoring was a negative relationship with team-oriented emergence scores. Finally, the within

exercise emergence scores were not consistently correlated with any individual difference measure.

In conclusion, the emergence correlates with the individual difference measures were disappointingly small. Our hope was that the double rotation design would produce more reliable across situation scores, that would in turn lead to stronger correlations with individual difference measures. Instead, the cadet performance in the second rotation was highly variable (most likely due to greater differences in motivation). The double rotation design would probably work much better in a situation where real consequences were associated with performance in the exercises.

In spite of these problems, the within situation scoring protocol showed promise. The team-oriented scoring procedure had the strongest and most consistent pattern of correlations with the individual difference measures. Also, the within exercise scoring protocol did not correlate well with the individual difference measures. This occurred in spite of the fact that the within exercise codings had the highest levels of reliability. This finding is consistent with our stated notion that the within exercise codings contain significant amounts of irrelevant systematic variance.

Method for Phase II: Emergence and Effectiveness

The second phase of the study monitored each cadet's progression in terms of leadership effectiveness in the Corps and to assess relationships between leader emergence and leader effectiveness. Personnel records for each cadet were reviewed at the end of the Spring semester of the Freshman year and all relevant performance data were recorded.

Objective Criteria

The frequency of demerits, reprimands, sanctions and positive incidents were recorded. Demerits refer to minor rules violations. Reprimands occur as demerits accumulate. Sanctions, the most serious misconduct other than dismissal, occur as reprimands accumulate or if a cadet violates a major rule (e.g., alcohol in his/her room). However, only one participant in the study was sanctioned, therefore, this variable was dropped from the analyses. Incident reports are also kept for when a cadet performs above and beyond the call of duty. We labeled this variable as "positive incidents". Cadet grade point averages (GPA's) were recorded from the fall and spring semesters. The most important effectiveness criteria was promotion. Each spring, promotions are awarded for the upcoming fall semester. Rising Sophomores are eligible for promotion to Assistant Team Leader and/or Assistant Staff Corporals. We recorded promotions as a dichotomous variable (0 = not promoted, 1 = promoted). Finally, we also recorded who withdrew from the Corps. as a dichotomous variable (0 = quit, 1 = stayed). This variable was labeled as "quit".

Subjective Criteria.

Each cadet is evaluated by three superiors (Squad leader, Platoon Leader, and Company Commander) in the Spring semester. The evaluation form contains five specific dimensions and an overall dimension. The dimensions include: Leadership, Human Relations, Job Performance, Cadet Behavior, and Cadet Image/Fitness. The rating dimensions are sixteen point scales on a range of one (poor performance) to four (good performance). The sixteen points are accomplished by breaking the rating scale into .2 increments. That is, 1.0, 1.2, 1.4, 1.6, and so on.

There were two minor problems with the rating data. First, the Cadet Image/Fitness dimension had almost no variance, therefore it was dropped from the analyses. Second, evaluators are allowed to give an "unknown" rating if they believe they can't accurately assess a cadet. This produced missing data, especially for the leadership dimension (perhaps Freshman cadets did not have many opportunities to lead). Each rank (i.e., Squad leader, etc.) had similar amounts of missing data, therefore, it was not feasible to rely on the ratings from one rating source to overcome this problem. Instead, we aggregated ratings across raters (this strategy produced as much data as any one rating source). Aggregation was justified in that the average within dimension intercorrelations from the three sources were all greater than .85.

Results of Phase II: Predictive Validity of Emergence

Correlations (Person r and Point-biserial), multiple regression, and discriminant analyses were used to estimate predictive validity. Our basic prediction was that the within situation emergence scores would be better predictors of leadership than either the across situation emergence scores and the within exercise emergence scores.

Across Situation Predictions

Tables 11 and 12 present the correlations of the across situation emergence scores with the performance criteria. Looking at the results for rotation one in Table

Table 11.

Correlations of Rotation One Across Situation Leadership Measures with Performance Criteria

Criteria	Peer Ratings		Behavioral Codings				
	Behaviors	General Leadership	Acknowledging Contributions	Clarifying Situations	Developing Ideas	Facilitating	Influencing
Objective							
Demerits	-.03	.06	.04	.06	.07	.13	.16
Reprimands	-.22*	-.25*	.18	-.20	-.21	-.18	.03
Positive Incidents	.00	.07	.19	-.12	.05	.08	-.08
Promoted[a]	.21*	.23*	.08	.00	.12	.07	.11
Q.C.A.-Fall	.07	.17	.06	.18	.26*	.16	.08
Q.C.A.-Spring	.02	.04	-.07	.11	.12	.19	-.02
Quit[a]	.09	.08	.01	.02	.18	.01	.15
Subjective							
Leadership	.11	.18	.15	.26*	.30*	.37*	.07
Human Relations	.06	.04	.11	.17	.33*	.29*	.04
Job Performance	.10	.06	.10	.14	.19	.28*	.04
Cadet Behavior	.00	-.03	.17	.09	.10	.20	.12
Overall	.11	.08	.22*	.17	.19	.24*	.00

[a] Correlations estimated with point-biserial.

* $p < .05$

** $p < .01$

Table 12.

Correlations of Rotation Two Across Situation Leadership Measures with Performance Criteria

Criteria	Peer Ratings		Behavioral Codings				
	Behaviors	General Leadership	Acknowledging Contributions	Clarifying Situations	Developing Ideas	Facilitating	Influencing
Objective							
Demerits	-.05	-.03	.06	.03	.09	.14	.20
Reprimands	.01	.01	-.08	-.24	-.01	-.17	-.21
Positive Incidents	.12	.07	.11	.02	.03	.20	.04
Promoted[a]	-.03	.00	.20	.13	.01	.07	.21
Q.C.A.-Fall	.16	.21*	.01	.04	-.10	-.16	.05
Q.C.A.-Spring	.16	.22*	.29*	.01	.05	.05	.07
Quit[a]	-.03	-.08	-.12	.12	-.07	-.03	.00
Subjective							
Leadership	.09	.14	-.01	.26*	.21	.26*	.23
Human Relations	.06	.02	.01	.18	.13	.22	.12
Job Performance	.11	.07	-.06	.14	.12	.21	.06
Cadet Behavior	.10	.06	-.14	.07	.01	.07	.05
Overall	.08	.04	-.11	.10	.10	.10	.00

[a] Correlations estimated with point-biserial.

* $p < .05$

** $p < .01$

11, the leadership scores generated from the rotation analysis did not predict any subjective criteria, but did predict the important dimensions of promotions and reprimands. Cadets with high behavioral and GLI leadership scores received fewer reprimands and were promoted more often. In contrast, the behavioral codings from the trained observers did not predict objective criteria well, but did predict subjective criteria. The coded emergence dimension of facilitating predicted all subjective criteria, except Cadet Behavior. Only the coded emergence dimension of influencing failed to predict at least one subjective dimension.

Multiple regression/discriminant analyses were used to see if improvements could be made on the predictive accuracy of the bivariate relations. Improvements were seen in the prediction of reprimands, and the dimensions of Human Relations and Cadet Behavior. The Acknowledging Contributions and Developing Ideas dimensions predicted reprimands ($\underline{R}$ = .33). Influencing and Facilitating predicted both Human Relations ($\underline{R}$ = .39) and Cadet Behaviors ($\underline{R}$ = .35).

Not surprising, the predictive validity of the emergence scores from rotation two were poor (See Table 12). Only four bivariate relationships were significant and the multiple regression/discriminant analyses showed no improvements.

Within situation predictions

Table 13 presents the predictive validity of the emergence scores generated in the initiating structure exercises. The leadership scores consistently predicted GPA, but little else. The only consistency in the behavioral coding data was the prediction of

Table 13.

Correlations of Initiating Structure Leadership Measures with Performance Criteria

Criteria	Initiating Structure Peer Ratings: Behaviors	Initiating Structure Peer Ratings: General Leadership	Initiating Structure Behavioral Codings: Acknowledging Contributions	Initiating Structure Behavioral Codings: Clarifying Situations	Initiating Structure Behavioral Codings: Developing Ideas	Initiating Structure Behavioral Codings: Facilitating	Initiating Structure Behavioral Codings: Influencing
Objective							
Demerits	.08	.11	.15	.12	.20	.18	.46**
Reprimands	-.07	-.09	.16	-.15	-.11	-.15	.05
Positive Incidents	.12	.09	.13	.00	.10	.13	.07
Promoted[a]	.21*	.12	.13	-.03	-.08	-.03	.04
Q.C.A.-Fall	.21*	.23*	.10	.17	.20	.16	.09
Q.C.A.-Spring	.21*	.21*	.11	.18	.08	.23*	.07
Quit[a]	.08	.04	.01	.16	.21	-.03	.08
Subjective							
Leadership	.02	.10	.02	.33*	.28*	.27*	.05
Human Relations	-.11	-.08	-.10	.15	.22	.16	-.08
Job Performance	.03	.00	-.06	.16	.09	.17	-.17
Cadet Behavior	.00	-.05	-.04	.10	.03	.10	-.19
Overall	.01	-.03	-.11	.17	.06	.11	-.13

[a] Correlations estimated with point-biserial

* $p < .05$

** $p < .01$

the Leadership dimension. The emergent predictors of Clarifying Situations, Developing Ideas, and Facilitating all predicted Leadership ratings. Multiple regression analyses improved on the prediction of three criteria. The same two dimensions of Facilitating and Influencing improved the predictions of reprimands ($\underline{R}$ = .31), Cadet Behaviors ($\underline{R}$ = .33) and Job Performance ($\underline{R}$ = .41).

Table 14 presents the predictive validity of the team-oriented emergence measures. The leadership scores from the rotation analyses did not predict any criteria. In contrast, the behavioral codings from the trained observers consistently predicted criteria. Not counting GPA, three of the five objective dimensions were correlated with at least one dimension. Three dimensions predicted promotions (Clarifying the Situation, Facilitating, and Influencing). Bivariate results were even stronger for the subjective criteria. Facilitating predicted all subjective dimensions. The dimensions of Clarifying the Situation, Developing Ideas, and Facilitating were consistent predictors of the subjective criteria, whereas, Acknowledging Contributions and Influencing did not predict any subjective criteria. Multiple regression analyses showed improvements on the prediction of positive critical incidents, Human Relations, Job Performance, and Overall Performance. The combination of Acknowledging Contributions, Clarifying Situations, and Facilitating predicted positive critical incidents ($\underline{R}$ = .43). The dimensions of Facilitating and Influencing improved the predictions of Human Relations ($\underline{R}$ = .53), Job Performance ($\underline{R}$ = .46), and Overall Performance ($\underline{R}$ = .44).

Table 14.

Correlations of Team Leadership Measures with Performance Criteria

Criteria	Team Peer Ratings		Team Behavioral Codings				
	Behaviors	General Leadership	Acknowledging Contributions	Clarifying Situations	Developing Ideas	Facilitating	Influencing
Objective							
Demerits	-.14	-.07	-.03	.00	.02	.12	.07
Reprimands	-.08	-.14	.05	-.29*	-.19	-.23	-.17
Positive Incidents	.05	.05	.27*	-.07	.08	.18	-.10
Promoted[a]	-.06	.08	.15	.24*	.18	.26*	.36**
Q.C.A.-Fall	.06	.17	.0?	.10	.03	-.04	.06
Q.C.A.-Spring	.04	.10	.14	.14	.12	.15	.10
Quit[a]	-.04	-.06	-.15	.06	.04	.07	.19
Subjective							
Leadership	.16	.22	.08	.25*	.32*	.43**	.20
Human Relations	.20	.12	.11	.25*	.31*	.43**	.04
Job Performance	.17	.13	.09	.23	.2?*	.36*	.11
Cadet Behavior	.12	.08	.06	.12	.16	.25*	.01
Overall	.15	.15	.18	.20	.23	.33*	.04

[a] Correlations estimated with point-biserial.
* $p < .05$
** $p < .01$

Exercise Scores

The predictive validity of exercise scores was also examined (See Table 15). For the objective criteria, there was no consistent pattern of bivariate relations. Only the manufacturing game from rotation two predicted two objective criteria (reprimands and Fall GPA). The lost on the moon exercise predicted four of five subjective performance dimensions, otherwise there was no consistent pattern of prediction. Multiple regression/discriminant analyses could not improve on the bivariate predictions.

Comparisons Across All Emergence Scoring Procedures

Table 16 presents a comparison of the correlations or multiple correlations of the best predictors of eight criteria. Quitting was not included in Table 16 because no emergence measured predicted who withdrew from the Corps. GPA's were not included because they are not direct measures of leadership effectiveness. Rotation two emergence scores also were excluded from Table 16. Comparisons show that the behavioral codings of team-oriented exercises were the predictive of seven of the eight criteria listed, and produced the highest predictive validity on six of the eight criteria. Furthermore, as seen in Tables 11 through 15, the behavioral codings of team-oriented emergence exhibited the most systematic pattern of relationships with the criteria.

Predictive Validity of the Individual Difference Measures

Although team-oriented emergence scores were the best predictors of performance criteria, there remains the issue of the predictive validity of the individual difference measures. Table 17 presents the predictive validity of the individual

Table 15.

Correlations of Rotation One and Rotation Two Behavioral Coding (Within Exercise) with Performance Criteria

Criteria	Rotation One				Rotation Two			
	Manufacturing Game	Tower Building	Admissions/ Placement	Lost on the Moon	Manufacturing Game	Tower Building	Admissions/ Placement	Lost on the Moon
Objective	r	r	r	r	r	r	r	r
Demerits	.17	.09	-.04	.21	.16	.34	-.06	.15
Reprimands	.07	-.10	-.33**	-.03	-.28*	.16	-.14	-.04
Positive Incidents	.10	.00	-.06	.10	.03	.20*	.06	.07
Promoted[a]	.10	-.01	.15	-.08	.00	-.03	.12	.17
Q.C.A.-Fall	.20*	.21*	.18	-.06	.00	-.18	.10	-.04
Q.C.A.-Spring	.14	.01	.17	-.02	.22*	-.02	.06	.02
Quit[a]	.15	.04	.00	-.05	.12	-.07	-.09	.07
Subjective								
Leadership	.08	.22	.21	.25*	.35**	-.12	.18	.04
Human Relations	.04	.13	.22*	.23*	.20	-.06	.16	.02
Job Performance	-.07	.20	.21*	.21*	.19	-.19	.15	-.02
Cadet Behavior	-.10	.14	.12	.19	.14	-.19	.10	-.06
Overall	-.05	.17	.23*	.28*	.18	-.16	.09	-.01

[a] Correlations estimated with point-biserial.

* $p < .05$

** $p < .01$

Table 16.

Correlations on Multiple Correlations Representing the Best Predictors of the Most important Criteria

Performance Criteria	Across Situations: Rotation One Behaviors/ General Leadership	Across Situations: Rotation One Leadership Codings	Across Situations: Rotation Exercise Codings	Within Situations: Peer Ratings of Initiating Structure	Within Situations: Peer Ratings of Team	Within Situations: Coding of Initiating Structure	Within Situations: Coding of Team
Objective	r/R	r/R	r/R	r/R	r/R	r/R	r/R
Demerits	NS	NS	NS	NS	NS	.46	NS
Reprimands	-.25	-.33	-.33	NS	NS	-.31	-.29
Positive Incidents	NS	NS	NS	NS	NS	NS	.43
Promoted	.23	NS	NS	.21	NS	NS	.36
Subjective Leadership	NS	.37	.25	NS	NS	.33	.43
Human Relations	NS	.39	.23	NS	NS	NS	.53
Job Performance	NS	.29	.21	NS	NS	.41	.46
Cadet Behavior	NS	.35	NS	NS	NS	.33	.25
Overall	NS	.24	.28	NS	NS	NS	.44

Table 17.

Correlations of Personality Measures and Cognitive Measures with Performance Criteria

	Objective Criteria					Subjective Criteria				
	Demerits	Reprimands	Positive Incidents	Promotion	Quit	Leadership	Human Relations	Job Performance	Cadet Behavior	Overall
Personality	r	r	r	r	r	r	r	r	r	r
Dominance	-.03	.04	-.05	.25*	.07	-.13	-.09	-.14	-.13	-.08
Femininity	-.02	.01	.08	-.05	.11	-.02	-.06	.00	.07	-.03
Self-Efficacy	-.06	-.15	.14	.26*	.08	-.04	-.12	-.11	-.06	-.02
Self-Monitoring	.04	.02	-.06	.10	.09	-.06	.00	-.15	-.10	-.08
Cognitive										
S.A.T.-Quantitative	.05	-.08	.10	-.15	.04	.18	.19	.05	.08	.09
S.A.T.-Verbal	.12	-.01	.12	.09	.02	.10	.01	-.09	-.03	-.07

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

difference measures. Perhaps most surprising is that SAT scores did not predict any performance criteria. Furthermore, only promotions were predicted by the personality data. Dominance and general self-efficacy both predicted promotions.

Finally, the individual difference measures were correlated with GPA's (See Table 18). The measures of femininity, general self-efficacy, and both SAT scores predicted the fall semester GPA. However, only general self-efficacy predicted the spring GPA.

Discussion of Phase II Results

Our general prediction was supported for the emergence measures generated from the team-oriented/consensus building leadership situations. The behavioral codings for team exercises clearly "out predicted" the across situation scores/codings and the within exercise codings. Naturally, there needs to be some caution given the small sample size involved and the lack of cross validation. Also, the multiple regression strategy used (enter all predictors and removing nonsignificant predictors in a backward, stepwise procedure) is a liberal regression strategy. However, the point was to give the emergence predictors the highest possible predictive relations with the effectiveness criteria. The opportunity to capitalize on chance relationships was equivalent across all emergence criteria, therefore, it is unlikely that the findings for the team emergence scores capitalize on chance more than the alternative emergence measures.

The initiating structure emergence measures and the across situation leadership measures from rotation one did reasonably well predicting effectiveness criteria.

Table 18.

Correlations of Individual Difference Measures with Grade Point Average

Individual Difference Measures	Q.C.A.	
	Fall	Spring
Personality	r	r
Dominance	.00	-.08
Femininity	.24*	.16
Self-Efficacy	.25*	.28**
Self-Monitoring	-.11	.01
Cognitive		
S.A.T. Quantitative	.29**	.19
S.A.T.-Verbal	.27*	.16

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

Given, the findings for team emergence measures, the fact that the initiating structure measures did not out predict the across situation measures suggests that the criteria for the Corps were most sensitive to measuring and rewarding cadet behaviors that facilitate effective team processes. That is, cadets who were able to work well with others and to facilitate group processes towards effective outcomes were most likely to be evaluated higher and to receive promotions.

The within exercise scores were poor predictors of criterion performance. Again, this finding was consistent with the notion that although within exercise scores exhibit a great deal of systematic variance in terms of dimension interrcorrelations, these exercise scores do not possess a great deal of relevant systematic scores in terms of predictive validity. Perhaps it is best to think of exercise scores as analogous to one item on a test designed to measure leadership ability (cf. Banks & Roberson, 1985), instead of multiple dimensions measuring leadership ability/emergence within a given exercise. As such, building predictive validity would require aggregating scores across exercises in order to increase the amount of relevant systematic variance.

Finally, the individual difference measures were poor predictors of criterion performance. Most surprising was the fact that intelligence did not predict any criteria directly related to leadership. Perhaps it was too soon in a cadet's career for intelligence to predict leadership effectiveness. The most accepted causal explanation of role of intelligence in job performance is that greater aptitude leads to acquisition of greater job knowledge which in turn leads to greater job performance. However, for freshman cadets, there is no "job" to do, where the acquisition of greater knowledge

leads to better performance. This is consistent with our finding of lower predictive validities for emergence scores from the initiating structure exercises.

General Discussion

Overall, the purpose of this study was to address two major themes. First and foremost, we wanted to connect the study of leadership emergence with the study of leadership effectiveness. The second theme was to see if methods used in leadership emergence studies have potential benefits for the utility and understanding of assessment center processes. Underlying both these themes was our belief that it is better to operationalize measures of emergence/leadership ability within a given leadership situation, instead of the typical strategy (used both in leadership emergence and assessment center research) of operationalizing measures across different leadership situations.

Linking Emergence and Effectiveness

The logic underlying this research is that individual differences cause people to behave differently in newly formed, leaderless work groups, and that these behavioral differences are interpreted by perceivers in terms of leadership perceptions. If the work group stays together, over time, there is a dynamic relationship between the ways that individuals behave and the manners in which others respond. Nonetheless, reasonably accurate perceptions of leadership can be formed in short periods of time. Therefore, these leadership perceptions in newly formed leaderless work groups should correlate with both the individual characteristics of participants and with long-term measures of leadership effectiveness in other situations.

However, emergence leadership research has struggled with the first half of the model (i.e., individual differences result in different behaviors, that perceivers accurately interpret in relation to leadership qualities) and has ignored the second half of the model (i.e., emergence in leaderless work groups is associated with leadership effectiveness in other situations). In contrast, assessment center research addresses the second half of the model. However, this research is not focused on the leader emergent/effectiveness relationship per se in that assessment center research utilizes multiple exercises that measure different managerial skills (i.e., not just leadership).

Our hope was to establish more clearly the linkage between leader emergence and leader effectiveness through better measurement of emergence. To this end, we utilized the double rotation design in the attempt to get more reliable measures of cross situation leader emergence, and in order to compute measures of emergence within specific leadership situations. Unfortunately, the across situation measures of leadership (especially those based on peer perceptions) were not reliable and were not particularly effective in terms of correlating with either individual difference characteristics or leader effectiveness criteria.

Fortunately, results for our within situation measures were better. The emergence scores from the team exercises correlated the best with the individual difference measures and the leadership effectiveness criteria. Dominance and intelligence were related to emergence in team exercises, and emergence in team exercises was related to leader effectiveness. This suggests support for our two general predictions that emergence in short-term, leaderless groups systematically is linked to both individual difference antecedents and leader effectiveness consequence

in other situations. Further, that detection of this linkage is more likely when emergence is measured within leader situations, instead of across different types of leader situations.

Implications for Assessment Centers

Our first goal in relation to assessment center processes was to examine if peer ratings provided by the participants in the exercises were as reliable and valid as observer ratings. The utility of using peer ratings as opposed to observer ratings in an assessment center is obvious. However, our results clearly showed that the observer ratings were typically more reliable and predicted criteria better than peer ratings. As mentioned before, motivation of the cadets in the second rotation appeared to play a role in the low reliability and validities of the peer ratings. As such, the idea that peer ratings might work in an assessment center situation should not be abandoned. However, it should be recognized that even if peer ratings had done well in the current study, the motivation of participants is still a critical issue. In the current study, cadet motivation most likely waned due to the lack of consequences. In assessment centers, candidates typically are aware of the implications of their performance, and such motivations could profoundly affect peer ratings.

Results for the second goal of comparing alternative scoring protocols for observer ratings were more positive. Traditionally, assessor ratings are aggregated within dimension, across exercises (i.e. across situations) or within exercise, across dimensions (i.e., exercise scores). We tested the notion that ratings aggregated within dimensions, across only those exercises requiring specific leadership abilities (i.e., within situations) would provide better predictive validities than traditional scoring

protocols. In terms of team oriented leader exercises, we found that our within situation scoring protocol was superior to both the across situations and the within exercise protocols

Results were elusive in terms of our final goal related to better understand of the construct validity of assessment center ratings. Much as with the emergence/effectiveness linkage issue discussed above, we believed that better measurement of emergence might help with understanding the construct validity issues in assessment center research (e.g., Brannick, Michaels, & Baker, 1989; Schneider & Schmitt, 1992; Shore, Thorton, & Shore, 1990). However, the reliability of our measures were not that good, which clearly limits what can be said about construct validity. Also, we have no doubt that if there were more participants in the current study, a factor analysis would find the typical exercise factors instead of dimension factors.

Conclusions

Perhaps more than anything else, the most important outcome of the current study is that it gives clear guidance about how to establish a stronger linkage between emergence and effectiveness in future research. First, the utility of rotation designs may be limited in this domain. Although the rotation analyses can systematically detect stability in terms of which individuals emerge (e.g., Zacarro et. al. 1991), the general reliance on peer ratings in typical rotation designs appears problematic (cf. Ilgen & Fujii, 1976). The rotation generated leadership scores did not converge highly with the behavioral codings of emergence, and the leadership scores did not predict effectiveness criteria well.

Second, in rotation designs, the random assignment of individuals to groups probably causes too much noise to produce high relations between emergence measures and individual difference measures. For example, if a high dominance individual is grouped with two low dominance individuals, then the high dominance person is likely to emerge. In contrast, if three high dominance people are grouped, then emergence will depend on other characteristics (e.g., intelligence). A more fruitful strategy may be to assign individuals to groups based on individual difference patterns predicted to cause emergence, then rotate groups through different exercises while maintaining the structure of the individual difference characteristics. For example, Smith and Foti (1997) found much higher correlations between emergence and personality when using this type of design.

Third, the results of this study suggests that correlations between emergence and individual differences will be stronger if emergence is measured within specific leadership situations. That is, multiple exercises that tap a specific leadership behavior (e.g., initiating structure or consensus building) should be used. Relative to across situation emergence measures, operationalizing emergence within situations appears to produce more relevant systematic variance in relation to correlating both with individual differences and leader effectiveness.

Fourth, our results also suggest that serious thought should be given to the nature of the type of leadership behaviors important in the leader situation from which the criteria are collected. We examined the performance evaluation instrument used by the Corps prior to choosing the types of leader situations to use in the emergence phase. We concluded the initiating structure situation was appropriate given the

definitions of the "Job Performance" and "Leadership" dimensions and that the team oriented consensus building task was appropriate given the "Human Relations" dimension. As it turned out however, it appeared that team process leader behaviors were the primary types of leader behaviors recognized and rewarded in the Corps. As such, our results may have been better if we had used two team process leader situations (e.g., a conflict resolution task along with the consensus building task).

Finally, we had some success in linking leader emergence and effectiveness, and our belief in the fundamental model that emergence is the critical process variable that is linked to both antecedent individual differences and leadership consequences remains strong. We are confident stronger linkages between individual differences, leader emergence, and leader effectiveness will be found in future research that 1) uses trained observers, 2) groups participants by leadership profiles, 3) measures emergence within leadership situations, and 4) matches the emergence exercises to those leadership behaviors most relevant to the situation where the effectiveness criteria are collected

References

Ajzen, I, Timko, C., & White, J. B. (1982). Self-monitoring and the attitude-behavior relation. Journal of Personality and Social Psychology, 42, 426-435.

Banks, C. G., & Roberson, L. (1985). Performance appraisers as test developers. Academy of Management Review, 10, 128-142.

Bass, B. M. (1981). Stogdill's handbook of leadership. New York: The Free Press.

Bell, G. B., & French, R. L. (1950). Consistency of individual leadership position in small groups varying membership. Journal of Abnormal Social Psychology, 45, 764-767.

Berger, C. R., & Douglas, W. (1981). Studies in interpersonal epistemology: III. Anticipated interaction, self-monitoring, and observational context selection. Communication Monographs, 48, 183-196.

Berscheid, E., Graziano, W. G., Monson, T., & Dermer, M. (1976). Outcome dependency: Attention, attribution and attraction. Journal of Personality and Social Psychology, 34, 978-989.

Borgatta, E. F., Bales, R. L., & Couch, A. S. (1954). Some findings relevant to the great man theory of leadership. American Sociological Review, 19, 755-759.

Brannick, M. T., Michaels, C. E., & Baker, D. P. (1989). Construct validity of in-basket scores. Journal of Applied Psychology, 74, 957-963.

Carter, L. F., & Nixon, M. (1949). An investigation of the relationship between four criteria of leadership ability for three different tasks. Journal of Psychology, 27, 245-261.

Dobbins, J. H., Long, W. S., Dedrick, E. J., & Clemons, T. C. (1990). The role of self-monitoring and gender on leader emergence: A laboratory and field study. Journal of Management, 16, 493-502.

Deci, E. L. (1975). Intrinsic motivation. New York: Plenum.

Ellis, R. J. (1988). Self-monitoring and leadership emergence in groups. Personality and Social Psychology Bulletin, 14, 681-693.

Ellis, R. J., Adamson, R. S., Deszca, G., & Cawsay, T. F. (1988). Self-monitoring and leader emergence. Small Group Behavior, 19, 312-324.

Fiedler, F. E., & Garcia, J. E. (1987). New approaches to effective leadership: Cognitive resources and organizational performance. New York: Wiley.

Foti, R. J., & Cohen, B. (1986, August). Self-monitoring and leadership emergence. In J. Kennedy (chair), Attempting to solve the Leadership Puzzle. American Psychological Association, Washington, D. C.

Foti, R. J., Fraser, S. L., & Lord, R. G. (1982). Effects of leadership labels and prototypes on perceptions of political leaders. Journal of Applied Psychology, 67, 326-333.

Garland, H., & Beard, J. F. (1979). Relationship between self monitoring and leader emergence across two task situations. Journal of Applied Psychology, 64, 72-76.

Gough, H. (1990). Testing for leadership with California Psychological Inventory. In K. Clark & M. Clark (Eds.), Measures of Leadership (pp. 355-380). West Orange, NJ: Leadership Library of America.

Hills, D. A. (1985). Prediction of effectiveness in leaderless group discussions with the adjective checklist. Journal of Applied Social Psychology, 15, 443-447.

Hogan, R., Raskin, R., & Fazzini, D. (1990). The dark side of charisma. In K. Clark & M. Clark (Eds.), Measures of Leadership (pp. 343-354). West Orange, NJ: Leadership Library of America.

Hunt, J. G. Boal, K. B., & Sorenson, R. L. (1990). Top management leadership: Inside the black box. Leadership Quarterly, 1, 41-65.

Ilgen, D. R., & Fujii, D. S. (1976). An investigation of the validity of leader behavior descriptions obtained from subordinates. Journal of Applied Psychology, 65, 642-651.

Katz, D. & Kahn, R. L. (1978). The social psychology of organizations. (2nd. edition) New York: Wiley.

Kenny, D. & Hallmark, B. W. (1991). Rotations designs in leadership research leadership. The Leadership Quarterly, 3, 25-42.

Kenny, D. A., & Zaccaro, S. J. (1983). An estimate of variance due to traits in leadership. Journal of Applied Psychology, 68, 675-685.

Kenny, D. A. (1988). Interpersonal perception: A social relations analysis. Journal of Social and Personal Relationships, 5, 247-261.

Kenny, D. A. (1989). ROTO: A FORTRAN program for the analysis of group rotation data. Storrs, CT: University of Connecticut.

Kotter, J. P. (1988). The Leadership Factor. New York: Free Press.

Lennox, R. D. & Wolfe, R. N. (1984). Revision of the self-monitoring scale. Journal of Personality and Social Psychology, 46, 1348-2364.

Lippa, R. (1978). The effect of expressive control on expressive consistency and on the relation between expressive behavior and personality. Journal of Personality, 46, 438-461.

Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. Organizational Behavior and Human Performance, 34, 343-378.

Lord, R. G., De Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relationships between personality traits and leadership perceptions. Journal of Applied Psychology, 71, 402-410.

Lord, R. G., & Maher, K. J. (1991). Leadership and information processing: Linking perceptions and performance. Boston: Unwin Hyman.

Mann, R. D. (1959). A review of the relationships between personality and performance in small groups. Psychological Bulletin, 56, 241-270.

Morrow, I. J., & Stern, M. (1990). Stars, adversaries, producers and phantoms at work: A new leadership typology. In K. Clark & M. Clark (Eds.), Measures of Leadership (pp. 419-440). West Orange, NJ: Leadership Library of America.

Schneider, J. R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. Journal of Applied Psychology, 77, 32-41.

Sgro, J. A. (1982). US Army ROTC leadership development: A performance-based program.

Shore, T. H., Thorton, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. Personnel Psychology, 43, 101-116.

Snyder, M. (1974). Self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30, 526-537.

Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of literature. Journal of Personality, 25, 35-71.

Stogdill, R. M. (1963). Manual for the leader behavior description questionnaire-Form XII. Columbus: Bureau of Business Research, Ohio State University.

Zaccaro, S. J., Foti, R. J., & Kenny, D. A. (1991). Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group rotations. Journal of Applied Psychology, 76, 308-315.