Winter 2000

Toward a Model of Team Situation Awareness

Alexis Anne Fink
Old Dominion University

TOWARD A MODEL OF TEAM SITUATION AWARENESS

by

Alexis Anne Fink
B.A. December, 1993, DePauw University
M.S. May, 1996, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
December, 2000

Approved by:

Debra A. Major (Director)

William Fals-Stewart (Member)

Mark Scerbo (Member)

Alan Pope (Member)

ABSTRACT

TOWARD A MODEL OF TEAM SITUATION AWARENESS

Alexis Anne Fink
Old Dominion University, 2000
Director: Debra A. Major

Situation Awareness (SA) is a construct that is considered important to safety in dynamic, risky, time-constrained and complex environments, such as military aviation, nuclear reactors and emergency management. Research consideration of SA is complicated by the fact that there is no clearly superior methodology for SA measurement. Typically, SA is considered at the individual level; however, the nature of the SA context often requires more than one individual for safe and effective operations. Team SA is a qualitatively different phenomenon than individual SA. Few models of team SA have been proposed. The primary purpose of this paper was to develop and test a model of team SA. Existing models of team SA were reviewed, an integrated model was put forth, and each of the models was tested. Additionally, the paper explored and compared several methods for quantitatively assessing SA. Results indicate that one measure of SA, SALIENT (Muniz et al, 1997) has the best measurement characteristics. Model testing revealed that all models put forth fit the data adequately, but the summation model yielded the best fit to the data. Implications and suggestions for future research were outlined.

For my husband and daughter, Tom and Kristiane Keolker

and

For my grandpa, Charlie Fink

ACKNOWLEDGEMENTS

First, my family for their patience, sacrifice and encouragement. This effort has consumed four years of my family's life, and my husband and daughter have felt the weight of my work as much as I have. Kristiane, this is Mom's Book that you've been helping with for almost as long as you can remember.

Second, I need to acknowledge my committee for their guidance and wisdom. In particular, I need to acknowledge Debra Major for her tireless reviews of more drafts than I care to count, and her thoughtful and thorough feedback on each. Her guidance has been instrumental throughout this project. I also need to acknowledge the statistical and moral support offered by William Fals-Stewart. Mark Scerbo and Alan Pope both contributed insights during the design of this research that led to a strong final product.

Third, I would be remiss if I did not acknowledge the generous support of Paul Schutte, my NASA technical advisor on the GSRP that helped make this research effort possible.

Finally, I need to acknowledge the influence of my late grandfather, Charlie Fink. By his example, he taught me to love learning, and inspired me to pursue education with enthusiasm. I couldn't have done it without you, Grandpa.

TABLE OF CONTENTS

Page

LIST OF TABLES viii

LIST OF FIGURES xi

INTRODUCTION 1
Situation Awareness 4
Difficulties in the SA Literature 12
Teams 37
Models of Team Situation Awareness 65
Statement of the Problem 76

METHOD 79
Participants 79
Measures and Materials 80
Procedure 96

RESULTS 99
Measure Validation 99
Bivariate Hypotheses 114
Model Testing 116

CONCLUSIONS 126
Measurement Validation 126
Bivariate Hypotheses 136
Model Testing 140
Summary 145
Study Limitations 146
Conclusions 152

REFERENCES 154

APPENDICES
A. Situation Awareness Rating Scale (SARS), Caretta, Perry & Ree, 1994 168
B. Items from Situation Awareness Rating Technique (SART), Taylor, 1989 170
C. Situation Awareness Linked Indicators Adapted to Novel Tasks (SALIENT), Muniz, Stout, Bowers & Salas, 1997 172
D. Communication Evaluation 174
E. Communication Score Sheet 177
F. Participant Training Material 179

Page

G. Descriptive Statistics for Measures 182
H. Permission Letters 184

VITA 189

LIST OF TABLES

Page

1. Situation Awareness Rating Technique 33
2. Situational Awareness Linked Indicators Adapted to Novel Tasks 35
3. Types of Decision Making Teams 43
4. Role Ambiguity and Role Conflict "True r" Estimates 60
5. Measures Included in the Research 82
6. Proximal SA Questionnaire 84
7. Role Checklist 91
8. Items from Role Conflict and Ambiguity Scale 93
9. Experience Items 94
10. Performance Scale – Self 95
11. Performance Scale – Partner 95
12. Principle Components Analysis for Proximal SA Measure 101
13. Principle Components Analysis for Proximal SA Measure, Without Cued Item 101
14. Principle Components Analysis for Situation Awareness Rating Technique (SART): Game Ratings 103
15. Principle Components Analyses for Situation Awareness Rating Technique (SART): Individual Ratings 103

Page

16. Internal Consistency Reliability Estimate and Correlation with Performance for Situation Awareness Rating Technique (SART) Subscales 103

17. Principle Components Analyses for Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT) 104

18. Internal Consistency Reliability Estimate and Correlation with Performance for Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT) Subscales 105

19. Principle Components Analysis for Communication Measures 106

20. Internal Consistency Reliability Estimate and Correlation with Performance for Communication Measures 106

21. Principle Components Analysis for Role Ambiguity and Role Conflict 109

22. Internal Consistency Reliability Estimate and Correlation with Performance for Role Ambiguity and Role Conflict 109

23. Principle Components Analysis for Rizzo et al. (1970) Measure of Role Ambiguity and Role Conflict 110

24. Internal Consistency Reliability Estimate and Correlation with Performance for Rizzo et al.(1970) Measures of Role Ambiguity and Role Conflict 110

25. Principle Components Analysis for Experience Items 111

26. Internal Consistency Reliability Estimate and Correlation with Performance for Experience Measures 112

27. Principle Components Analyses for Subjective Performance, Without Item 6 113

Page

28. Internal Consistency Reliability Estimate and Correlation with Objective Performance for Subjective Performance Measures 113

29. Correlation Table for Measures Included in Model Testing 114

30. Goodness of Fit Indices for Summation Model 119

31. Goodness of Fit Indices for Non-Recursive Communication Model 120

32. Goodness of Fit Indices for Communication Model 121

33. Goodness of Fit Indices for Shared Understanding Model 123

34. Goodness of Fit Indices for Integrated Model 124

35. Fit Indices for Models 125

LIST OF FIGURES

Page

1. Three types of work groups 41
2. Overview of hierarchical team decision making theory 50
3. Examples of job-role combinations 57
4. Summation model 67
5. Non-recursive communication model 69
6. Communication model 71
7. Shared understanding model 74
8. Integrated model 76
9. Standardized regression estimates for summation model 119
10. Standardized regression estimates for non-recursive communication model 120
11. Standardized regression estimates for communication model 121
12. Standardized regression estimates for shared understanding model 122
13. Standardized regression estimates for integrated model 124

INTRODUCTION

Human operators in a wide array of occupations are increasingly asked to perform tasks that are more complex and to perform them with fewer errors. Technology at once increases our capabilities and increases the demands placed on human operators. Surgeons of centuries ago performed their procedures without benefit of monitors and modern surgical gadgetry, but their outcomes were significantly poorer than those we expect and accept today. Similarly, when the Wright Brothers made their famous flight, they were unconcerned by landing gear indicators, autopilot functionality, GPS or LORAN data, horizon indicators, electrical malfunctions or evading enemy detection. Yet, the simplicity of the system in which they operated also limited them significantly.

While technological advances provide us with increased information, capabilities and safety, they also exact a higher demand on our cognitive abilities. Where once tasks were easily managed by a solo operator, often with cursory training, now multiple operators, each extensively trained, are often required. While the demands may be manageable given the luxury of time and optimal operating conditions, when elements of risk and time-constraint are added, the difficulty of operations in these complex environments increases exponentially.

Clearly, to be effective, human operators in these situations must monitor numerous systems for changes, constantly updating their understanding of relevant pieces as well as the whole and using this updated information to guide their decisions and

---

The Journal model for this dissertation is: Murphy, K. R. (2000). *Journal of Applied Psychology, 85*(1).

actions. Failure to do so effectively can have disastrous consequences. For example, crew error has been identified as the primary causal factor in 70% of aviation accidents (specifically, hull loss accidents with known causes for the period 1988-1997; Boeing Commercial Airplane Group, 1998). Situation awareness (SA) is the term used to describe the state of consciousness, both process and product, required by these contexts.

SA has proved a rich ground for both theoretical and empirical work. Yet, despite the fact that teams are often necessary to operate effectively in these SA environments, research on team level SA is sparse. Furthermore, research that has addressed team level SA has generally ignored social and team processes as a means of understanding and facilitating SA team performance. This dissertation represents an effort to clarify the process by which team SA is achieved. The ultimate goal of this research was to empirically identify a model or models of team SA.

A few models of team SA have been discussed in the literature. This dissertation reviewed these models, and built an integrated model that drew from each of the models present in the literature. Specifically, this dissertation reviewed a summation model of team SA, two communication models and the shared understanding model of team SA.

The summation model essentially holds that team level SA is the sum of individual level SA. The communication models adds one level of complexity by proposing that team level SA is built by communication within the team. Third, the shared understanding model suggests that team SA is developed through shared understanding among the team. The integrated model builds upon elements of each of the models extant in the literature. Clearly, individual level SA is a necessary component of team SA. However, the integrated model suggests that good SA at the individual level is not sufficient to ensure good SA at the team level. Rather, as each team member will possess important information that is not possessed by other team members, communication within the team is required for good team SA. Third, the integrated model includes the concept of shared understanding as an important element of team SA.

This dissertation tested each of the models presented in the literature, as well as the integrated model developed in this dissertation. The strategy was designed to reveal which model accounted for the greatest proportion of variance as well as revealing which individual elements in the models make the greatest contribution to SA at the team level, and ultimately to team performance.

This dissertation also had a secondary aim. Although multiple measurement strategies for SA have been proposed, there is scant data with regard to the measurement characteristics, such as reliability and validity, of these measures. Thus, it is challenging for the SA researcher to select measures of SA with confidence. The secondary aim of the research was to empirically test the measurement characteristics of these metrics.

This secondary aim was also expected to support the primary aim of the dissertation, which was model testing. On the basis of the empirical evaluation of the metrics included in the study, it was possible to identify those measures which appear to be most reliable and valid. By including only those observed variables that possess good measurement characteristics in model testing, it was possible to eliminate some random error in each of the models described. Thus, the metric validation component of the study provided information that was vital to the central purpose of this dissertation.

Situation Awareness

Conceptualizing SA

Over the years, authors have put forth a wide variety of different definitions for SA. The definition has evolved since Bolman's (1979) theory of the situation was first proposed. Shrestha, Prince, Baker and Salas (1995) reviewed the definitions outlined by a host of authors and determined that five main attributes have been identified: 1) environmental and 2) temporal awareness in dynamically changing situations, 3) mission or goal awareness, 4) the ability to "observe, integrate, assess and act upon" (Shrestha et al., 1995, p. 51) the relevant information (environment, temporal, and mission or goal data), and 5) anticipation of future changes and events. Nearly all the definitions reviewed included awareness of surroundings. Thus, SA can be described as the state of one's awareness of elements in the context (surroundings, temporal place and goals or mission), the interactions of those elements and anticipation of future changes in those elements.

This description makes clear the importance of cognition in SA. Reviewing the elements identified by Shrestha et al. (1995), several cognitive factors are relevant; specifically, at a minimum, attention, working memory and mental models contribute to SA. Durso and Gronlund (1999) suggested that, in addition to these three, pattern recognition and naturalistic decision making are important cognitive factors in SA. However, Wickens (1999) was careful to clarify that cognitive processes such as attention are not SA in an of themselves, but rather processes that support SA.

Context is an important element to consider in researching SA. A certain amount of SA is required for even mundane tasks such as walking across one's living room; Durso and Gronlund (1999) point out that divided attention may provide a floor effect, maintaining a minimal level of SA even in the absence of operator attention. However, SA is generally researched in relationship to the requirements to operate in complex environments, as a failure of SA in a complex environment is likely to have more dire consequences than a failure of SA in a more mundane environment. Sarter and Woods (1991) point out that the contexts in which SA is most critical have several characteristics that make them particularly difficult to work within. In general, these contexts are characterized by multiple elements that are in a constant state of change, they typically have an element of substantial risk, and they are time-constrained (Sarter & Woods, 1991).

Given these attributes, developing and maintaining good SA can become quite literally a matter of life and death. Gaba and Howard (1995) further specified that critical elements or contextual cues may be challenging to perceive and integrate into overall SA. Cues may be available only briefly, or they may be subtle but meaningful deviations from normalcy. Further, certain combinations of elements, while individually unimportant, may have critical significance when combined. Thus, these are cognitively demanding contexts. The environment must be constantly scanned for relevant information, which must be processed, prioritized and incorporated in planning.

There is some disagreement in the literature as to whether SA is a state or a process. Although most authors treat SA as a state, some treat SA as a process (e.g., Sarter & Woods, 1995). Other authors have argued for the term situation assessment for the process and situation awareness for the state (e.g., Billings, 1994; Pew, 1994). Still others argue that the process/state debate is an artificial division, and that SA is in fact an element of consciousness (Smith & Hancock, 1995). Smith and Hancock (1995) argued that SA "is a dynamic concept that exists at the interface between the agent and its environment" (Smith & Hancock, 1995, p. 139). In the interest of clarity and consistency with the majority of the literature on this subject, this dissertation will treat SA as a state. However, it is explicitly recognized that the processes involved in developing and maintaining SA are critical to the state of SA.

Another important conceptual point must be made with regard to whether or not to include outcomes, such as decision making and performance, as a part of SA. Certainly, Bolman's (1979) discourse on theories of the situation included decision making as a part of SA, and other authors have treated decision making and action as incorporated into SA (e.g., Flach, 1995; Smith & Hancock, 1995). However, many authors (e.g., Endsley, 1995a) have taken the position that SA is distinct from decision making as well as performance. Endsley (1995a) pointed out that, "Even the best-trained decision makers will make the wrong decision if they have inaccurate or incomplete SA. Conversely, a person who has perfect SA may still make the wrong decision (from a lack of training on proper procedures, poor tactics, etc.) or show poor performance (from an inability to carry out the necessary actions)" (Endsley, 1995a, p. 36).

In defining SA to exclude outcomes, however, is it important not to discount the critical role that SA plays in those outcomes. SA is a necessary but not sufficient factor for good decision making and performance. SA not only provides the foundation for decision making, but may actually influence the decision making process (Endsley, 1995a). It has long been recognized that framing has a significant impact on decision making (Tversky & Kahneman, 1981); here, one's level of SA provides the decision frame, and thus does exert a direct influence on the decision. Adding another layer of complexity onto the matter, Venturio, Hamilton and Dvorchak (1989) found that both SA and decision making predicted performance. Thus, although each of the concepts (SA, decision making, and performance) are distinct, they are also intimately intertwined.

There may be considerable overlap in the processes that lead to the state of high SA and the processes that lead to good decisions. Adams, Tenney and Pew (1995) suggested that there is much to be gained through thoughtful examination of the interdependence of the processes which contribute to SA; this is also the case with the interdependencies in decision making processes. There is perhaps even greater benefit in examining the interdependencies among the processes that SA and decision making have in common. Ultimately, although good decisions and good performance are not in and of themselves SA, they may be considered post hoc indicators of SA, to the extent that SA provides the framework upon which decisions and ultimately performance are based. SA acts as a limiting factor on decision making and performance; outstanding decision makers and individuals with considerable expertise in the task at hand can only perform as well as their SA will permit.

Before leaving the topic of definitions, it is important to discuss the matter of terminology. Although the majority of the SA literature discusses *situation* awareness, there are some authors who use the term *situational* awareness. In the interest of clarity and consistency, this work uses the former term, situation awareness. Endsley (1994) pointed out that the former term (situation awareness) is the more precise term, in that it literally means "awareness of the situation," whereas the latter term (situational awareness) is an noun modified by the adjective "situational," and literally means "a type of awareness relating to situations." The term situation awareness is more direct and clearer than situational awareness, and thus, in the interest of parsimony, the better term to use.

Importance of SA

Accidents and mishaps in the types of contexts where SA is critical end all too often in tragedy. Aviation mishaps are particularly costly, both in financial (lost or damaged equipment) terms as well as in human (casualty) terms. The scientific community has recognized for quite some time that the majority of aviation accidents stem from human error (Foushee, 1984). Over the past decade, flight crew errors have accounted for seven times as many major aviation incidents (aircraft missing or beyond repair) as airplane failures, the next most common cause (Boeing Commercial Airplane Group, 1998). Of the 149 hull loss (missing or beyond repair) accidents for which causes have been identified during this period, 105 were primarily due to flight crew factors (Boeing Commercial Airplane Group, 1998). Over a 10 year period, 6,792 fatalities resulted from aviation accidents, and 213 aircraft were lost or damaged beyond repair (Boeing Commercial Airplane Group, 1998).

Overall, aviation safety records are extremely good (1.2 accidents per million departures for scheduled passenger operations and 4.4 accidents per million departures for all other operations; Boeing Commercial Airplane Group, 1998). However, despite the fact that aviation accidents are fairly negligible in a relative sense (very few incidences as a percentage of flights), the absolute impact is substantial, in terms of needless loss of life, financial losses due to missing or irreparable aircraft and possibly due to lost aviation industry revenue as a result of public perceptions of unacceptable safety.

Controlled flight into terrain, where a crew pilots an aircraft that is functioning normally into the terrain, e.g., the ground or a mountain, is the number one cause for aircraft incidents, by a substantial margin (Boeing Commercial Airplane Group, 1998). Controlled flight into terrain is a classic example of a loss of SA by the flight crew; typically the crew simply does not know where they are or what is happening. Controlled flight into terrain accounted for 36 (17%) of the accidents, and 2806 (41%) of the fatalities over the past decade (Boeing Commercial Airplane Group, 1998). The next most common cause, loss of control in flight, accounted for 31 accidents and cost 1932 lives (Boeing Commercial Airplane Group, 1998). There were only 2 incidents of the third most common cause of death, midair collision, which accounted for 506 deaths (Boeing Commercial Airplane Group, 1998). Together, these three accident categories, all of which suggest a loss of SA by the crew, cost 5244 lives, and accounted for 77% of all aviation-related deaths over the last decade.

Although the accident rate dropped substantially between 1959 and 1971 (approximately 45 hull loss or fatal accidents per million departures in 1959 versus fewer than 5 hull loss or fatal accidents per million departures in 1971; Boeing Commercial Airplane Group, 1998), accident rates have remained relatively stable since that time. Thus, it is reasonable to suggest that further improvement in accident rates will require a new approach to accident reduction.

Given that the vast majority of significant aviation incidents are attributable to flight crew failures, improving flight crew performance appears to be the best leverage point for improving accident rates. Elimination of 50% of the hull loss errors attributable to flight crew factors of the past decade would have prevented 52 of these accidents, or

35% of all commercial hull loss incidents worldwide over that 10 year period. A 35% accident rate reduction would have saved 2445 lives and 75 aircraft this decade.

A great deal of intellectual and financial capital has been devoted to research aimed at improving aviation safety through aircrew training, and development of superior cockpit systems. However, as discussed above, human error remains a large factor in aviation mishaps; much of this error can be characterized as deficiencies in SA.

The most famous example of lost SA by a flight crew in the SA literature is the tragic 1972 Eastern Airlines crash in the Florida Everglades. Here, the crew became preoccupied with an indicated nose landing gear malfunction and failed to notice that the autopilot had become disengaged. As with many tragic mishaps caused by poor SA, the Eastern Airlines crash of 1972 ended in controlled flight into terrain.

While SA has typically been addressed as important in an aviation context, it does have applications and implications beyond this. Anesthesiology and other medical systems (Gaba & Howard, 1995; Garland, Endsley, Andre, Hancock, Selcon, & Vidulich, 1996), nuclear power (Garland et al, 1996), air traffic control (Endsley, & Rogers, 1996; Hopkin, 1994; Mogford, 1994; Rantanen, 1994), advanced manufacturing systems (Garland et al, 1996), emergency services and emergency management (Companion, 1994; Schenk, 1994), battle fields (Garland et al, 1996), aircraft maintenance (Endsley, in press), and automobile driving (Garland et al, 1996; Gugerty & Tirre, 1996), have been identified as additional fields in which operators must have high levels of SA to perform safely and effectively. Each of these contexts contains the major elements that comprise an environment in which SA is critical - high information load, high cognitive demand, complexity, time constraint and presence of risk. Thus, although the models and research have typically focused on a fairly narrow set of applications, SA is relevant beyond the cockpit.

Although most of the statistics cited have been in reference to SA in an aviation context, this is largely due to the availability of data with regard to aviation accidents. Thorough investigations are conducted and the results published when aviation accidents occur. For obvious legal reasons, the same data are not available for medical incidents. However, this is not to suggest that SA is not relevant in these contexts or the others mentioned above.

Early Work

Two decades ago, Bolman (1979) proposed the "theory of the situation" as an important factor in aviation accidents. Bolman's theory of the situation is "a set of beliefs about what was happening and what actions it was appropriate to take" (Bolman, 1979, p. 34). Bolman further suggests that a theory of the situation "is a short term theory used by an individual to analyze and make decisions about the immediate environment" (Bolman, 1979, p. 35). Bolman proposed that aircrews test assumptions and look for differences between their espoused theories and their theories-in-use. Here, aircrews proactively seek out information to ensure that they are operating on the most correct information.

Bolman (1979) identified four factors that increase the probability that aircrews will be able to discover and correct faulty theories of the situation (or SA). The first is essentially the pilots' training and experience, and the extent to which pilots rely on inquiry and testing in situations that are not optimal. Second, Bolman suggested that a crew's ability to detect and correct faulty SA is determined by their ability "to combine skills in advocacy and inquiry" (Bolman, 1979, p. 32). Bolman's third major predictive factor is the management skills and style of the captain. Finally, Bolman's fourth major factor is "the degree to which the role system in the cockpit is well understood, and procedures for role-modification are mutually shared" (Bolman, 1979, p. 32).

Perhaps the most interesting thing about Bolman's seminal work on SA is the fact that it focused on SA as a team level phenomenon, and that the factors he identifies as important for maintaining high levels of SA (or accurate theories of the situation) are social ones. It is curious that the SA literature has such emphasis on individual-level SA, and cognitive and technological approaches to increasing SA, given this jumping off point.

Levels of SA

As discussed above, SA has three main components: awareness, integration and anticipation. Endsley (1988, 1995a) defined these as three levels which build upon each other. Level 1 SA is the awareness of elements in the environment. Level 2 SA is the integration of those elements into a single coherent picture. Level 3 SA is the projection of that integration into the future, and planning to address this anticipated future.

Clearly, each of the levels builds upon the preceding one in such a way that the preceding one is necessary but not sufficient for each next higher level SA. One cannot integrate elements of which one is not aware (transition between levels 1 and 2), nor can one formulate an effective plan based on a faulty comprehension of the overall situation (transition between levels 2 and 3).

Difficulties in the SA Literature

Perhaps one of the reasons that the concept of SA has not gained more currency in the scientific literature is the presence of a few persistent problems in the SA literature. These difficulties generally fall into three major categories: methodological, conceptual and operational. Somewhat tongue-in-cheek, Billings (1994) mused, "situation awareness (SA) is a process – or is it a product? Or is it both, or neither? It either is, or is not, critical to define SA precisely. It either can, or cannot, be quantified, but if it can be, the Heisenberg principle probably applies, and we alter it in the process of measuring it." (Billings, 1994, p. 321). Clearly, challenges such as these can cast a pall over a line of scientific inquiry. Compounded, difficulties in all three of these domains (i.e., methodological, conceptual and operational) can render a topic "unresearchable."

Methodological Difficulties

Because of the unique nature of SA and the specialized nature of tasks for which SA is most critical, researchers face some special methodological challenges. First, SA researchers are limited in their ability to assess operators' SA in context; the element of risk and time constraint that characterize environments in which SA is critical are barriers to allowing a researcher to collect SA data while operators are performing their tasks. Thus, researchers are generally limited to using simulations or relying on post hoc or retrospective measures.

Second, because the individuals who perform these tasks are generally members of select groups (e.g., fighter pilots, air traffic controllers, anesthesiologists), SA researchers are frequently limited in their ability to collect data from large samples. Thus, one of the other limitations on much SA research is the small sample size usually associated with empirical studies of SA (Carretta, Perry & Ree, 1994). In their review of the empirical SA literature, Carretta et al. (1994) found an average sample size of 21.75. Sample sizes are generally limited in that researchers have drawn largely from expert operators (primarily pilots) in their research on SA. However, recently SA researchers have been using participants, although small sample sizes are still prevalent (e.g., 32 in Bolstad & Endsley, 1999, 12 in Farley, Hansman, Endsley, & Amonlirdviman, 1999).

Conceptual Difficulties

The conceptual difficulty in studying SA lies in the separation of SA from the processes it is meant to influence (decision making and performance, principally). Indeed, aside from its ability to have an effect on these critical outcomes, SA is an esoteric concept, interesting only to theorists. Earlier in this paper, SA was discussed as both a state and a process by which that state is obtained. Other authors appear to include outcomes in their conceptualization of SA, such that adverse outcomes are suggestive of faulty SA. This dissertation has argued for defining SA as distinct from outcomes, such as decision making and performance, on theoretical grounds; there are practical grounds for this position as well.

Including outcomes in the conceptualization of SA introduces a considerable amount of error into the equation. For example, it includes situations that may be beyond the operators' control. Yet, one can also argue that exclusion of extenuating circumstances merely restricts the variability in SA. For example, if an aviation crew fails to complete its mission due to engine failure, the first perspective might only consider the crew's SA prior and subsequent to the engine failure and their integration of the engine failure into a plan to return safely. The second perspective might go further in including the crew's performance in completing the flight safely in assessing SA, or even consider the engine failure itself as a mark against their SA (e.g., they were not able to predict and prevent the problem). As discussed earlier, the processes that lead to the state of SA and the outcomes of decision making and performance are intimately intertwined. However, this dissertation maintains the position that SA is the state of awareness (here, recognition of engine failure and integration of that failure into a plan to meet objectives), and that decisions and ultimately performance are indicators of SA, but are not SA themselves.

Operational Difficulties

Beyond these conceptual difficulties, there is another major category of difficulties in SA: problems related to the operationalization of SA. The central problem is that because SA is a cognitive process, it can only be inferred and never directly observed, although the inputs and outputs can be directly assessed. In addition, some argue that SA is too subjective for quantitative measurement (Flach, 1995; Gilson, 1995). Although many methods have been developed, there is no consensus in the SA literature with regard to the validity and applicability of the various techniques (Garland et al, 1996). However, moving the study of SA beyond an amusing theoretical exercise requires that we be able to test our theories, and that requires reliable and valid measures of SA (Garland et al., 1996).

Although Endsley's (1995b) piece in the *Human Factors* special issue on SA is probably the most comprehensive treatment of SA measurement, several authors have examined and categorized the various attempts to empirically capture SA. These authors all propose substantially different taxonomies for SA measurement, thereby muddying the waters for the student of SA. Ultimately, two major approaches to SA measurement have emerged: proximal and distal, although a third approach, a middle-range or meso approach, has also been introduced in the SA literature.

Each of these three approaches to SA measurement, proximal, distal and meso, asks a different question with regard to SA. Proximal approaches to SA measurement ask the question, does the operator have the basic building blocks necessary for good SA? Thus, proximal SA measures operate at Level 1 SA, perception of elements in the environment.

Distal measures, on the other hand, ask the question, when all is said and done, was the operator's level of SA sufficient to do the job? In the earlier discussion as to whether or not to include decision making and performance as SA per se, the present author outlined several reasons why decision making and performance were not part of SA itself, although they are important consequences of SA. Distal measures, therefore, address the products of SA, rather than SA itself; that is, distal SA measurement approaches actually capture outcomes of SA, rather than measurements of SA itself.

Meso measures, or middle-range indicators, attempt to assess SA itself, rather than focusing on inputs, like proximal measures, or outputs, like distal measures. Extending this illustration, meso measurements focus on throughputs. However, the nature of SA makes meso indicators inherently challenging to collect. It was stated earlier that one of the major difficulties in SA measurement is that SA cannot be measured directly, yet this is precisely what meso indicators attempt to do.

Additionally, attempts to categorize SA measurement paradigms can be dichotomized as either objective or subjective. This is a basic operationalization decision that must be made in nearly all quantitative research. When researching SA, objective data may be such things as mission goals achieved or percent of time at the prescribed altitude. Subjective measures of SA, on the other hand, typically involve observer ratings of an operator's SA.

Any dichotomization of measurement approaches belies the complexity that has arisen around the issue of SA measurement. A host of measurement paradigms has emerged in the study of SA. Indeed at times it appears that each author, frustrated with the inadequacies of existing SA measurement techniques, developed her or his own operational definition and measurement strategy.

Proximal Measures of SA

A review of the definitions of SA revealed that one of the elements of SA upon which most authors and theorists agree is that of awareness of elements in the environment. Thus, this variety of immediate awareness seems a logical starting place for SA measurement. Level 1 SA, the level generally targeted by proximal measures, is a limiting factor on higher levels of SA, and ultimately, on performance. However, this is not a guarantee that high levels of level 1 SA will lead directly to high levels of overall SA and performance.

Objective approaches to proximal SA measurement. Perception research has long relied on physiological measures. Thus, it seems logical that, when searching for objective means by which to assess an operator's perception of elements in the environment, physiological techniques should be explored. Physiological techniques, such as electroencephalography (EEG) and eye-tracking devices allow researchers to assess whether elements have been sensed (i.e., whether or not the subject's eyes were pointing to a particular location), although sensation does not necessarily mean that the elements were attended to or understood. As Endsley (1995b) pointed out, physiological techniques provide no information with regard to the amount or accuracy of the information retained in the operator's memory, or of what meaning the operators attach to the information that is perceived. Thus, Endsley (1995b) dismissed physiological measures as "not very promising" for SA measurement.

Another perceptual theory that has been considered in SA research is signal detection theory (e.g., Atkinson, 1963). At a very basic level, signal detection theory's hits, misses, correct rejections and false alarms are very applicable to SA research. In some ways, this is the very essence of level 1 SA. This has been discussed as an implicit approach to SA measurement (Brickman, Hettinger, Stautberg, Haas, Vidulich & Shaw, 1999; Garland et al., 1996). Typically, these implicit measurement approaches apply signal detection theory to assess the operator's perception of and reaction to prespecified critical events and cues. For example, if an operator uses any radar setting other than that which has been previously specified as correct during a mission segment, inadequate SA would be recorded for that item during that segment. Thus, correct actions are scored as indicating adequate SA and incorrect ones are scored as indicating inadequate SA during the preprogrammed mission segment.

Although signal detection theory provides an important analysis tool, in a panel discussion, Vidulich (Garland et al, 1996) questioned the scope of SA that signal detection theory captures. In a later publication, Vidulich and his colleagues (Brickman et al., 1999) suggested that their Global Implicit Measurement technique has several advantages in that reactions to the prespecified events can be weighted to yield a more realistic assessment of SA.

Endsley (1995b) addressed the use of questionnaires in assessing SA. Of all the approaches to SA measurement discussed, Endsley suggested that this one held the most promise. Because the questionnaires that she discussed involve assessing objective information such as number of enemies currently threatening, they are more likely to be valid. In addition, questionnaires are more able to tap directly into an operator's SA, rather than simply inferring it.

Endsley (1995b) discussed three approaches to questionnaires: posttests, on-line assessment and the freeze technique. Posttests, administered directly after a trial, offer the advantage of permitting subjects to respond to long and involved questions. However, Endsley pointed out that posttests probably only assess SA at the very end of the trial, as events unfold and earlier misconceptions are addressed. On-line assessment techniques, which layer SA assessment questions on top of the operators' tasks, offer the advantage of real-time assessment but they also draw needed attentional resources away from the task at hand and may actually shape SA by essentially giving the operators hints as to what information is important.

Finally, Endsley (1995b) discussed her preferred technique, the freeze technique. Here, the action is momentarily frozen, and the operators are probed for their SA with regard to a certain prespecified element in the environment. While this has the obvious drawback of being unavailable in "real" settings (one cannot simply freeze a dogfight and ask the respective pilots about their SA), it does allow the experimenter an opportunity to assess SA information in real time. Furthermore, Endsley's (1995b) research showed that the technique did not appear to be intrusive on SA. Her technique, the Situation Awareness Global Assessment Technique (SAGAT), thus appears to offer some advantages in the study of SA.

It must be recognized, however, that the freeze technique and SAGAT are distinct. The freeze technique is a methodology, and SAGAT applies that methodology with a particular type of question. Endsley's research suggests that this content-methodology combination is effective. However, SAGAT can be challenging to deal with from a conventional measurement paradigm, in that the different items included within a SAGAT protocol, as Endsley uses them, tend to be fairly independent; "This means that trying to compile SA queries on different situational aspects into one combined SA variable is not supported" (Endsley, 1998, p. 85.).

There is nothing inherent in the freeze technique, or the proximal approach to SA measurement that categorically precludes the development of fairly unitary scales, however. In fact, there are examples in the literature (e.g., Chaparro, Groff, Tabor, Sifrit, & Gugerty, 1999) where a probe/recall technique has been employed using a composite score approach.

Both on-line assessment and Endsley's freeze technique are representative of memory-probe approaches to SA measurement. Most authors who address the issue of SA measurement include probe techniques in their taxonomies (e.g., Garland et al., 1996). The primary concern with the probe technique is that it disrupts operator performance, offering respite from the time constraint of the SA context and possibly breaking the operator's concentration. Endsley's (1995b) pair of studies on the freeze technique suggested that for breaks of up to 5 or 6 minutes, operators were able to use this technique without adverse affect on memory or performance.

On-line assessment requires much smaller "breaks" than the freeze technique. Therefore, given the finding that a break of 5 to 6 minutes does not affect outcomes such as memory or performance, researchers should also feel confident that the shorter breaks used in on-line assessment will not impair memory or performance either. However, Sarter and Woods (1991) challenged this conclusion, suggesting instead that, despite the claim and aim that these techniques assess what a particular operator is aware of at a given moment in time, these probe techniques are intrusive and in fact can only measure what the operator can recall once removed, even temporarily, from the context.

Subjective approaches to proximal SA measurement. Although most probe techniques, discussed by Sarter and Woods (1991) as intrusive techniques, collect objective data, this is not a requirement of the technique. It is also possible to collect subjective data, such as ratings of performance, using this technique. For example, one way to circumvent the criticism that any measure taken post hoc is biased by outcome knowledge would be to take performance assessments or ratings using a probe technique. Similarly, it would be possible to collect meso-level data subjective data, such as assessment regarding most imminent threats, using a probe technique. Essentially, this simply means that subjective data of any type might be collected during brief breaks in the task, rather than after a simulation or task is completed.

Summary. Proximal measures such as queries are the most common measures of SA (Tenney, Adams, Pew, Huggins, & Rogers, 1992). However, the specific data requested in the probe vary in complexity from spatial location of a single aircraft (e.g., Fracker, 1989) to very complex recall tasks. The central difficulty with this technique is the fear that artificially stopping the simulation for operators to respond to the probe items is disruptive, and will affect the operator's SA. It is possible that this could either artificially enhance operators' SA by briefly relieving the time pressure and giving cues as to important elements, or that it could prove a detriment to SA by interrupting the operators' concentration and train of thought. However, as mentioned above, at least one study (Endsley, 1995b) tested this hypothesis and found that operators' level of SA was not altered by pausing the simulation and probing for SA. Although Endsley's (1989) SAGAT has been discussed above, it is important to remember that SAGAT is but one of a number of measurement protocols based on the probe technique.

Distal Measures of SA

The other major approach appears to take the perspective that outcomes equal SA. This group uses distal indicators of SA such as overall performance, e.g., goals achieved. Despite the earlier debate on whether SA included outcomes such as decision making and performance, this approach has an appealing logic to it. After all, we are interested in SA only to the extent that it contributes to performance. Therefore, variation in SA that is not related to performance is less compelling and perhaps no great loss is suffered if we fail to capture those data. However, this approach is ultimately tautological; performance is treated both as SA itself and as an outcome of SA. That is, an operator has low performance because his or her SA (as measured by performance) is low. In addition, the number of intervening variables introduces a great deal of measurement error into this indicator. While some measurement error is unavoidable, excess amounts obscure the true score of the variable of interest.

Objective approaches to distal SA measurement. The most obvious approach to objective distal SA measurement is to identify outcomes of interest and measure performance against them. For example, one might measure performance in terms of mission goals achieved in military applications, traffic incidents per thousand miles driven for automobile studies, or fatality and damage statistics for emergency management SA. However, this sort of global assessment of performance is the most rife with difficulties.

Clearly, as measures of SA, global measures of performance are subject to large amounts of contamination; innumerable factors, aside from the operators' SA, can affect the overall performance level (Endsley, 1995b).

At least three approaches resolving this difficulty in performance measurement have been discussed in the literature. One method to overcome the difficulty of excessive random error is to restrict the performance evaluation to one or a few key elements that contribute to overall performance. This technique, which Endsley (1995b) refers to as imbedded task measurement, evaluates performance against prespecified subtasks. For example, in an imbedded measurement paradigm, one might assess performance in terms of maintaining a particular altitude or stopping for red lights. This maintains the "naturalness" of global performance measures, while eliminating some of the random measurement error. The primary problem with this technique is that SA information and tasks are highly interdependent, and in focusing on one, the experimenter may miss critical changes in SA on other elements.

The second approach to eliminating the random measurement error that plagues global performance assessment as an indicator of SA is to artificially manipulate the simulation and evaluate performance in terms of perception of and reaction to the manipulated events. This sort of external performance measurement might take the form of a "disappearing" aircraft during a simulation. The operator's perception of and reaction to this event would then be measured. Although this technique has the attractive element of control, Endsley (1995b) pointed out that such artificial manipulation of the scenario may fundamentally change the operator's SA as well as decision making.

Tenney et al. (1992) proposed a third approach to SA measurement using performance. Although this technique also relies on artificial manipulation, there is a key difference from the approach discussed above: where the previously discussed approach relies on manipulations that essentially "trick" the operator, the manipulations proposed by Tenney and her colleagues are intended to test SA by requiring deeper processing to comprehend the important elements of the scenario or requiring the operator to shift attention or focus to perform successfully. By increasing the challenges to SA through manipulation, the theory is that the sensitivity of the performance measure to differences in SA is increased.

In a panel discussion, Vidulich (Garland et al., 1996) addressed the difficulty of mitigating circumstances from another perspective. He simply categorized performance-based measures into measures of *effectiveness* and the more "data-dense" measures of *performance*. The fundamental difference between the two, as discussed by Vidulich, is that measures of performance take into account mitigating circumstances whereas measures of effectiveness do not. That is, in order to be a true measure of performance, mitigating circumstances must be included in measurement. While the increased precision that measures of performance theoretically offer is an advantage, accurately operationalizing mitigating circumstances presents a significant challenge to the SA researcher. Thus, despite the fact that measures of performance theoretically have greater predictive efficacy, their operational challenges may exclude them from practical use. Measures of effectiveness, then, may actually be better than measures of performance as indicators of SA.

Subjective approaches to distal SA measurement. Endsley (1995b) also discussed subjective measures. She divided her discussion between self and observer ratings. The primary difficulty with both methods is the same, however; in neither method does the rater have access to all the necessary information. Self-ratings may be inflated due to ignorance of the true reality. While observers may have the luxury of knowing exactly what the "reality" of the situation is, the observer does not have access to the operator's cognitive experience of it. She pointed out that self-ratings are probably contaminated by performance. Endsley suggested that the demonstrated relationship between self-ratings of SA and performance (Venturio et al., 1989) is more likely due to the fact that, when the operator's gave their self-ratings of their SA, they knew the outcomes of their missions. Thus, they were able to evaluate their SA based on the outcome. From this perspective, it is little wonder that SA and performance have been found to correlate.

However, Endsley (1995b) did not address the fact that Venturio et al. (1989) found that observer ratings correlated more strongly with performance (r = .85) than did self-ratings (r = .60). Although the subjective SA ratings used in the experiment were supposed to be estimated independently of the final engagement outcome, due to the absence of other data, it is unclear whether the observers were biased by the outcomes of the trial, and thus basing their SA estimations on overall performance, or whether they were able to provide more objective estimations of SA than operators themselves. Regardless, the disparity between observer and self-estimations of SA is large enough to warrant further investigation into the foundations for the differences; indeed, the observer estimations of SA accounted for approximately twice as much of the variance in performance as the self-estimations.

Sarter and Woods (1991) took a simple approach to categorizing attempts at measuring SA. They dichotomized SA measurement into two types: intrusive assessment techniques and after-the-fact data collection. Intrusive techniques are those like Endsley's SAGAT, and after-the-fact methods are those like pilot debriefings. Each has its own distinct flaws. As mentioned earlier, despite their intention to measure an operator's "real time" SA, probe measures may only capture the information to which an operator has access when extracted from the situation. Similarly, after the fact measures rely on context-deprived recollections. Thus, although posttests offer the advantage of permitting subjects to respond to long and involved questions, as discussed earlier, posttests have a critical flaw in that they rely on assessments of SA from only one point in the trial. Furthermore, this single assessment is based on the very end of the trial, after the end results are often known.

Summary. Contrary to Endsley's (1995b) assertion that performance-based measures reveal little about an individual's SA, Tenney and her colleagues suggested that "much can be revealed about the situation awareness of the crewmember through performance measures alone." (Tenney et al., 1992, p. 13). However, Tenney et al. went on to clarify that this is the case only when the demands on attention can be manipulated.

Meso-Measures of SA

Even given this varied menu of measurement techniques, it seems that each leaves something to be desired. One additional avenue to pursue in measuring SA is to revisit the promixal/distal debate and attempt to find, quite literally, the middle ground. Middle range indicators (meso-indicators) of SA should be more appropriate and perhaps more precise as indicators of SA than either proximal or distal measures. Meso-indicators should suffer neither the myopia of proximal measures nor the random error of distal measures.

Objective approaches to meso SA measurement. Objectively attempting to capture the whole of SA is a trickier proposition than simply attempting to objectively measure SA's inputs or outputs. The most complex of these is the model-based approach to SA measurement. Tenney et al. (1992) suggested the use of human performance models that operate in parallel to a crew of human operators in a flight simulation. During a simulation, the parallel system collects complete data on the crew's activities. From this record, a performance model is developed. In essence, the researchers suggested developing an artificial intelligence model based upon the crews' actions and reactions to environmental events. Of course, since these data are collected in a simulation, the universe of external events are also recorded, and the datasets are integrated into the performance model. At the end of the development process, the experimenters would then have a complete model of each crewmember's "consciousness" at every given point in time throughout the simulation. Although this is clearly a highly complicated and sophisticated technique, if achievable, the resultant model could yield volumes of SA information to researchers. More importantly, it would yield information not only about the inputs to the operators' SA and the outcomes of it, but of the SA itself. The goal of this technique is to yield information on the moment-to-moment perceptions and processes during the simulation.

Clearly, there are some technical challenges in this model-based approach. In the air traffic control domain, one way to gain the richness of data offered by the model-based approach to SA research, without the technological challenges is through a method proposed by Rodgers and Duke (1994): Situation Assessment Through Re-Creation Of Incidents (SATORI). Here, as the name suggests, real incidents are re-created for review and analysis. Specifically, the SATORI system records both radar and verbal air traffic control information. This information is then synchronized and played back, or re-created, at a later time. One key advantage is that these are natural data rather than simulation data, with all the richness that natural data provide. Because the method captures all of the data available to the air traffic controller, in the sequence in which it was made available, as well as the communications from the air traffic controller, there is significant opportunity to capture SA data. While the method was developed specifically for use in air traffic control, it offers promise for use with other systems, particularly those that are largely computer based, such as nuclear power operations.

Despite the appeal of these model-based approaches to SA measurement, there are substantial barriers to their implementation. First, both operator and environment data must be reliably captured. Second, those data must be integrated into a coherent whole (not unlike the achievement of Level 3 SA). Finally, due to the highly specialized nature of the models that are created, the results of such studies may not be generalizable.

There are, however, several methods by which holistic meso assessments of SA can be made which do not require the technological sophistication of the model-based approach or SATORI. There are at least three objective approaches to meso SA measurement that are not so technologically complex. Each of these in some way taps higher-level SA (Level 2 or Level 3). First, the operator's response time and accuracy in assessing situations can be measured. Second, there are some direct measures of information seeking that can be indicators of higher-order SA. Finally, display recall can be used as an indicator of higher-order SA. Each of these are discussed in greater detail below.

The first of these, speed and accuracy in assessing situations, gets at the core of SA in many regards. The defining feature of Level 2 SA is integration of elements in the environment into a cohesive whole. Level 3 is the level of judgements and planning for future events. Thus, an operator's ability to correctly and swiftly assess the normalcy of a given situation can be taken as an indicator of higher order SA. Tenney et al. (1992) proposed using this technique as a direct measure of SA. The drawback here is that the technique is only applicable in controlled settings, where one outcome or situation assessment can be objectively defined as "correct," and any other judgement can be objectively characterized as "incorrect."

The second of these less technologically advanced objective measures of SA at the meso level is information seeking. Tenney at al. (1992) suggested that information seeking is a viable SA metric. It is only through the processes of information integration and planning for future events that operators can generate questions and test hypotheses. Although the perception of information is an element of Level 1 SA, information *seeking* is driven by higher level SA. That is, the operator's current understanding of a situation will drive her or his efforts to seek confirmatory or contradictory evidence. Thus, information seeking could be used as an indicator of higher-order SA.

Tenney et al. (1992) discussed two approaches to the measurement of information seeking as an indicator of SA: making all information available only by request and recording eye-movements. Using the technique of making all information available only by request actually includes elements that are both subjective and objective. Here, the operator is effectively isolated; all data are provided only on request. Thus, the number and content of the operator's requests can provide good insight into her or his current model of the situation, as well as providing a perfect record of exactly what information that model is based upon. It is also possible that this concept could be applied to an operator's or team's communication pattern, such that information seeking could be seen as an indicator of team SA. However, the delays involved in actively requesting all information, as well as the lost opportunity for serendipity seriously challenge this as a useful method for quantifying SA. The "real-time" quality of the SA context suggests that the timing delays inherent in such a system may compromise generalizibility.

Real-time measures of information seeking have been discussed, however. A more objective, but possibly less accurate meso-level SA metric is recording eye movements during a real-time simulation. Here, a device records exactly what the operator looks at, and for how long. However, this method is not without its flaws. While it provides very detailed data about what elements in the environment are receiving the operator's visual attention, there is no information about whether the information actually received the operator's attention. Furthermore, the metric does not clearly delineate among the levels of SA, thus rendering "muddled" data. With this technique, the researcher cannot distinguish whether attention paid to a given element is indicative of mere perception of that element, suggesting Level 1 SA, or due to information seeking as a part of Level 3 SA. That is, it is not possible to establish comprehension or strategy on the basis of eye movement data. Thus, this method does not seem particularly appropriate for judging information seeking per se. However, if it is assumed that eye movement tracking captures only data related to perception, it may function as a good measure of proximal SA to the extent that environmental data are often visual in nature.

The third objective approach to meso-level SA measurement involves display recall. One of the classic studies in cognition establishes the link between pattern recall and meaningfulness of the pattern (Chase & Simon, 1973). In essence, when patterns are meaningful, the pattern is the unit of recollection, rather than the individual element. Meaning is layered onto "raw" perceptual SA data at Level 2 SA. Thus, recall of displays should vary as a function of meaningfulness, and thereby be a good indicator of Level 2 SA (Tenney et al, 1992). Tenney et al. treat display recall as a category separate from the simple queries addressed in the earlier section on proximal measurement. Thus, of the objective approaches to meso-level SA measurement, this seems the most promising. However, the success of this measure depends upon some reliable way to quantify display recall accurately. Operationalizing display recall poses significant challenges to the SA researcher.

Subjective approaches to meso SA measurement. Subjective measures are ideally suited for quantifying SA at the meso-level. The central difficulty in measurement of SA itself, which led to the proximal and distal approaches in the first place, was that SA was difficult to observe directly. However, given the flaws inherent in proximal and distal measurement strategies, there is an appeal in attempting to capture SA itself, and subjective ratings of SA as a whole seem an attractive means by which to achieve this. Subjective ratings of SA are a promising measurement approach to the extent that the ratings are based on sound theory.

Several authors have taken this general perspective and proposed subjective SA measurement instruments. Carretta et al. (1994) proposed an 8 dimension, 31-item rating scale for use in assessing SA (see Appendix A). The Situation Awareness Rating Scale (SARS) is designed for use both as a supervisory rating scale and a self-rating scale, and assesses such things as "time-sharing ability," "communication quality," and "defensive reaction." The SARS is scored relative to other F-15C pilots, from "acceptable" to "outstanding." Unfortunately, Carretta and his colleagues have not published any data on the reliability, validity or predictive efficacy of the measure that they propose. In addition, SARS is specifically written for combat pilots, thus limiting is applicability to other populations that require SA.

Five years earlier, Taylor (1989) developed an objective rating scale that is both more concise than Caretta et al.'s (1994) SARS and more broadly applicable (see Table 1). Taylor's 10-D Situation Awareness Rating Technique (SART) relies on 10 empirically derived items, such as "complexity of situation" and "concentration of attention." SART is scored from low to high on each of the items. Interestingly, Taylor's scale captures both the level of challenges present in the environment and the operator's ability to meet those challenges. Selcon and Taylor (1989) have shown preliminary evidence that SART is a valid measure of SA. Endsley's (1998) comparative study of SAGAT and SART suggested that SART was highly correlated to a rudimentary measure of SA, although no relationship was established between SAGAT and SART.

One meso-level measurement strategy is available that represents a fairly dramatic shift from the approaches discussed thus far, in terms of the content that is evaluated. Some behaviors are indicative of SA, and to the extent that they can be codified and captured, behavioral indicators hold promise for fruitful SA measurement. Situational Awareness Linked Indicators Adapted to Novel Tasks, or SALIENT (Muniz, Stout, Bowers & Salas, 1997), is a set of behaviorally based indicators of SA (see Table 2). SALIENT holds promise for quantitatively assessing SA as a meso-indicator.

Table 1
Situation Awareness Rating Technique (Taylor, 1989)

Dimension	Item
Demands on attentional resources	
	Instability of situation: Likeliness to change suddenly
	Complexity of situation: degree of complication
	Variability of situation: number of variable/factors changing
Supply of attentional resources	
	Arousal: degree of alertness: readiness for activity
	Concentration of attention: degree to which thought is brought to bear
	Division of attention: distribution/spread of focus of attention
	Spare mental capacity: mental ability available for new variables
Understanding of situation	
	Information quantity: amount of knowledge received and understood
	Information quality: goodness or value of knowledge communicated
	Familiarity: degree of prior experience/knowledge

SALIENT (Muniz et al., 1997) is a list of behaviors proposed to be indicative of good SA (see Table 2). These behavioral indicators were drawn from the SA literature. In their paper outlining SALIENT, Muniz et al. utilized a five phase process to arrive at a final checklist or observation form based on SALIENT. They first identified behaviors that indicate high levels of SA; their list of behaviors is displayed in Table 2. The second phase involved developing scenario events that would allow for the display of SA-related behaviors. Third, Muniz and her colleagues identified specific behaviors that are based on the behavioral indicators identified in Phase 1. For example, in this phase, "reporting problems" was operationalized as "team member verbalized they have lost radio contact with AMC (air mission commander)" (Muniz et al., 1998, p. 11-5). In phase four, Muniz and her colleagues developed a script for experimenters to use in administering SALIENT. Finally, the fifth phase of SALIENT development yielded an observation form for use in capturing SA data. Here, the specific behaviors that were believed to indicate high levels of SA in each segment were listed on a checklist. The checklist includes a column next to each specific behavior where the experimenter indicates the presence of each specific behavior.

One of the limitations outlined by Muniz and her colleagues (1997) addresses this behavioral specificity. They acknowledge that the identification of the "best" responses during each scenario segment is a long and arduous process. Furthermore, they acknowledge that some of these specific examples of the behavioral indicators "may not occur naturally" (p. 11-6), and that alternative responses are also possible. Thus, the level of specificity offers both advantages and disadvantages.

It is possible, therefore, that the benefits of their behavioral focus may be obtained through less constrictive means than the specific behaviors outlined in their full methodology. That is, it may be constructive to use their list of behavioral indicators as a traditional behavioral checklist, keeping the general categories intact. The development process speaks to the content validity of these dimensions. Applying a more generalizable rating or checklist approach to that content may be one means by which to assess SA in a wider variety of contexts or experimental settings than might be possible with the full SALIENT instrument as described by Muniz and her colleagues.

In addition to the measurement strategies outlined above, however, it is possible to take middle-range subjective measures of SA in forms other than rating scales. Notably, it is possible to apply the probe technique, the most used in SA of all measurement techniques (Tenney et al., 1992), to meso-measurement. Tenney and her colleagues were

Table 2
Situational Awareness Linked Indicators Adapted to Novel Tasks (Muniz et al., 1997).

Dimension	Item
Demonstrated Awareness of Surrounding Environment	
	Monitored environment for changes, trends, abnormal conditions
	Demonstrated awareness of where he/she was
Recognized problems	
	Reported problems
	Located potential sources of problem
	Demonstrated knowledge of problem consequences
	Resolved discrepancies
	Noted deviations
Anticipated need for action	
	Recognized a need for action
	Anticipated consequences of action and decisions
	Informed other of actions taken
	Monitored actions (self & others)
Demonstrated knowledge of tasks	
	Demonstrated knowledge of tasks
	Exhibited skilled time sharing attention among tasks
	Monitored workload (self & others)
	Shared workload within station
	Answered questions promptly
Demonstrated awareness of information	
	Communicated important information
	Confirmed information when possible
	Challenged information when doubtful
	Re-checked old information
	Provided information in advance
	Obtained information of what is happening
	Demonstrated understanding of complex relationships
	Briefed status frequently

somewhat disparaging of the type of information collected in many applications of the probe technique, such as one's altitude at a given time. They suggested that, if data collected via this technique are to have more than face validity, researchers must focus on such elements as goals, diagnoses, and crew responsibilities and knowledge. Clearly, these data are indicative of higher order SA. Thus, utilizing a probe protocol with content in these areas would yield SA data at the meso level.

Summary. Meso indicators of SA are promising in that they avoid the major challenges faced by proximal and distal SA metrics. Specifically, they are not subject to the contamination that plagues distal measure of SA, nor do they suffer from the restriction of proximal measures.

Other Approaches to SA Measurement

It has been suggested that generalized intelligence (*g*), psychomotor skills and conscientiousness may be useful as predictors of SA (Caretta et al., 1994). Although these clearly represent a different approach to dealing quantitatively with SA than the techniques discussed above, there is merit in considering the potential efficacy of these fairly universal predictors in SA research. Given that SA is a cognitive construct, separate from performance, it seems as if psychomotor skills might be less directly related to SA (although it is likely related to overall task performance). However, it is impossible to determine empirically the efficacy of any predictor of SA until adequate methods for quantifying SA itself can be established.

Finally, some qualitative methods of data gathering have been proposed for use in SA research. Baker, Stout, Salas, Fowlkes, and Cannon-Bowers (1998) proposed using structured interviews and guided verbal reports. Their research suggested that these two methods yielded reliable results that hold great potential utility. Unfortunately, due to the qualitative nature of these data, it is not possible to assess the validity of the measures. Baker and his colleagues emphasize that, while these methods yield rich data that are consistent across techniques, it is important to recognize their limitations. Ultimately, as with qualitative techniques in general, these techniques are most useful for revealing knowledge, rather than establishing the amount or extent of knowledge.

Summary

Despite the volumes of attention paid to different approaches to capturing SA quantitatively, to date there is no authority greater than any one author's opinion as to the most efficacious method of SA measurement. That is, most authors' efforts seek to establish the legitimacy of their own methods, rather than to demonstrate their merits relative to other available methodologies. Thus, one of the goals of this work is to identify the efficacy of a variety of indicators of SA.

Teams

The team literature is substantially larger and better developed than the SA literature. In recent years, team research has evolved considerably. Given the nebulousness of the concept of "team," and some of the controversies in the team literature, it is important to first define teams, and explore some team taxonomies that are relevant to the proposed research. In addition, it is important to consider team processes, and to address the relevance of these concepts to team SA. Finally, several models of team SA are considered.

Definitions

Before we can begin a meaningful discourse on teams, it is useful define the term. One of the frequently cited definitions of teams is "a distinguishable set of individuals who interact interdependently and adaptively to achieve specified, shared and valued objectives" (Morgan, Glickman, Woodard, Blaiwes & Salas, 1986, p. 3). Although brief, there are some important elements in this definition.

First, a team consists of a distinguishable set of multiple individuals. This essentially speaks to team composition. Thus, it must be possible to distinguish both who is, and who is not, a member of a given team. In the same vein, one must establish some parameters for team size. One of the controversies in the team literature is how many people are required to classify multiple individuals working together as a "team." To those researchers whose primary interests are in certain social processes (e.g., coalition formation), three people are required for a group of individuals to be classified as a team. Other models suggest that two individuals are sufficient. Given that minority/majority influence and coalition formation are not included in the models discussed in this dissertation, two individuals will be considered a team for the purposes of this research.

Second, teams work interdependently. This second element addresses team processes. This interdependence means that there is an element of reliance and shared fate among team members. While multiple individuals can work "together" in time or space towards the same or similar goals, unless the element of interdependence is present, those individuals would not be classified as a team.

However, this interdependence can come in several forms. It is possible for teams to be interdependent in either their processes or in their outcomes, or in both. Each has a different implication for the level of interaction required. Process interdependence requires more interaction on the part of team members, as the very way in which they achieve their goals is intertwined. Outcome interdependence, on the other hand, requires only that each team member "carry their own load" and contribute their portion to the final team outcome. While these teams require fewer interactions among the team members, none can succeed while another fails.

The last element of this definition speaks to a cognitive aspect of the definition of teams. Here we recognize the importance of goals and goal-directed activity. Thus, interdependence is a necessary but not sufficient element of teams. Symbiotic organisms are interdependent, but are in pursuit of differing goals and therefore could not be classified as a team.

Finally, it is important to address some of the social processes associated with teams and effective teamwork. The necessity of coordinating multiple individuals each of whom is likely to have unique information and task priorities can be challenging. However, successful team performance requires coordination.

Types of Teams

Clearly, the definition above leaves considerable room for variability. Meaningful discussion on teams requires more specificity with regard to the particular defining characteristics of the teams under consideration. Two taxonomies are presented that contribute clarity to the definition of teams. The first is noteworthy for its elegance and parsimony in making distinctions among types of teams on the basis of their central features, within a fixed universe of features. The second is presented for its critical analysis of the differences among decision making teams. Although we have discussed SA as a construct distinct from decision making, it is important to bear in mind that teams which require SA are also required to make decisions. Thus, discourse on decision making teams is relevant.

Basic Types of Teams

While numerous taxonomies have been presented, one in particular clearly lays out the key elements that distinguish among major types of teams. McGrath, Berdahl and Arrow (1995) suggested that work groups (teams) are distinguished by relative emphasis that is placed on one of three key elements: members, tasks or tools. Thus, a given work group is characterized most by the individuals who comprise it, the specific tasks the work group performs, or the tools and technologies it employs to attain its goals.

These three key elements are not distinct, but overlap. The elements at the intersections of these primary elements are also interesting. At the intersection of members and tasks we find the division of labor as an important element. Job structure is at the intersection of tasks and tools, and the intersection between tools and members is defined as the role network.

The model presented by McGrath et al. (1995; see Figure 1), addresses three different types of mutually interdependent work groups with singularity of purpose. The defining characteristic when distinguishing among these is the primacy of one of the three key elements of members, tasks or tools. The remaining elements, then, are less central and less immutable.

The first type of work group outlined by McGrath et al. (1995) is a team. According to their taxonomy, the primary component of a team is the individuals involved. The secondary element of a team is the technology that it uses, and the least central component is the particular task at hand. When primacy is put on the individuals involved, teams will evolve to accomplish different tasks and use different technologies as necessary.

The focal element of a team is a relatively stable population; the team members do not change drastically over time.

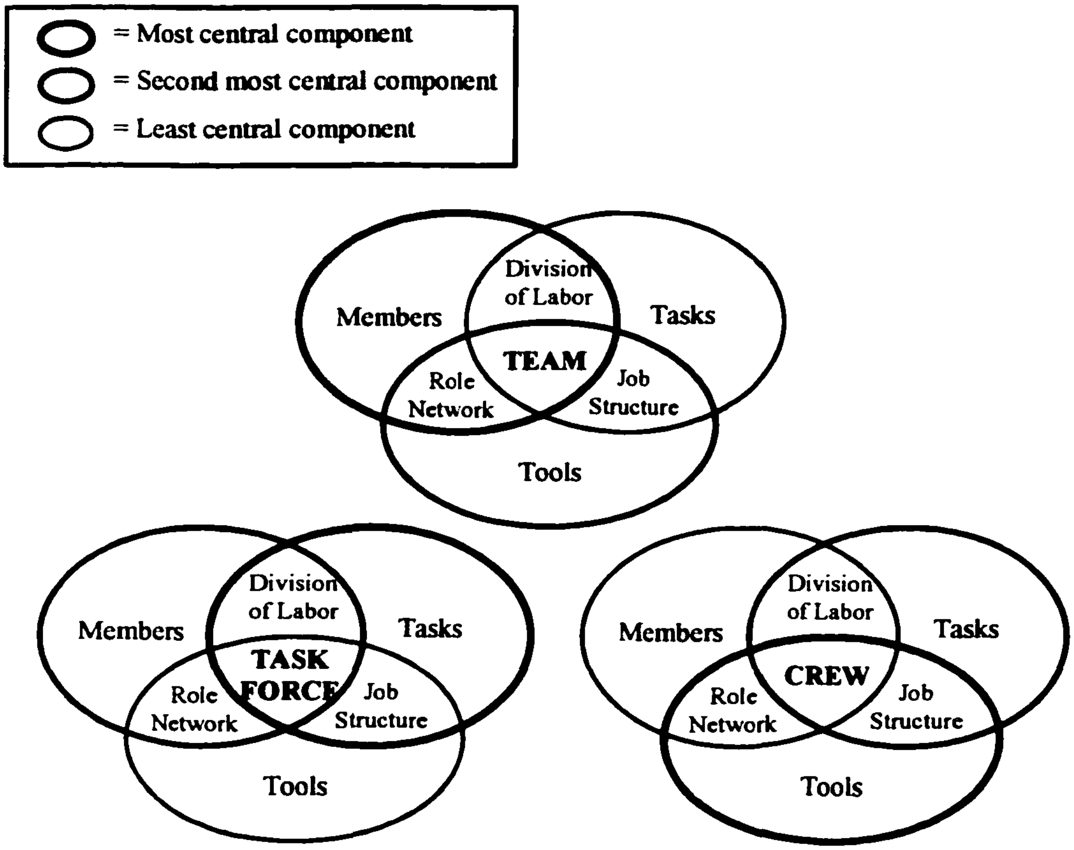


Figure 1. Three types of work groups.

McGrath, J. E., Berdahl, J. L, & Arrow, H. (1995). Traits, expectations, culture and clout: The dynamics of diversity in work groups. In S. E. Jackson & M. N. Ruderman (Eds.), *Diversity in work teams* (pp. 17-46). Washington, DC: American Psychological Association. Copyright 1995 by the American Psychological Association. Reprinted by permission.

Following the hierarchy of elemental importance in each type of work group, the intersection elements discussed above also vary in importance in the different types of work groups. In a team, the role network and division of labor are the most important elements. In a task force, the division of labor and the job structure are the most important elements. In a crew, the job structure and role network are the most important elements.

Decision Making Teams

In general, teams are formed with a purpose or goal in mind. Often, decision making is a significant component of teams' activities. Many different types of teams must make decisions in their daily function. McGrath et al. (1995) presented some key differences among major types of work groups. However, considerable room for refinement in the definition of teams and the distinctions among them remains. Klimoski and Jones (1995) present five types of decision making teams. The types of teams presented by Klimoski and Jones (1995) are: command and control teams, production teams, customer-service teams, professional/technical decision making teams, and executive teams. Each is distinct from the others, and each faces unique challenges. Table 3 outlines some of the key differences among these five types of teams.

Command and control teams are characterized by highly specialized and interdependent jobs. This means that team members cannot "pick up the slack" if one member fails to perform adequately. Command and control teams generally perform highly coordinated tasks in response to the environment. Generally, their decisions must be made in real time, and consequences for decisions are immediate.

Production teams, autonomous production teams in particular, are characterized by broad responsibilities and flexible interdependence. While broad responsibilities and skills allow flexibility in goal accomplishment, this flexible interdependence increases the complexity in team coordination. Unlike command and control teams, where task responsibilities are clearly laid out as a consequence of specialization, these production teams must address the matter of who will perform what functions, in a team where several may be able to perform each task.

Table 3
Types of Decision Making Teams (Klimoski & Jones, 1995).

Team Type	Coordination	Decisions	Specialization	Interdependence
Command & Control	Highly coordinated	Real time with immediate consequences	Highly specialized	Highly interdependent
Production	Need to coordinate cross-functional interdependence	Real time, immediate or long-term consequences	Broad responsibilities	Flexible interdependence
Customer Service	High need for coordination	Complex, based on perception of client needs, feedback is not consistent in form or timing	Flexible, broadly skilled	High interdependence
Professional	(varies)	Varies widely, significant impact	Highly specialized	(varies)
Executive	(varies)	Poorly defined problems, significantly delayed feedback	(varies)	(varies)

Third, Klimoski and Jones (1995) addressed customer service teams. They discussed the “handoff” approach to customer service, where one individual acts as a initial point of contact, and then “hands off” the situation to another employee or group of employees. This is a highly interdependent process and successful execution of this process requires a great degree of coordination. Furthermore, the decision tasks faced by customer service teams are complex. The stimuli upon which decisions are based may be subtle or subjective. In addition, the feedback is not consistent, taking many forms and operating on a variable timeline from immediate to significantly delayed.

Professional and executive teams are more variable than the previous three types. Klimoski and Jones (1995) suggest that executive teams are a type of professional team. The primary defining characteristic of both of these types of teams are their decisions. Professional teams come in many permutations, and thus, their decision tasks vary considerably as well. However, one constant is that their decisions generally have significant impact. The situation is more complex for executive teams. They tend to be faced with poorly defined problems, and feedback on their decisions may be delayed by months or even years.

Teams that Require Situation Awareness

The complexity and constraints presented by the SA context often preclude a single individual working alone from performing successfully. The demands are too high and the operations are too complex. Thus, teams are often required to successfully operate in the SA context. While multiple individuals working together increase the resources that the unit can devote to operations in the context at hand, this also presents new challenges of communication and coordination. The teamwork required of teams in contexts that require is highly sophisticated.

Teams that operate in contexts where SA is critical must demonstrate exquisite coordination. The time-constraint and risk often associated these contexts leaves little room for error or confusion. While a good deal of this can be established in advance (as in a mission brief, or through extensive experience as a team), the complexity of these contexts means that unexpected turns of events are a distinct possibility.

Acquiring and maintaining good SA places some additional demands on those teams which need high levels of SA to function maximally. Team SA, then, falls at the intersection between two complex topics. Although a few perspectives or theories on team SA have emerged, it is important to first discuss some of the unique characteristics and demands that teams which must operate in the complex environments that require SA must face.

Several authors have presented taxonomies of teams, and it is clear that some types of teams are more likely than others to find themselves in contexts that require SA. These contexts are often very technically demanding. Operators must be expert at their respective jobs to the extent that they can detect and analyze even minor deviations from normalcy that might indicate a problem down the road.

Following McGrath et al.'s (1995) taxonomy, then, the SA context is most likely to be a factor for crews, as opposed to teams or task forces. This has interesting implications for SA research. According to McGrath et al.'s (1995) model, the technology or tools are the most important element in a crew, followed by the task at hand. The members of the crew are the least central component.

Let's review a context that requires high levels of SA against this prescription. Taking surgery as a prototype, we find that the technology at hand is the central or defining element of the group's existence. Secondary to that, we find a particular task at hand, in this case a particular operation. The element of tertiary interest is most subject to change. Indeed, we find that several physicians may execute different elements of the surgical procedure, with little disruption in the overall process. It is far more challenging to imagine the team of surgeons shifting from one technology to the next (for example, collectively abandoning surgery to pursue commercial fishing), than it is to imagine replacing on member of the surgical team with another.

This characterization of teams that require SA as crews under the McGrath et al. (1995) model, also has interesting implications arising from the intersection elements. In a crew, the job structure and the role network are the two most important intersection elements. This suggests that teams (or crews) facing an context that requires SA must address these fundamental issues for successful performance.

Following Klimoski and Jones's (1995) taxonomy of team types, different attributes become important for teams likely to find themselves in contexts that require SA. According to this taxonomy, command and control teams are the most likely to require SA in their operations. These teams were characterized by high levels of coordination, real time decisions with immediate consequences, high levels of specialization and high levels of interdependence. Thus, it can be inferred that these attributes would also be important for teams in the SA context.

Together, these perspectives paint a fairly clear picture of the characteristics of a team in a context that requires SA. The high level of specialization in command and control teams is a requirement of the technological focus of a crew. The time-constrained nature of the SA context drives the requirement for real-time decisions with immediate consequences. The high levels of coordination and interdependence are means by which to meet the need for job structure and clarity in role networks.

Team Processes

To this point, our discussion of teams has been primarily descriptive. If we wish to be prescriptive with regard to teams, for example, to improve team effectiveness, it is important that we be able to understand team processes. From understanding, we can move to prescribing mechanisms and techniques for improved performance. Two perspectives that hold promise for improving team performance via process prescriptions are discussed: team decision making and role theory. In addition, several behaviors that contribute to team effectiveness have been identified.

Team Decision Making

Having differentiated among several types of decision making teams, it is important to examine the important elements that contribute to good team decision making. It goes without saying that team decision making will be similar to and different from individual decision making in important ways. While several good individual decision making models exist, one team decision making model is particularly clear in the constructs it proposes, the linkages among those constructs, and the empirical evidence to support the validity of the model. This is Hollenbeck, Ilgen, Sego, Hedlund, Major and Phillips' (1995) hierarchical team decision making model.

Hollenbeck et al.'s (1995) team decision making theory was developed for one specific type of team, although it may be applicable to other varieties as well. Specifically, their model addresses decision making in hierarchical teams with distributed expertise. These are command and control teams with two primary distinguishing factors: status differences among team members and differential expertise among team members. Although one leader in these teams has the final decision authority, each team member brings unique expertise. Thus, the leader must gather, integrate and act upon information from each of these distinct information sources.

Hollenbeck et al. (1995) explicitly recognize the importance of distinct levels of the decision making process: decision level, individual level, dyadic level and team level. However, the bulk of their work focuses on the core team level constructs of team informity, staff validity and hierarchical sensitivity. Although, in general, these have analogues at the other levels (decision, individual and dyad), it is the team level constructs that are of interest for purposes of the present paper.

Team informity addresses the extent to which the team as a whole is aware of all relevant information. Staff validity, Hollenbeck et al.'s (1995) second construct, addresses the accuracy of team members' judgements. Finally, hierarchical sensitivity assesses the team leader's accuracy in weighting the inputs of the various team members in arriving at a single team decision. The central thesis in this theory is that these three core team level constructs are the most direct causes of accurate team decision making.

Although their effects were not directly tested in Hollenbeck et al.'s (1995) piece on the hierarchical team decision making model, the model also includes six non-core constructs, believed to influence decision making accuracy indirectly, through the core constructs. Each of these six non-core constructs influences one or two of the core constructs. Specifically, Hollenbeck et al. suggest that roles influence both staff validity and hierarchical sensitivity. The social environment, a non-core construct, is hypothesized to affect hierarchical sensitivity. The behavior setting (physical proximity) affects both hierarchical sensitivity and team informity. The physical/technical environment is also proposed to affect team informity. Tasks are thought to affect both team informity and staff validity. It is suggested that factors within the person influence staff validity as well. These concepts and the relationships among them are displayed in Figure 2.

Empirical research by Hollenbeck and his colleagues (1995; Hollenbeck, Ilgen, LePine, Colquitt, & Hedlund, 1998) supports the primary contention that the three core team level constructs (informity, validity and sensitivity) are the most direct causes of accurate team decision making. In the seminal piece on this theory, Hollenbeck et al. (1995) report two studies. The first study by Hollenbeck et al. (1995) examines the contributions of the core constructs over a series of six task sessions, each of which was three hours in duration. The second study by Hollenbeck et al. (1995) relies on a more typical team research paradigm, examining the results of a greater number of participants, each whom made a much smaller number of decisions. Thus, each of the studies drew its statistical power from a different source. The first study by Hollenbeck et al. (1995) relied upon many data points for a relatively small number of individuals (within-subject variation), and the second study relied upon relatively few data points for a substantially larger participant group (between subject variation). Both studies relied upon university students who received training in a simulation task.

Study 1 by Hollenbeck et al. (1995) revealed that the core constructs accounted for nearly 50% of the variance in team performance. The lion's share of this variance came from team informity and the interaction between staff validity and hierarchical sensitivity, each of which accounted for about 20% of the variability in overall team performance. Furthermore, the constructs were related to task success in two other interesting ways. First, teams that were high on all three of the core constructs outperformed those that were low on all three core constructs by about two standard deviations. Second, teams that were low on all three of the core constructs were five times more likely to experience a disaster than those that were high on all three.

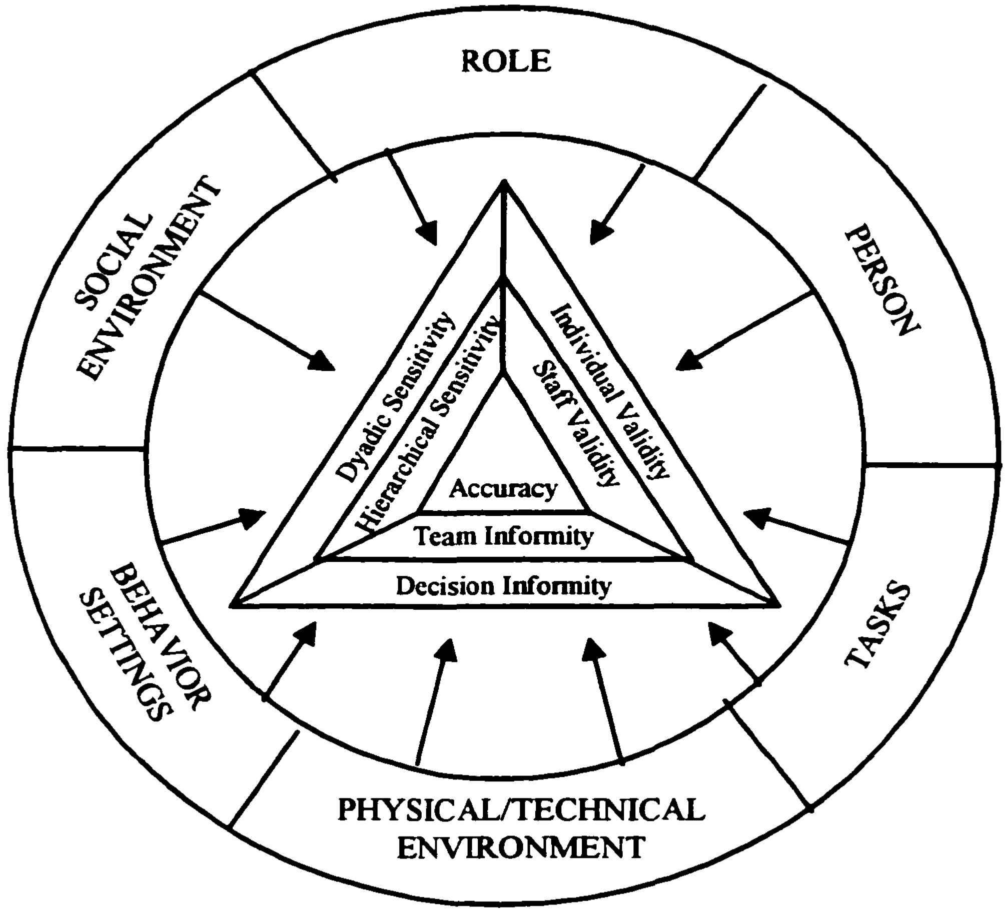


Figure 2. Overview of hierarchical team decision making theory

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*, 292-316. Copyright 1995 by the American Psychological Association. Reprinted by permission.

In Hollenbeck et al.'s second (1995) study, this finding was even more pronounced; none of the 14 teams that were high on all three characteristics experienced a disaster, whereas the teams that were low on all three core constructs experienced a 3% disaster rate. This second study also revealed the importance of team cohesiveness in contributing to team success; team cohesiveness was the only non-core variable to significantly predict all three of the core constructs, as well as overall team decision accuracy.

More recently, Hollenbeck et al. (1998) have shown that, together, these core constructs account for 63% of the variance in team decision accuracy. In terms of unique variance, the lion's share was contributed by team informity (32%) and staff validity (23%).

There are several lessons to be taken from these findings. First, the ability to collect and distribute information is important for accurate decision making. Second, it is important for teams to have members who can integrate raw data into meaningful and accurate judgements. This was seen as especially important, given the distributed expertise within the teams; substantial benefit was gained from team members processing and integrating information prior to passing it along to another team member who may not have been expert in the topic at hand. Third, it is important for accurate decision making that information from various sources be weighted accurately. Furthermore, within-team variance on these constructs can be substantial, particularly for informity.

Finally, these studies yielded some important observations in the area of communication. Although communication data were not collected as a central variable, the nature of the task used by Hollenbeck et al. (1995) permitted them to collect extensive communication data. While the objective quantity of communication did not differ between high and low performing teams, Hollenbeck et al. made two interesting observations regarding the nature of those communications.

First, Hollenbeck et al. (1995) noted that high and low performing teams differed in the extent to which they were proactive or reactive in their communications. High performing teams in their studies were much more proactive, sending unsolicited relevant pieces of communication to their teammates. Low performing teams, on the other hand, were reactive in their communication patterns; rather than proactively sending out information, these teams devoted much of their communication effort to asking one another questions. Furthermore, in the low performing teams, many of the questions went unanswered. Thus, fewer of the low performing teams' communications actually yielded information.

The second interesting observation made by Hollenbeck et al. (1995) was that high and low performing teams differed in the extent to which they processed and added meaning to information prior to communicating it. Where low performing teams tended to pass along information in a raw data form, high performing teams processed the data prior to communicating it. Then, the individuals in high performing teams tended to pass along the processed information in a form that was meaningful to the recipient of the communication.

Decision Making and Team Situation Awareness

The high demands placed on teams by the contexts that require high levels of SA require very effective decision making strategies. Those decisions that can be planned in advance (programmed decisions, Simon, 1960) offer a reasonable expectation for success, given their developed structure, clarity of criteria and viable alternatives. However, the advance planning that characterizes programmed decisions is not always available. Nonprogrammed decisions are those that must be made in the moment, without the benefit of advance planning. In the face of nonprogrammed decisions, often necessary in situations that are novel and/or ill-defined, the likelihood of success is an unknown quantity. Critical cues may be missed, and time and resources may be insufficient to thoroughly explore alternative courses of action.

This state of affairs leads to a bounded rationality with regard to decision making. When unusual situations arise in the SA context, teams must select a course of action and initiate it very quickly. Thus, the team must use effective communication, and rely on a common understanding of the situation and each team member's part in the solution. Thus, team processes must be adequate to perceive, process, plan and perform smoothly in a very short time span, with limited information and resources.

The fact that many decisions within an SA context will be made using bounded rationality makes high levels of SA all the more important. When crises strike, teams do not have the luxury of reviewing all the possibly important information categories, initiating a search of the environment, and synthesizing information from the environment search prior to reaching a decision. Rather, they must make their plans on the basis of the information that is available at the time. It follows, then, that the better the quality of that information, the better the decisions that will come out of the situation.

This argument returns us to the importance of SA at levels 1, 2 and 3, and the core constructs in Hollenbeck et al.'s (1995) hierarchical team decision making model. Hollenbeck and his colleagues have demonstrated the contribution that informity makes to accurate team decisions. Although we have established that decisions are not a component of SA, decision accuracy is a consequence of SA. Major and Fink (1998) have argued that informity and Level 1 SA are analogous constructs. If level 1 SA and team informity are analogous constructs, and team informity has been empirically demonstrated to predict team decision accuracy, then it follows that Level 1 SA has a significant contribution to team performance. This rebuffs the arguments that probe measures of SA that assess operator's knowledge of elements in the environment (e.g., altitude) are high in face validity but impoverished as measures of SA. To the extent that informity and Level 1 SA overlap, Level 1 SA is indeed an important contributor to decision accuracy, a consequence of SA. It is difficult to conceive of a means by which an element of SA could contribute to a consequence of SA without influencing overall SA levels.

Role Theory

One of the challenges in effective teamwork is distributing the tasks such that each is done most efficiently, none are omitted, and the level of redundancy is appropriate. An equally important, but often overlooked, challenge is establishing not only who will attend to the explicit task elements, but also who will attend to the implicit or emergent elements of the team's work. Role theory addresses these challenges.

Defining "role." As with many terms that are used both in the scientific literature and common parlance, it is important to define the term precisely prior to engaging in any discussion on the topic. A "role" as it is discussed in role theory is a pattern or set of behaviors that are expected of a particular individual (Biddle, 1979). Ilgen and Hollenbeck (1991) further clarify that roles are cognitions, and therefore exist only in the minds of people.

It is important to make the distinction between "job" and "role." The essential difference is that roles are more fluid than jobs. Where a job will consist of established task elements that exist independent of the incumbent, a role is comprised of emergent task elements that are "subjective, personal, dynamic and specified by a variety of social sources" (Ilgen & Hollenbeck, 1991, p. 174). On the other hand, Ilgen and Hollenbeck (1991) define jobs as objective, bureaucratic, static and specified by a prime beneficiary. Where roles are subjective and exist in the minds of people, jobs are objective and can be documented in formal job analyses and job descriptions. Jobs are also bureaucratic, where roles are personal. This means that jobs are independent of any given incumbent; a "job" exists as a set of defined task elements even when there is no one performing those task elements. Further, although the role may change as incumbents change, the job remains relatively static. Finally, where roles are developed on the basis of social information from many sources, jobs are created by an individual or entity in a position of authority who will benefit by the performance of the job.

The "space" between the defined and static job and the demands placed by a dynamic environment is filled by roles. Because this space is defined by a set of changing circumstances, roles themselves are fluid and subject to a great deal of change. Thus, emergent task elements are added to jobs, and the entire set is referred to as a role.

Ilgen and Hollenbeck (1991) discuss several different job-role combinations, varying from the highly prescribed bureaucratic prototype, where the conceptual space is almost entirely comprised of fixed established task elements (the job) and very little space is devoted to emergent task elements (the role), to the "loose cannon" prototype, where nearly the entire job-role space is comprised of emergent (role) elements, surrounding a very minimal core of established (job) task elements. In addition, they discuss the fact that a fixed set of established task elements (a job) will have a different set of emergent (role) task elements associated with it, depending upon who fills the role. That is, the environment may make different demands on different job incumbents. Figure 3 displays Ilgen and Hollenbeck's examples of job-role combinations.

Developing roles. Understanding the content of roles is of little instructional value for the team researcher without companion knowledge regarding how roles develop and change. The classic model of role development (Kahn, Wolfe, Quinn, Snoek & Rosenthal, 1964; Katz & Kahn, 1978) describes a process by which an individual's superiors, peers and subordinates, collectively referred to as the role set, communicate expectations to the focal individual. The focal individual's behavior in response to these expectations influences the role set's expectations. The cycle of send – respond – alter expectations is then repeated. This non-recursive process is known as a role episode.

The alternative perspective, presented by Graen and his colleagues (Dansereau, Graen, & Haga, 1975; Graen, 1976; Graen & Scandura, 1987) argues for a more active role for the focal individual. Graen and his colleagues argued that the focal individual negotiates with the role set to develop a role that meets the needs of both the focal individual and the role set, rather than passively receiving role information from the role set.

Particular attention has been paid to the influence of supervisors on subordinates, and role theory is no exception. In their Leader-Member Exchange (LMX) theory, Graen

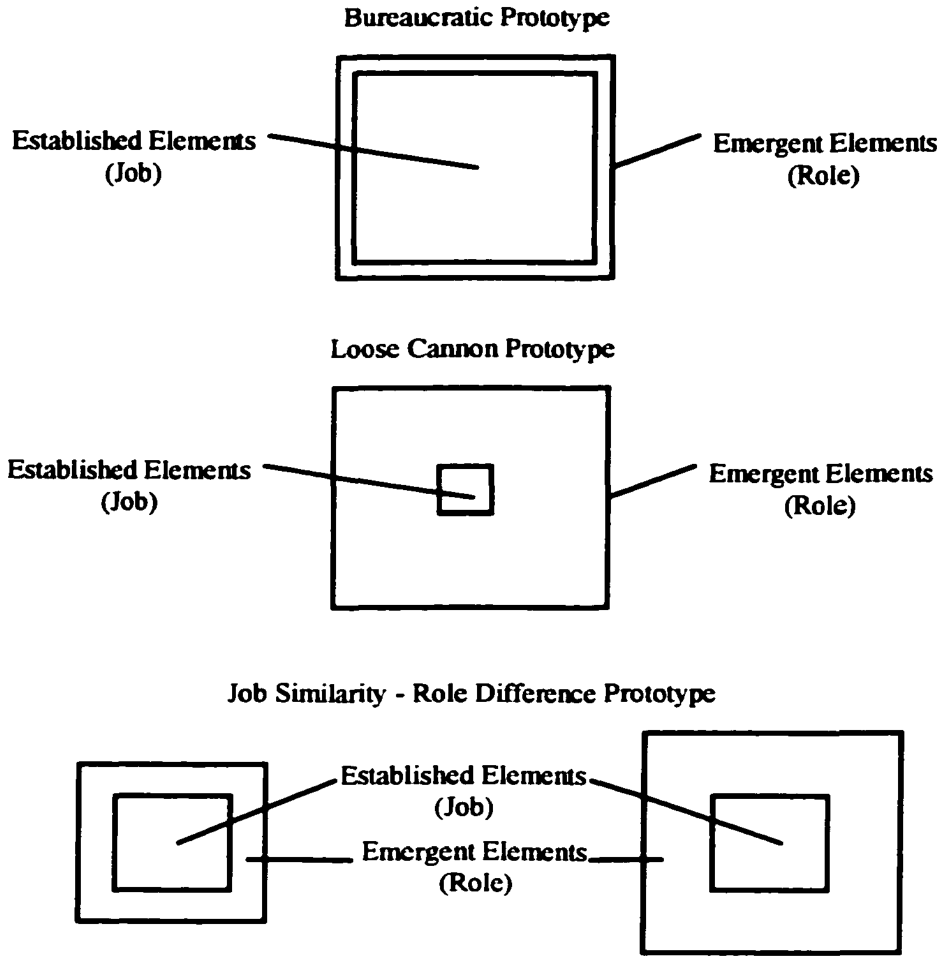


Figure 3. Examples of job-role combinations (Ilgen & Hollebeck, 1991)

and Scandura (1987) describe a process of role development from an incumbent's first encounter with her or his supervisor until role behavior stabilizes. This state of role routinization is achieved when the incumbent and supervisor have reached a shared understanding with regard to the incumbent's role that is agreeable to both parties.

In addition to the attention paid to the crucial role that supervisors play in the development of roles, Seers (1989) explicitly addresses the importance of team members in the development of roles. Although peers lack the formal authority and control over organizational resources that are vested in superiors, peers may exert a greater level of pressure on focal individuals through their more frequent presence and sheer numbers. That is, the constant influence of several less powerful peers may sum to a greater impact on role development than the distal, but more powerful, influence of the supervisor. Seers (1989) demonstrated the importance of team member exchange quality for job attitudes and performance. Thus, Seers argued that peers have a systematic and measurable effect on role outcomes.

All of these role development processes are bounded by a finite set of relevant task elements. That is, the universe of possible elements that could be incorporated into a role is limited to those that are germane to the job that forms the core of the role. In addition, roles are based on jobs, the task elements of which are non-negotiable (Ilgen & Hollenbeck, 1991). Thus, the role negotiation process, whether viewed as one of passive role taking or active role making, is always limited to that conceptual space between the demands of the environment and the static job upon which the role is built.

Role conflict and ambiguity. Unfortunately, the demands of multiple role senders in the role set and a continually changing environment contribute to difficulties in the role development process. Generally, the role theory literature has discussed these failures as one of two varieties: either a lack of clarity in the roles or incompatibility among the roles.

A lack of clarity in the expectations that comprise a role is referred to as role ambiguity. This ambiguity may arise from either a lack of clarity with regard to the behavioral expectations associated with a given role; alternatively, the ambiguity may result from confusion around the consequences of particular behaviors (Cook, Hepworth, Wall & Warr, 1981). Thus, either expectations or consequence may result in ambiguity.

Role conflict, on the other hand, arises from mutually exclusive or otherwise incompatible demands placed on the focal individual. These conflicting demands often originate from different expectations held by different members of a single role set (intra-role conflict), or the incompatible expectations may be held by member of differing role sets (inter-role conflict). Alternatively, it is conceivable that these conflicts emerge as a function of the dynamic nature of other elements in the environment. That is, as the space between a job and its environment changes, conflicting demands may emerge.

Although the concepts of role ambiguity and role conflict are interesting theoretically because they offer researchers insight into the processes by which roles develop and change, they also have practical implications. Jackson and Schuler's (1985) meta-analysis of role ambiguity and role conflict revealed significant negative relationships between the role constructs and many work related variables (See Table 4). For example, Jackson and Schuler (1985) found that role conflict and role ambiguity were significantly negatively related to every type of job satisfaction analyzed. In addition, both concepts were negatively related to feedback and leadership behavior (consideration and initiating structure) as well as to commitment and involvement. Conversely, role conflict and role ambiguity were significantly positively related to tension and anxiety and propensity to leave. Thus, the organizational impact of poorly developed roles can be substantial.

Rizzo, House and Lirtzman's (1970) measure of role conflict and ambiguity is by far the most frequently used (Van Sell, Brief & Schuler, 1981). Despite considerable scrutiny and questioning of the validity of the measure (e.g., Breaugh & Colihan, 1994; Schuler, Aldag & Brief, 1977; Tracy & Johnson, 1981), the Rizzo et al. (1970) ambiguity and conflict measure is still considered to be a satisfactory measure of role ambiguity and role conflict (Jackson & Schuler, 1985).

Table 4
Role Ambiguity and Role Conflict "True r" Estimates (Jackson & Schuler, 1985)

Correlate	Ambiguity "True r"	Conflict "True r"
Feedback from others	-.58	-.31
Feedback from task	-.41	-.25
Leader initiating structure	-.43	-.27
Leader consideration	-.44	-.42
Participation	-.55	-.37
Job satisfaction: general	-.46	-.48
Job satisfaction: supervision	-.53	-.53
Job satisfaction: work itself	-.52	-.49
Job satisfaction: co-workers	-.37	-.42
Job satisfaction: pay	-.26	-.31
Job satisfaction: advancement	-.40	-.38
Tension/anxiety	.47	.43
Commitment	-.41	-.36
Involvement	-.44	-.26
Propensity to leave	.29	.34

In response to the criticisms surrounding the Rizzo et al. (1970) measure, Ilgen and Hollenbeck (1991) suggested supplementing this classic measure with other metrics. Specifically, they suggested asking incumbents to indicate what elements comprise their role. By permitting incumbents to respond in a likert-type fashion (i.e., a rating scale anchored at one end by a statement like "certain that it is" and at the other by "certain that it is not"), the researcher may be able to make several interesting calculations regarding the focal individual's role.

Ilgen and Hollenbeck (1991) suggested that role ambiguity could be assessed through such an instrument by assessing the number of elements to which the focal individual responds that he or she is uncertain as to whether it is included in his or her role. That is, responses near the midpoint on such an instrument are an indication of role ambiguity.

Further, Ilgen and Hollenbeck (1991) suggested that role conflict could be assessed by comparing the responses of the role set to the focal individual's responses. To the extent that these descriptions of the focal individual's role differ, role conflict exists. In the case where multiple role senders are available, another dimension of role conflict could be assessed by comparing role descriptions between or among the role senders as well.

Role Theory and Team Situation Awareness

As discussed in the previous section, the contexts in which high SA is typically required presents several challenges to teams, beyond those faced in less demanding contexts. In addition, following the model presented by McGrath et al. (1995), the role network within a team is one of the most central elements in a team. Klimoski and Jones (1995) suggested that high levels of coordination and interdependence indicate that each team member knows what is expected of him or her and what he or she can expect from each of his or her teammates. Role theory, then, is applicable to teams that must operate in these contexts at a very basic level (Fink & Major, 1999; Major, Fink, & Stout, 1998/1999).

Role theory explicates the means by which the roles in a team are developed. Thus, role theory provides not an end, but rather a means to an end. As discussed earlier, role theory also indicates that roles will be fluid, and responsive to changing circumstances. Clearly, the circumstances in an SA context are highly subject to change. Thus, effective expectations for appropriate roles under different conditions, as well as protocols for efficiently shifting from one set of roles to another, are important in this context. Role theory provides the outline for how such processes may be utilized.

The perspective brought by role theory may have even greater importance for novice operators. Where expert operators are likely to already have well-established mental models with regard to team interactions as well as technical aspects of the task at hand, novice operators are likely to present more of a *tabula rasa*. Therefore, a structured process by which to gain shared understanding can spare such a team a great deal of trial-and-error learning. Additionally, those that may be technical novices no doubt still have some level of expertise or knowledge about social processes and interactions, making the content brought to light by role theory somewhat more within a "masterable" domain.

Essential Teamwork Behaviors

Ultimately, teamwork is a behavioral construct; it is a composite of behaviors that team members display to perform their own tasks and support their teammates in their tasks (i.e., work together) to achieve a common goal. However, much of the discussion thus far with regard to teams has not focused on behavioral dimensions of teamwork. While the two previously explored models of team process do much to illuminate segments of underlying team processes, there are still several elements of teams, teamwork and team processes that remain mired in confusion. Thus, it is important to review those teamwork behaviors that have been demonstrated to be essential to effective teamwork.

McIntyre and Salas (1995) helped to elucidate some of these essential teamwork behaviors. They examined command-and-control teams and outlined a series of behaviors and conditions that contribute to effective teamwork. Specifically, they identified four behaviors that are important for teamwork: monitoring, feedback, communication and back-up behaviors.

First, McIntyre and Salas (1995) suggested that monitoring one another's performance is central to teamwork. Their research suggested that, in effective teams, team members keep track of their teammate's work as well as performing their own. However, it is important that this monitoring is built on trust, with a goal of improving overall team performance, rather than an orientation toward catching another individual doing something wrong. McIntyre and Salas's research suggested that monitoring is actually built into the psychological contract among team members.

In addition, McIntyre and Salas (1995) contended that the giving and receiving of feedback is an essential teamwork behavior. They suggested that feedback is a natural extension of the monitoring discussed above. However, feedback is only effective to the extent that it can be given and received freely. Unfortunately, a free flow of feedback is sometimes hampered by such elements as status, rank and tenure or experience. McIntyre and Salas's research suggested that high performing teams have a climate such that these characteristics do not obstruct the flow of feedback among team members. Furthermore, they suggested that it is particularly important for team leaders to model an ability to accept constructive criticism. In this way, leaders can establish that such constructive criticism is appropriate within the team.

Communication is highlighted by McIntyre and Salas (1995) as central fixture in good teamwork. In particular, they cited closed-loop communication as important in situations where information must be communicated rapidly. They discussed closed-loop communication as a 3-step process. In the first step, the sender initiates the communication. Then, the receiver accepts the communication and provides feedback as a mechanism to indicate that the communication arrived successfully. Finally, the sender confirms that the correct or intended communication was in fact received by the receiver. McIntyre and Salas (1995) suggested that the importance of closed-loop communication is attested to by the fact that many military communications are based on proceduralized closed-loop communications.

Finally, they suggested that effective teamwork cannot occur without the "willingness, preparedness and proclivity to back up fellow members during operations" (McIntyre & Salas, 1995, p. 26). They suggested that this element "is perhaps at the very heart of teamwork, for it makes the team truly operate as more than the sum of its parts" (McIntyre & Salas, 1995, p. 26). Teams that exhibit back up behavior are populated by members that ably lend and receive help when needed. This is only possible, however, when team members are skilled not only in their own technical niche but are also competent at the tasks of their fellow team members.

McIntyre and Salas (1995) also outlined two conditions that enable the teamwork behaviors discussed above. First, they regarded it as essential that team members view themselves collectively, and that they consider their success to be contingent upon their interactions. Second, building upon the first condition, McIntyre and Salas suggested that within-team interdependence enables teamwork. To a certain extent this latter proposition goes without saying; interdependence is one of the central elements in the definition of teamwork.

Models of Team Situation Awareness

Despite the social bent to the inaugural paper on the topic of SA (Bolman, 1979), the majority of the SA literature is not written from a social perspective, but rather a cognitive or perceptual one. This perspective has influenced the manner in which models of SA at the team level have evolved. Consequently, sophisticated social processes do not appear in most models of team SA. Each of the models discussed below suggests different things about some aspects of SA measurement. However, other aspects of measurement will remain the same across models. For example, team SA is conceptualized as a second order variable in each of the models.

In their review of team SA models, Major, Fink and Stout (1998/1999) discussed three general models of team SA: team member contributions (summation), communication, and shared understanding. This paper will build on the foundation laid by Major and her colleagues in outlining different approaches to team SA in the literature.

Summation

The first model outlined by Major and her colleagues (1998/1999) addresses team member contributions to SA. Here, we are calling this the summation approach to team SA. This is the simplest approach to team SA. Here, team SA is considered to be the sum total of SA at the individual level. Endsley (1995a) exemplifies this approach. She states that team SA is "the degree to which every team member possesses the SA required for his or her responsibilities" (Endsley, 1995a, p. 39). Endsley explicitly stated that a lack in one team member's SA is detrimental to overall team SA. As seen in Figure 4, the summation model makes a direct leap from individual SA to team SA. This model translates into the following bivariate hypothesis:

Hypothesis 1: Individual SA is positively correlated with performance

This summation model is intuitively appealing both for its parsimony and its rationality. Here, the team level of the variable is simply the sum of the individual parts. This perspective is often applied when studying group phenomena. For example, if one is interested in a group's collective opinion on an issue, one might collect data on each group member's opinions and then take the sum or the mean of all the individual level data to get the group level data.

SA however, is a complex phenomenon, and when examined at the team level, we may expect it to be even more complex. While individual SA may be a limiting factor, much as is Level 1 SA (necessary but not sufficient), it is unlikely that it accounts for the entirety of SA at the team level. Indeed, if each individual has a very high level of SA for her or his area of expertise or responsibility, but is oblivious to all other information, it may be difficult for a coherent picture of the entire situation to be developed at the team level.

In presenting her summation model of team SA, Endsley (1995a) discussed the issue of redundancy, noting that some pieces of information may be necessary components of more than one team member's SA. However, she did not address the issue of judicious use of resources in terms of establishing an appropriate level of redundancy for team functions. She failed to outline processes by which the redundant information is gained. That is, she failed to consider that one team member might have primary responsibility for

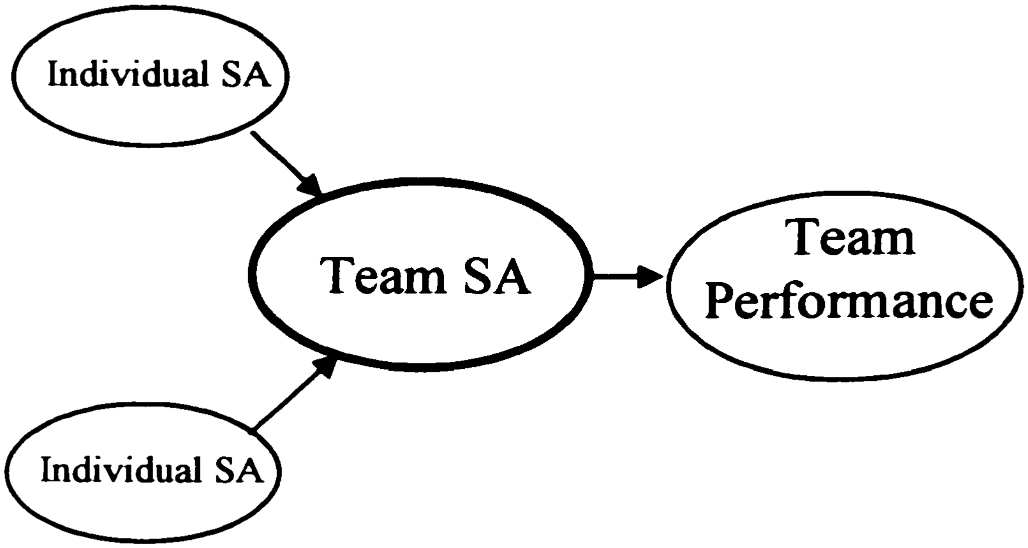


Figure 4: Summation model

monitoring a certain element of raw data in the environment, and then passing along processed information (Level 2 SA information) with regard to that element to an individual who must integrate that information with other environmental information (Level 2 and Level 3 SA responsibility). Thus, we are left with several team members independently developing their own SA and acting accordingly, perhaps with little knowledge of their teammates' actions. Furthermore, we are presented with a situation where the only guarantee against counterproductive actions is redundancy, an approach that leaves much to be desired, particularly in the resource-thin situation in which many teams find themselves.

Communication

Major and her colleagues (1998/1999) also discuss the communication perspective on team SA. Here, communication is taken as either a cause or an effect (or both) of good team SA. Applying Endsley's three levels of SA to the team level, addressing issues of communication is much like the integration and synthesis level of individual SA, Level 2.

However, the communication model of SA is conceptually imprecise. Communication is implicitly treated as both cause and effect, rendering any model difficult to test in a linear, empirical fashion. To the extent that communication is both sent and received, this is a natural result; communication may mean one thing for the sender and another to the receiver. Given that communication is not a strictly linear process, this cause/effect debate may be, to a certain degree, unavoidable.

Some authors address communication as an important concept in the study of SA, but avoid the cause-effect problem entirely. Shrestha et al. (1995) avoided the cause-effect debate by proposing that communication acts as a moderator for SA. Given the central role that communication plays in developing team SA (Shrestha et al., 1995), however, it seems unwise to relegate it to moderator status.

It is logically possible for communication to be both a cause and an effect of good team SA. Communication can function as a precursor to team level SA: the individual data points are communicated from individuals to the team at large, thus enhancing team level SA. However, it is equally logically attractive for communication to be a result of good SA; teams which have a coherent and unified understanding of a given situation should communicate more efficiently and effectively than those that are confused and uncoordinated.

The following solution to this tautology is proposed: communication is a cause of SA for the receiver of a communication and an indicator of SA for the sender of a communication. Thus, communication affects each individual's SA, rather than directly affecting team SA itself. Rather, the influence of communication on team SA is indirect, through its influence on individual SA. This non-recursive relationship can be seen in Figure 5, and translated into the following bivariate hypotheses:

Hypothesis 2: Individual SA is positively correlated with communication

Hypothesis 3: Communication is positively correlated with performance

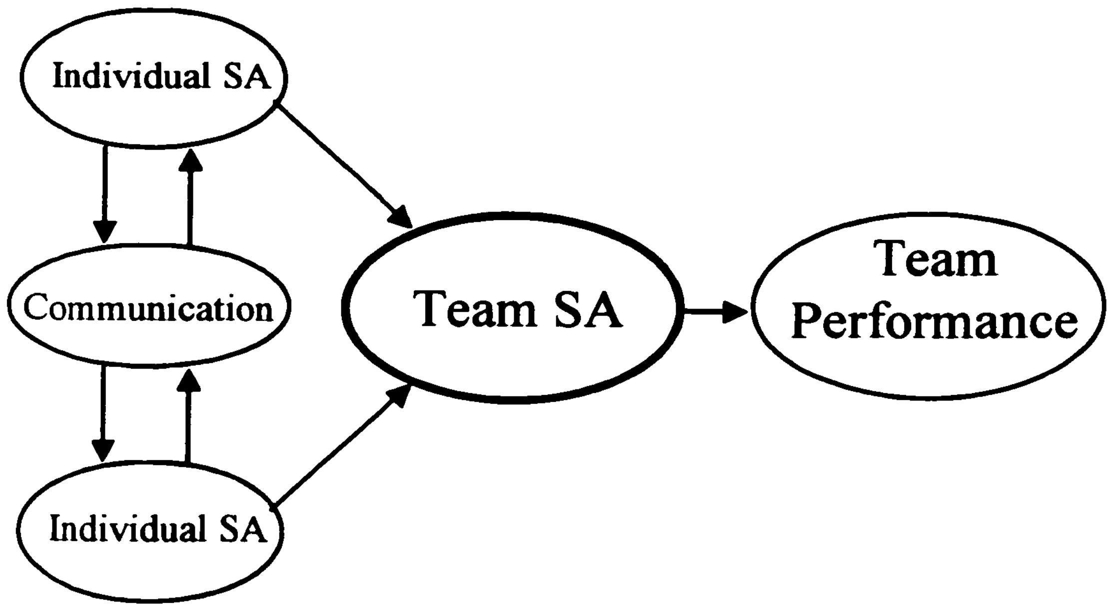


Figure 5. Non-recursive communication model

While such processes pose difficulties for those trained in cause-and-effect, linear models of science, cybernetics relies upon such feedback loops (see Richardson, 1990). Therefore, they are not without precedent in the social sciences.

Some authors have suggested that specific types of communication are indicative of team SA. Schreiber, Lee, Raspotnik, and Hubbard (1996) found that high SA pilots showed more overall communication, were more directive in their communication, made more requests, and were more specific in directing their communications at a particular individual. Several of the same researchers (Schreiber, Bell & Raspotnik, 1998) identified a similar pattern, finding that pilots high in SA provided more information, were more directive and requested more information.

Hollenbeck and et al. (1995) performed post hoc analyses on the communication patterns of high-performing v. low-performing teams. Contrary to Schreiber and his colleagues' (1996) findings that pilots high in SA showed more overall communication, Hollenbeck and his colleagues (1995) found that the amount of communication did not differ significantly between the high and low performing teams; rather, the nature of the communication differed. Specifically, low performing teams sent significantly more requests for information (low $M = 5.3$ v. high $M = 0.6$) and high performing teams sent significantly more facts about the situation (high $M = 15.9$ v. low $M = 9.6$). Furthermore, high performing teams were more likely to send communications that had already been processed and evaluated; that is, they were more likely to share Level 2 SA information.

Hollenbeck et al.'s (1995) finding that low performing teams made more requests for information appears to be in contrast to Schreiber et al.'s (1996) finding that pilots high in SA made more requests, and Artman's (1999) finding that greater performance was actually related to fewer communications. However, Schreiber et al. (1996) did not specify what type of requests their high performing pilots made. That is, it is possible that the pilots were making requests for action rather than requests for information, thus eliminating the contradiction.

Hollenbeck et al. (1995) went so far as to suggest that this difference in communication patterns may be the primary factor that distinguishes high performing from low performing teams. Similar to the overall cause-effect debate about communication, sending information rather than requests represents both a cause of higher performance (time is spent efficiently) and an effect of good team processes or high SA (team members are able to anticipate the needs of their team members).

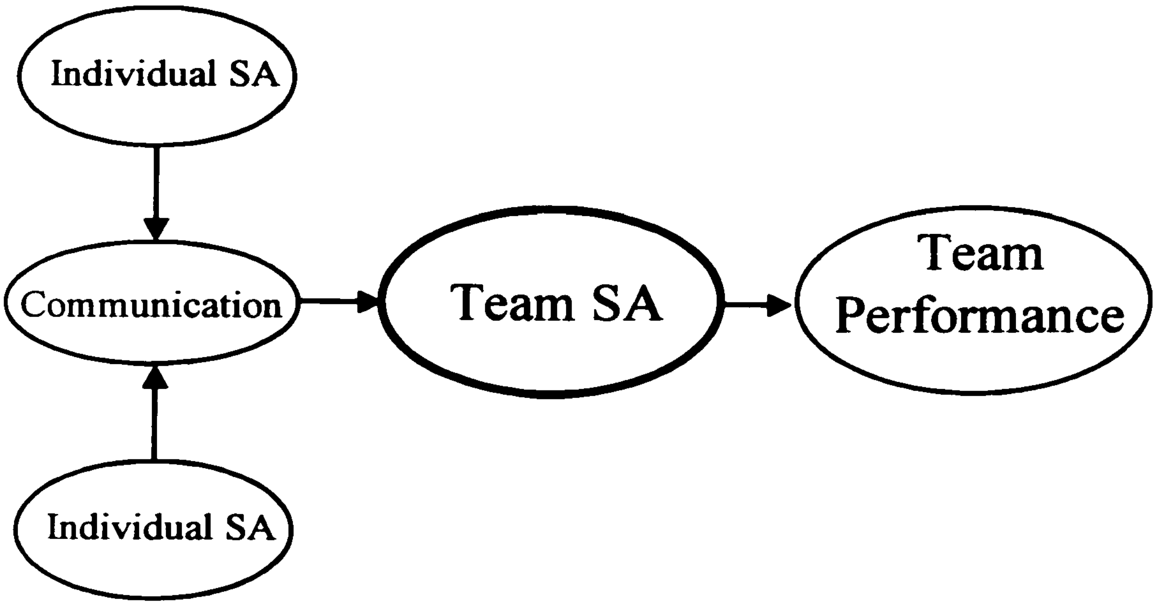


Figure 6. Communication model

The communication perspective suggests something different about the development of team level SA than the summation perspective does. Where team level SA is a direct result of individual level SA in the summation perspective, communication is the only route by which team SA can be achieved under the communication perspective (see Figure 6).

The central difficulty with the communication perspective on team SA is its reliance on overt communication. Major and her colleagues (1998/1999) pointed out that high team SA may actually decrease a team's need to communicate. In fact, it seems reasonable that one of the key distinguishing characteristics of teams with high levels of SA may be the seamless fashion in which they "change gears;" where a team with low SA may suffer through a period of confusion when conditions change, teams with high SA may have a much smoother transition. Wellens (1993) suggested that the quantity of communication within the team decreases due to the high demand that is characteristic of the SA context. Thus, Major et al. (1998/1999) concluded that communication is insufficient to explain or describe the phenomenon of team SA, although they acknowledged that communication is essential to team SA.

Shared Understanding

The shared understanding approach is fundamentally a mental models approach to SA at the team level. Shared understanding is built upon "mental maps" that are shared among team members. These mental maps may develop either through explicit attention to their formation or through more implict means, such as experience with a given team and/or situation. Recent research (Bolstad & Endsley, 1999; Farley, Handsman, Endsley & Amonlirdviman, 1999) has shown that shared information, specifically shared displays, has a beneficial impact on SA and on performance, lending some support to the theory that shared understanding should increase team SA. Major et al. (1998/1999) suggested that the processes by which these cognitive constructs are formed and shared can also be social and behavioral.

In proposing the shared understanding approach to team SA, Stout, Cannon-Bowers and Salas (1996/1997) suggested that there is more than one route to high levels of team SA. Essentially, there are only two major elements in this model: individual SA and shared understanding. Stout et al. suggested that high levels of team SA can be achieved either through high levels of individual SA or through high levels of shared understanding (see Figure 7). This suggests the following bivariate hypotheses:

Hypothesis 4: Individual SA is positively correlated with shared understanding

Hypothesis 5: Shared understanding is positively correlated with performance

According to Stout et al., the same level of team SA may be achieved by either of these routes to team SA, and different situations will determine the extent to which each route is central to the formation and maintenance of team level SA. This is in direct contradiction to the summation perspective, where high individual levels of SA were proposed as the sole means to high levels of team SA.

The mental models, or shared understanding, approach to team SA does a fine job of identifying some of the specific types of information that must be shared in order for a team to operate effectively in the SA context. Thus, this cognitive perspective provides a great deal of guidance on the content of shared understanding. Role theory, however, focuses more on the processes by which mental models come to be shared among team members (Major et al., 1998/1999). While the social perspective brought by role theory adds another wrinkle to the content piece of shared understanding, a large part of its contribution is the process complement to the content component provided by a focus on mental models. Thus, role theory is an integral part of the shared understanding approach to team SA. In addition to illuminating the process by which shared understanding comes to be shared, role theory also provides content for shared understanding, in that roles are a major component of shared understanding.

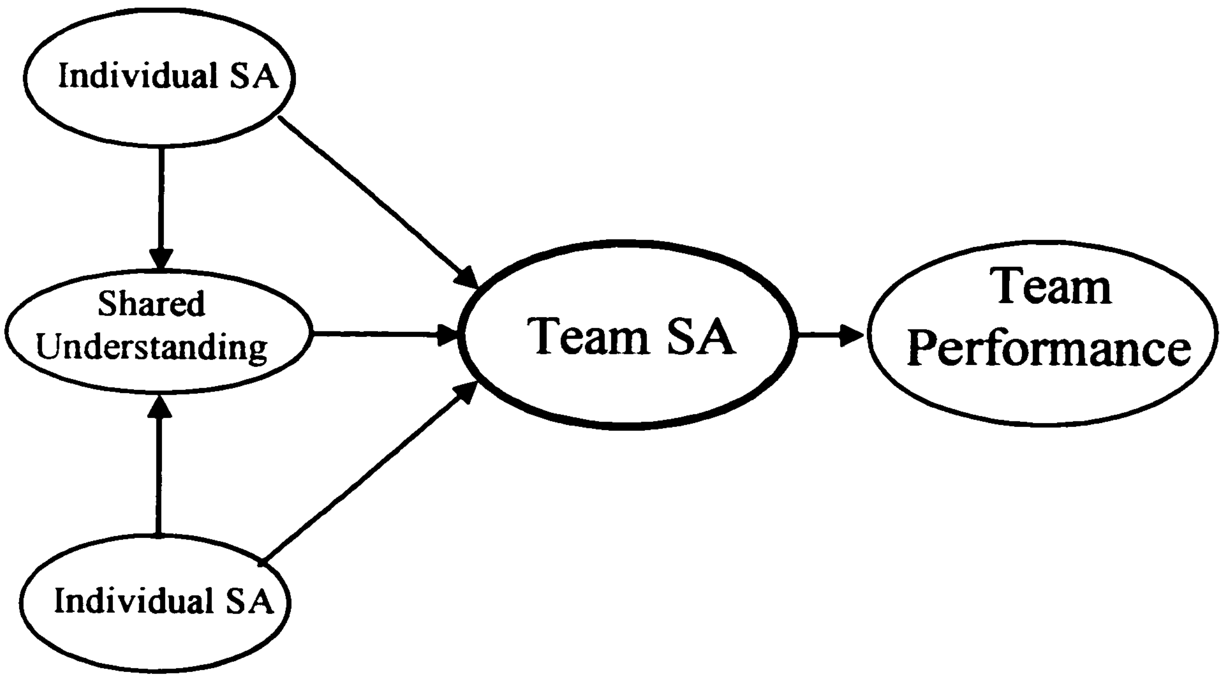


Figure 7. Shared understanding model

Integrated Model

It is often the case that several theories have something of substance to offer the scholar of a particular topic. The models presented above are not necessarily mutually exclusive, although certain elements among them are contradictory. However, to account for a maximum amount of the variance in team SA, it seems advantageous to draw upon each of the previously outlined models of team SA to develop one integrated model of SA at the team level. Thus, we find ourselves with a model that contains elements from the summation, communication and shared understanding models (see Figure 8), which suggests this final bivariate hypothesis:

Hypothesis 6: Communication is positively correlated with shared understanding.

The summation model (Figure 4) suggested that a direct link must exist between individual SA and SA at the team level. We reviewed two different approaches to incorporating communication into a model of SA at the team level. One considered communication to be a non-recursive process, where communication acted as both a cause and an effect of SA (Figure 5). The second suggested that communication exerts a direct influence on levels of team SA (Figure 6). Elements of both of these approaches to incorporating communication into a model of overall team SA are included in the integrated model. Communication is shown both resulting from individual SA and contributing towards individual SA, and it is shown influencing team SA, albeit indirectly.

This non-recursive representation of communication adds to the shared understanding model, in that shared understanding is explicitly influenced both by individual SA as well as communication. Stout et al.'s (1996/1997) contention that team level SA is a function of either high levels of individual SA or high levels of shared understanding is represented in this model as well; team level SA is directly influenced by both individual SA and shared understanding in the integrated model. Thus, we see in the integrated model the central elements of all three of the models of team SA previously reviewed.

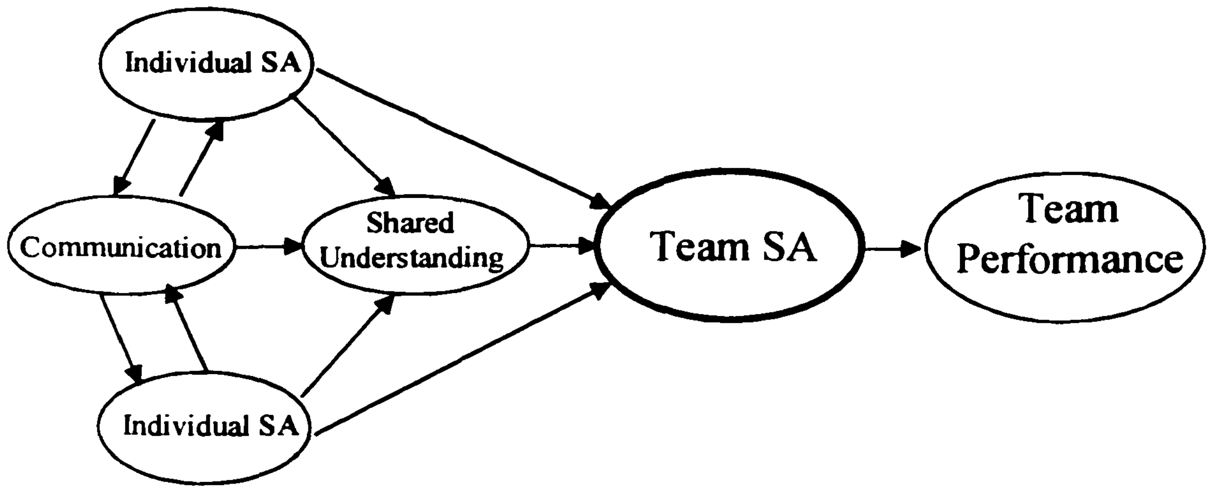


Figure 8. Integrated model

Statement of the Problem

This research had a twofold goal. The first goal was to evaluate the relative efficacy of several SA measurement methods established in the literature. Second, the three models of team SA proposed in the literature were tested, as was as the integrated model proposed above. In addition, this discourse was intended to add some clarity to the SA literature as a whole and the team SA literature in particular.

Evaluating SA Measures

As outlined in the section on SA measurement, a wide variety of measures of SA have been proposed. The present paper utilized several of these measures. Further, we conducted analyses to determine which among them provided the greatest predictive efficacy. Despite the variety in SA measures and approaches to measurement, no rigorous study comparing the various methods has been conducted previously. Without data comparing the techniques, it is difficult to generalize between and among studies to speak authoritatively on the matter of SA.

Furthermore, this research was intended to make a contribution to the SA literature by analyzing SA measurement techniques according to their type. That is, we analyzed proximal, meso and distal measures of SA. Given the nature of these different approaches to SA measurement, we treated proximal and meso measurements as predictors of SA and distal measures as outcomes of SA. Thus, we were able to establish predictive models.

Evaluating Models of Team Situation Awareness

This paper outlined several models of team SA that have been proposed in the literature. To date, however, none of these have been subjected to rigorous empirical investigation. By developing an integrated model, as well as testing the models which are currently present in the literature on team SA, we had two routes by which to establish the relative validity of the various models. Of course, testing each model independently yielded fit statistics that indicated the correctness of each model. In addition, the integrated model offered an opportunity to establish which elements account for the greatest amount of variance in the variable of interest, team level SA. It was expected that the integrated model, drawing on several theoretically based models of SA at the team level, would yield the best fit to the data.

METHOD

Participants

Participants ($N = 272$) were drawn from a large southeastern university. Mean age of participants was 21.4 ($SD = 6.2$), although participants ranged in age from 18 – 74, and 42 (15%) did not report their age. One hundred eighty (66.2%) participants were female, 80 (29.4%) were male and 12 (4.4%) did not indicate their sex. Participant self-identification of racioethnic identity was 48.9% (133) Caucasian, 27.9% (76) African-American, 9.9% (27) Asian, 5.1% (14) Hispanic, and 5.1% (14) Native American/Pacific Islander. Eight participants (2.9%) declined to report their racioethnic background.

Participants were volunteers from the student body at a large, urban, southeastern university. The university review process approved the use of human subjects for this research. Participants received their choice of experimental credits in eligible courses, $10 or an entry into a lottery for $100 as an incentive for their participation. In addition, because this study used a commercially available flight simulator game, it was observed that the participants derived intrinsic enjoyment from participation.

Participants had the option to sign up to participate individually or in pairs. As a means to control variance with no relationship to the theoretical model, same sex pairs were used. Participants who signed up to participate as individuals were randomly paired with another same-sex individual to complete the experimental trials. Participants who signed up in pairs were permitted to complete the trials as intact pairs. One of the models of team SA, the shared understanding model, predicts that one means by which shared understanding may be achieved is through experience with team members. Thus, it was necessary that there be some variability in team members' familiarity with one another to adequately test this proposition.

Measures and Materials

Materials

The simulation used in this study was a commercially produced helicopter flight simulation game, *Werewolf v. Comanche* (NovaLogic, 1995). Jentsch, Barnett, Bowers, Hicks, and Sierra (1997) evaluated commercially available flight simulation software for appropriateness in team SA research and determined that *Werewolf v. Comanche* was the best package for this purpose. Due to the fact that SA is generally of greatest interest in contexts that are risky and time-constrained, most SA research is conducted using simulators, rather than in the actual task environment. Gugerty and Tirre (1996) compared SA in a PC-based simulator and found that it correlated significantly with performance in a higher-fidelity simulator and actual task performance. Thus, PC-based simulators have been shown not to distort SA.

Jentsch et al. (1997) evaluated 55 commercially available aviation simulations in a multi-phase procedure. First, they identified those simulations that met 8 criteria: 1) elicitation of team SA behaviors/processes, 2) allow experimental manipulation, 3) allow data collection, 4) allow control of extraneous variables, 5) simulate military aviation context, 6) easy to learn and operate, 7) minimum of hardware and peripherals, and 8) inexpensive. Simulations that met these criteria were evaluated by subject matter experts (experienced aviators) for ease of use my aviation-naïve participants. *Werewolf v. Comanche* was identified as the best platform for team SA research in the final phase, which evaluated simulations for their controllability, scenario complexity, realism and multiplayer capability.

Jentsch et al. (1997) determined that *Werewolf v. Comanche* was the easiest simulation to control, stating that "this simplicity and intuitiveness of use, together with the multiplayer capabilities and the graphics which appealed to participants were the final reasons that *Werewolf v. Comanche* was ranked highest" (p. 24). Jentsch et al. (1997) reported that, after a brief factual introduction to the simulation, "most participants were able to successfully complete most tasks in their first attempt" (p. 31). All participants in Jentsch et al. (1997) were able to perform critical tasks such as takeoff, hover, forward flight, turns, selecting weapons, and firing gun and destroying target on their first attempt, without any practice. Those tasks for which one or more participants required a second attempt or a hint were of a more sophisticated nature, such as "selecting one specific gas tank in a refinery" (p. 32). Similar ease of use was observed for participants in this research.

When networked together, players can "see" one another in the simulation, and are two players in a single simulation. Therefore, if one player destroys a target, the target is destroyed for both players. Networked players may either play "with" each other or "against" each other. Jentsch et al. (1997) revealed that the simulation performed reliably in the multiplayer mode.

The simulation offered several missions for each of several levels of difficulty. The summary of each mission is displayed on the screen after it is selected. The mission summary must be acknowledged by the participants before the simulation begins; that is, the players must click a button indicating that they accept the mission before the simulation will begin. The simulation includes both stationary and moving targets in the form of mission goals (stationary) and enemies (moving). Additionally, the simulation tracks each team member's performance in terms of mission goals destroyed and enemies killed.

Measures

Several criteria were applied in the selection of measures for this study. First, only measures or measurement approaches that had been previously described in the literature were used. Second, only measures that had either some established theoretical basis or

Table 5
Measures Included in the Research

Latent Trait	Measure	Rater
Individual SA		
	Proximal SA Questionnaire	Self
	SART	Self, Partner, Experimenter
	SALIENT	Self, Partner, Experimenter
Communication		
	Communication estimates	Experimenter
	Communication behavioral observation	Experimenter
Shared Understanding		
	Role checklist	Self, Partner
	Role conflict and ambiguity scale	Self
	Task/partner experience	Self
Performance		
	Objective performance	(simulation recording)
	Subjective performance	Self, Partner

previously established empirical validity were used. Third, only measures that were applicable to, or could be adapted for, the simulation task in this research were used. Fourth, only measures that fit within the models discussed were collected. Measures were selected to ensure a balance between objective and subjective measures, and to ensure representation of proximal, meso and distal indicators.

The set of measures that satisfied these criteria is shown in Table 5. Furthermore, as the majority of these measures had not been subjected to rigorous previous validation efforts, preliminary analyses were conducted on each of the measures to establish the relative merits of each. Only those measures that displayed evidence of good measurement characteristics and preliminary evidence of validity were included in model testing.
It is important to note that no direct measures for team SA were included in the proposed research. Rather, team SA was a second order variable, built from other latent variables in the various models.

Individual SA Measures

Both proximal and meso level SA were included in the individual SA measures. Proximal SA was assessed using brief questionnaires in a probe or freeze technique. Endsley (1995b) suggested that this was the most promising technique by which to assess SA. Furthermore, Endsley (1995b) has demonstrated that briefly pausing a simulation, as this technique requires, does not adversely affect SA. Hollenbeck et al.'s (1995) construct of informity has been shown to significantly contribute to team decision performance. Thus, questions for this measure were derived from task-relevant information present in the simulation. The experimenter collected data on each of these items for each participant as well. Items are presented in Table 6.

One of the criticisms of probe measures of SA is that simply by asking a particular question, we have changed the operator's SA. In effect, this criticism addresses a cueing effect; that by asking a question about something, we convey to the operator that is it important, and thereby cue the operator to attend to that piece of information in the future. As a test of this, one item was repeated across breaks in the proximal SA measure.

Table 6
Proximal SA Questionnaire

Break	Items
Break One (at 5 minutes)	
	How many enemies are currently threatening?
	What is your current weapons setting?
	How many mission goals (targets) are currently within range?
Break Two (at 15 minutes)	
	What is your current speed?
	What is your current heading (direction)?
	How many enemies are currently threatening?

Proximal SA items were calculated as a difference score between the participant's recall of the item during the break and the experimenter's recording of the true state of affairs. This difference scoring approach is similar to that taken by Chaparro et al. (1999). Weapons setting was coded as true/false; either the participants were aware of their weapons setting or they were not. The score for mission goals and for enemies within range was the difference between the participants' reports and the experimenter's record of actual targets in range. The score for current speed was calculated by taking the absolute value of the difference between the participants' report of their speed and their actual speed at the time of the break, and dividing this number by 5. Finally, current heading (direction) was based on an 8 point heading system, using 8 basic directional points as anchors: North, Northeast, East, Southeast, South, Southwest, West and Northwest. In this scheme, East is 2 points away from North, as is West, with a maximum difference of 4 (i.e, South is 4 directional points away from North – NE, E, SE, S).

Because of the difference scoring, low scores are indicative of accuracy and are therefore desirable.

Meso level SA measurements were also used to indicate individual level SA. These measurements were calculated using items from two of the three scales discussed in the introduction: Taylor's (1989) Situation Awareness Rating Technique (SART, Table 1) and Muniz et al.'s (1997) Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT, Table 2). Both of these measures are grounded in SA theory and were developed with the assistance of subject matter experts. Additionally, SART has been subjected to extensive factor analyses, although no reliability data are available. These measures were completed by the experimenter and participants as a rating of self SA and as a rating of partner SA. The exact items used can be found in Appendices B and C.

Taylor (1989) specified that the SART was scored on a 7-point rating scale, where 1 indicated "low" and 7 indicated "high." Therefore, this scoring scheme was preserved for the purposes of this study. Participants rated themselves and their partners. Additionally, the experimenter rated each participant using this scale.

SALIENT is a behavioral approach to SA measurement. The original SALIENT methodology was devised for a very circumscribed simulation for use with skilled pilots. During the simulation, the crew experiences a series of events. For each of these events, they have the opportunity to display their SA behaviorally in a 5 x 25 matrix of categories. In theory, then, there are 125 opportunities to demonstrate SA at each juncture. In the original proposition of the SALIENT methodology, subject matter experts and team task analysis were used to identify a "correct" behavior in each of the categories for each segment of the simulation. During the simulation, an experimenter notes the number of times the participants' actions correspond to the behavior identified *a priori* as "correct."

Two factors make SALIENT, as originally designed, inappropriate for our use. First, the novice status of the participants means that, for example, it may be unreasonable to expect participants to demonstrate the "correct" behavior during the simulation. Second, the simulation used in this research allows for variability in the simulation. That is, while each scenario progresses according to a predetermined pattern, participants were not given a precise flight plan, as the participants envisioned by Muniz et al. (1997) would be given. Thus, it was impossible to predict, for example, from which direction enemies would appear in this study, as participants may have chosen a circuitous route. This presented a challenge to the researcher in developing a clear sequence of events for the purposes of creating an observation form of the type described by Muniz et al. (1997).

Therefore, the SALIENT instrument was adapted for the purposes of the study. The content was preserved, but rather than used in an observation format, the dimensions were used as stems for a rating scale. Participants rated themselves and their partners on each of the items, using a 1 to 9, likert-type rating scale, where 9 indicates "consistently did this" and 1 indicates "never did this." Thus, the adaptation of SALIENT used in this research essentially captures the extent to which a particular behavior was characteristic of a participant, rather than whether precise actions or reactions indicative of one of the SALIENT dimensions were taken at a particular point in the simulation. The experimenter rated participants using this scale as well.

For model testing and statistical analyses, the meso level constructs theoretically load onto the latent variable "individual SA." For both measures (SART and SALIENT), the participant's self ratings, her or his partner's ratings of the participant, and the experimenter's ratings of the participant were used as observed variables, in a 2 (test) x 3 (rater) matrix: "self rating – SART," "partner rating – SALIENT" etc. Thus, three of these observed variables (self, partner, experimenter) are variations of SART and another three are variations of SALIENT, and the six observed variables in turn indicate the latent variable "individual SA." However, prior to being used in model testing, these measures were subjected to preliminary analyses, evaluating their measurement characteristics, such as factor structure and reliability, as well as preliminary indications of their validity.

Communication Measures

Communication was operationalized both objectively and subjectively, through experimenter ratings (subjective) and a communication sample collected by the experimenter. Communication was the only measure in the proposed research that was collected solely by the experimenter. The subjective communication estimates focused on issues relevant to building shared understanding as well as assessing overall communication levels. Although research on SA using communication as a variable has addressed communication in terms of frequency (e.g., Schreiber et al., 1996), the special characteristics of SA, team SA, and propositions of role theory suggest specific data collection categories. Therefore, in addition to rating the overall frequency of communication for each participant, a rating scale was developed that included ratings for five additional dimensions of communication.

The first additional communication dimension was on-task communication of a mundane nature, such as current conditions (speed, altitude, weapons). These items indicated the same underlying construct as the proximal SA measure. The second additional dimension also addressed on-task communication, but it captured communication related to emergencies, such as attacks. Third, a dimension was added to reflect the participants' extent of role-related communication. Roles and role theory were hypothesized to provide content that contributes to shared understanding, and it seemed reasonable to assess communication that reflected continued role negotiation or references to established roles. Similarly, shared understanding was also hypothesized to be influenced by the relationship between team members. Thus, the fourth additional dimension addressed communication that was team related, such as rapport-building communication and support statements. Finally, McIntyre and Salas (1995) suggested that closed-loop communication was an essential teamwork behavior, particularly in time-constrained situations. Thus, the fifth additional communication dimension addressed the participants' use of closed-loop communication. Given that good role-related communication, for example, can also be closed-loop, it was possible that a single communication could be included in multiple categories.

All of the communication estimates were made by the experimenter, at the completion of the trial. The experimenter made communication estimates for each participant individually. The estimates were made on a percentage basis, using a base of 10. An end point of zero was included to indicate that the participant never exhibited a particular communication type. That is, the experimenter rated the percentage of time that each participant engaged in each type of communication. If a participant spent 60% of her or his time engaged in role related communication, then he or she would have received a communication estimation rating of 6 for that dimension.

Despite the fact that this communication measure was grounded in SA theory, it was unvalidated. Thus, these data were subjected to preliminary analyses to determine the measurement characteristics and efficacy of the measure. Factor analyses and reliability analyses were run on this measure.

In addition to the communication estimates, the experimenter collected a 10 minute communication observation sample between the two breaks. That is, from minute 5 of the simulation to minute 15 of the simulation, the experimenter tallied each communication and categorized it. Additionally, the trial segment between probe breaks was tape recorded as a back-up for communication observation.

Following Hollenbeck et al. (1995), the communication sample was based on the frequency of particular types of communication. Hollenbeck et al.'s post hoc observations suggest that particular types of communication that appear to contribute to team success. Thus, the communication sample collected communication data in three content categories: information requests, raw data about situation facts and processed (qualitative) data about situation facts. These data were collected as frequency data. That is, a score of 0 in one of the categories meant that, during the 10 minute sample period, the participant exhibited no communication of that type.

Communication items and data collection formats can be found in Appendices D and E. As shown in the communication models, the team level latent variable of communication also predicts and is predicted by individual level SA. Further, given that communication is the primary means by which shared understanding is achieved, communication also predicts the latent variable "shared understanding" in team level analyses of the integrated model. Finally, communication indicates the latent variable "team SA" in the communication models.

Shared Understanding Measures

Shared understanding is contributed to by roles and also, according to Stout et al. (1996/1997), experience with team members or the task at hand. As the integrated model discussed in the introduction displayed, communication was also hypothesized to contribute to shared understanding.

Roles were assessed in two ways. First, a role checklist (Table 7) was used. Ilgen and Hollenbeck (1991) recommended a procedure in which focal individuals indicate how sure they are that a particular task element is included in their role. Additionally, they recommend that another individual from the focal individual's role set do the same for the focal individual's role. Thus, a list of role elements was constructed for the purposes of the proposed research. Participants indicated how confident they were that each item was included in their role, and they also indicated how confident they were that the items were included in their partner's role. Ratings were made on a 9 point, likert-type scale, where 1 indicated "completely confident that this is not part of the role," and 9 indicated "completely confident that this is a part of the role."

Following Ilgen and Hollenbeck's (1991) recommendations, role ambiguity was operationalized as a neutral response to the question of how confident a participant was that a given element was part of a particular role. For ease of understanding, it is preferable that role conflict and ambiguity be negatively related to shared understanding, given that they are inversely related. Thus, low role conflict and ambiguity scores would correlate with high shared understanding scores, due to the fact that those with high levels of shared understanding should agree on their roles, and therefore have low levels of role ambiguity and role conflict.

Table 7
Role Checklist

Role or Responsibility
Leader
Back-up person
Responsible for locating and identifying mission goals or targets
Responsible for attacking mission goals or targets
Responsible for locating and identifying enemies
Responsible for disabling enemies
Responsible for setting strategy
Devil's advocate
Moral support
Navigate - decide where to go
Weapons expert - decide which weapons to use
Responsible for making task assignments
Responsible for monitoring partner's activities
Responsible for keeping perspective
Responsible for keeping task fun
Trainer – teach partner

To calculate role ambiguity, role checklist items were recoded into "confidence" scores for the purposes of calculating role ambiguity. That is, items were turned from a 1 to 9 scale that ranged from "Completely confident that this is *NOT* a part of role" to "Completely confident that this *IS* a part of role" into a 1 to 5 scale that ranged from great confidence (1) to no confidence (5). Items that were scored originally 1 to 5 maintained their original scoring, but 6 to 9 were recoded in descending order, such that the greatest confidence in either direction is coded as a 1. Thus, scores were transformed into a 1 to 5 scale, where 1 was very confident (positively or negatively), and 5 was no confidence as to whether the item is included in the role. In this way, a high score indicated high levels of role ambiguity, and a low score indicated low levels of role ambiguity (or high levels of role clarity).

Role conflict was calculated using difference scores between self and partner ratings, using the original ratings. Thus, if the self rating for a role element was 7 and the partner rating for the same element was 9, there was a difference score of 2. The smaller the difference between self and partner ratings on a given role element, the greater the degree of agreement between team members and the lower the level of role conflict. Thus, low scores in role conflict indicated low role conflict, and high difference scores indicated high levels of role conflict. It is important to bear in mind, however, that this methodology only captured role conflict in terms of disagreement between the focal individual and the role set; it did not capture intra-role conflict.

Second, roles were assessed by several items from Rizzo et al.'s (1970) role conflict and ambiguity scale. These items are presented as Table 8. Only those items which were applicable to the task used for the proposed research were included. As a result, the abbreviated scale includes 5 items assessing role ambiguity and 3 items that assess role conflict.

Despite the challenges to this measure, it was indisputably the most commonly used measure of role ambiguity and conflict. At a minimum, the implementation of both methods for role assessment allowed the opportunity to quantitatively compare them. The items on the Rizzo et al. (1970) measure were scored on a 1 to 5 Likert-type agreement scale, where 1 indicates strongly disagree and 5 indicates strongly agree. Thus, the Rizzo et al. measure was scored in the same direction as the role checklist; 1 indicated low role conflict or ambiguity, and 5 indicated high role conflict or ambiguity.

Table 8
Items from Role Conflict and Ambiguity Scale (Rizzo et al., 1970)

Dimension	Item
Role Ambiguity	
	I feel certain about how much authority I have.
	I have clear, planned goals and objectives for my job.
	I know what my responsibilities are.
	I know exactly what is expected of me.
	I have to work under vague directives or orders.
Role Conflict	
	I work on unnecessary things.
	I receive a assignment without the ability to complete it.
	I work under incompatible guidelines.

For the purposes of model testing and statistical analyses, the role measures theoretically indicate (albeit negatively) shared understanding. Preliminary analyses were used to determine the measurement characteristics of each measure. Those measures that were determined to have satisfactory measurement characteristics were included in model testing. When scored for role ambiguity, the role checklist was an observed variable "role ambiguity." When scored for role conflict, the role checklist was the observed variable "role conflict." Role conflict and role ambiguity indicated shared understanding, and were expected to have a negative relationship with shared understanding. The Rizzo et al. (1970) scale was also an observed variable that indicated shared understanding, and was also expected to be negatively related to shared understanding.

One of the models of SA suggests that shared understanding can be gained through either task experience or experience with team members. Therefore, task experience and experience with one's partner were also assessed. These items are shown in Table 9. Experience items were scaled from 1 to 5 in a Likert-type format, where the anchor for 1 was "not at all" and the anchor for 5 was "very/a lot."

For the purposes of model testing and statistical analyses, the first three items were treated as observed variables that indicated the latent variable "task experience." The observed variable "partner experience" and the latent variable "task experience" together indicated the latent variable "experience." The latent variable "experience" indicated the team level latent variable "shared understanding." As with the other measures included in the study, the experience measures were first submitted to preliminary analyses to determine their measurement characteristics.

Table 9
Experience Items

Dimension	Item
Task Experience	
	How would you describe your skill level at computer games like flight simulators?
	How much experience do you have with this particular flight simulator (Werewolf v. Comanche)?
	How much experience have you had using a joystick?
Social Experience	
	How much experience have you had working in teams?
	Please rate how well you know the person you will be participating with:

Performance Measures

Performance was assessed both via subjective means and via objective means. Objective performance data were collected from the mission statistics (enemies killed and mission goals achieved) reported by the simulation used in the study. Absolute numbers of mission goals achieved (targets destroyed) and enemies killed were collected and summed as "hits." Thus, an individual who achieved 15 mission goals and killed 7 enemies during the course of the simulation would be coded as having achieved 22 hits. These are true ratio level data.

Table 10
Performance Scale - Self (adapted from Wayne & Liden, 1995)

Item
Overall, to what extent did you feel that you were performing your job the way you would like it to be performed?
To what extent did your performance meet your own expectations?
Overall, to what extent do you feel you effectively fulfilled your roles and responsibilities?
Rate the overall level of performance that you observed for yourself:
What is your personal view of yourself in terms of your overall effectiveness?
If you had it to do over again, to what extent would you change the manner in which you did your job?

Table 11
Performance Scale - Partner (adapted from Wayne & Liden, 1995)

Item
Overall, to what extent did you feel that your partner was performing his or her job the way you would like it to be performed?
To what extent did your partner's performance meet your own expectations?
Overall, to what extent do you feel your partner effectively fulfilled his or her roles and responsibilities?
Rate the overall level of performance that you observed for your partner:
What is your personal view of your partner in terms of his or her overall effectiveness?
If you entirely had your way, to what extent would you change the manner in which your partner did his or her job?

Subjective performance measures were collected via questionnaires or rating scales. Both self and partner performance ratings were collected from participants. Performance scales were taken from Wayne and Liden (1995). The self version of the performance scale is found in Table 10 and the partner version of the performance scale is found in Table 11. Both instruments were scaled from 1 to 5, where 1 means "not at all" and 5 means “to a great extent.” Thus, a higher score indicated higher performance or distal SA, and a lower score indicated lower performance or distal SA.

All three performance measures, the objective measures and both self and partner subjective measures, were included in the analyses as observed variables indicating the latent trait performance for team level analyses.

Finally, limited demographic data, in the form of age, sex and race were collected for sample reporting purposes. These data were not included in statistical analyses or model testing, but were reported only in summarized form for the purposes of describing the sample used in the study. The last section of the questionnaire booklet contained items asking participants for this information.

Procedure

Participants completed the trial in teams of two. Participants arrived at the testing site, were introduced to the study and given 10 minutes of training on the simulation. The training involved a brief introduction to the simulation and an opportunity to practice “flying” the simulation. The training materials discussed the mechanics of controlling the simulation, including information on weapons options. The training materials can be found in Appendix F. As discussed earlier, research by Jentsch et al. (1997) demonstrated that this simulation is very easy to learn; in Jentsch et al. (1997), all participants were able to perform a core of operations on the first attempt without practice.

The training materials introduced participants to the idea that they would be operating together in the same virtual environment. Participants were instructed to work as a team to complete their mission, although they were not given specific instructions with regard to their teamwork, such as how they should divide their tasks or roles. The experimenter brought the mission assignments up on the screen, and participants were told to click the "accept" button on the computer screen when they were ready to begin. In this way, participants were given the opportunity to strategize or otherwise prepare prior to initiating the simulation. However, they were given no instructions to prepare in any particular way; they were simply instructed to "click the 'accept' button when you are ready to begin."

As a means to control variability, all participants completed the same scenarios or "missions," with the same goals, weather conditions, terrain and enemy threats. However, they were able to freely maneuver within the simulation; they were not required to maintain a predetermined flight path. All participants were networked together as "friendlies;" that is, participants both flew the same type of aircraft on the same mission, and competed against the computer's pre-programmed responses, rather than being networked together as opposing forces.

The scenarios were selected in a predetermined order. The task was divided into three segments by the data collection breaks. A new scenario was started for each segment. If a team completed a scenario during a segment, a new scenario was started for the remaining time in the segment. Thus, each team had 20 minutes of game play, and an opportunity to play at least 3 scenarios.

Due to the fact that the primary aim of the research was model testing, no experimental manipulations were conducted. All participants received the same training, operated the simulation for the same duration and were subjected to the same measurement protocol. The aim of the research was to capture and analyze natural variations, rather than manipulated ones. The level of difficulty was manipulated by the scenario selection; each scenario progressed in a predetermined order, with identical goals and threats each time a mission was played. Thus, teams operated in a standardized environment. Teams were faced with progressively more difficult scenarios until their 20 minute trial period was up.

After training, participants played the simulation with their partner for a period of 20 minutes, during which period two brief probe measures of SA were collected. The probe breaks were scheduled at 5 and 15 minutes into the simulation, at which time the simulation was paused. Endsley (1995b) empirically demonstrated that pausing a simulation and probing for SA information does not adversely affect operator's levels of SA. The break periods did not count toward the 20 minute trial period. During the 10-minute period between probe breaks, the experimenter collected the communication sample. Tape recordings of experimental sessions were made to ensure the reliability of experimenter coding. Timekeeping resumed when the simulation was resumed after each break.

After the simulation period, participants were asked to complete a series of paper and pencil measures: partner performance, self performance, selected items from SART and SALIENT, the role checklist, role conflict and ambiguity scale and the task and partner experience items. Upon completion of the measures, participants were thanked, offered their choice of incentives and dismissed. While participants were completing their paper and pencil measures, the experimenter completed the SART, SALIENT and the communication rating scale for both participants. Results were submitted to statistical analyses.

RESULTS

Measure Validation

Statistical Analyses

Several sets of statistical analyses were conducted. Measures were validated using exploratory principle components analysis with varimax rotation, unless otherwise noted. Eigenvalues greater than 1 were used as the criterion for identifying significant factors in the principle components analysis. Items were placed in the factor for which they had the highest rating; these placements are indicated in bold in the tables. Tabachnick and Fidell (1983) suggest that a factor loading of greater than .30 is considered acceptable. Comrey (1973) suggests that factor loadings of .30 are considered to be poor, but that a factor loading greater than .45 can be considered fair. Therefore, a factor loading cutoff of .45 was used; no item was included in a factor if it had less than a .45 factor loading.

Reliability was assessed using Chronbach's coefficient alpha internal consistency reliability estimate. Although no firm minimum acceptable value of the internal consistency reliability estimate has been established, most test designers agree that .70 is acceptable (Golden, Sawicki & Franzen, 1990), and lower values are often used, even in top-tier journals (e.g., internal consistency reliability estimate of .62 in Earley & Mosakowski, 2000 and internal consistency reliability estimate of .61 in Garoznick, Brockner & Seigel, 2000). Thus, an internal consistency reliability estimate cutoff of .65 was used.

Analyses at this phase were conducted at the individual level of analysis; that is, each of the 272 participants was treated as an individual unit of analysis. In the interest of space and brevity, where multiple principle components analyses were conducted on different versions of the same measure (such as a participant, partner and experiment rating of using the same items), results of all principle components analyses are reported in a single table. The correlation between the refined scales and objective performance (hits) was reported as a preliminary indicator of measure validity. Descriptive statistics for all measures included in the study can be found in Appendix G.

Proximal SA

Although Endsley's (1989) SAGAT was dichotomously scored and was never intended to be a unitary measurement, there is no reason that scales devised using the probe methodology in general should not be unitary. The scoring scheme used for the proximal SA measure in this research was essentially an index of accuracy (how accurate was the participant's proximal SA). It is reasonable to expect accuracy to be fairly consistent across items. Thus, the proximal SA measure used in the present study was submitted to principle components analysis. This measure yielded mixed results, and was disappointing in terms of psychometric characteristics. Principle components analysis showed that the measure did not capture a single component; three significant factors emerged. Table 12 displays the factor loadings for all items on this scale.

A criticism of measures which interrupt the situation with probe questions is the possibility of a cueing effect. As a test of this hypothesis, one item was repeated across breaks. A t-test revealed that participants were significantly more accurate, $t\ (230) = 2.10$, $p < .05$ ($\eta^2 = .08$) the second time they were asked about the number of enemies currently threatening. During the first break, the mean difference between participants' estimates and experimenter's observations was 3.62 ($SD = 5.66$), whereas the mean difference was 2.78 ($SD = 4.41$) during the second break; thus, participants were significantly more accurate in their estimations during the second break, suggesting that they had been more closely attending to the information asked in that item. It is noteworthy that the repeated item loaded identically in the principle components analysis both the first and second time it appeared (factor loading of .70). Removing this cued item from the scale yields the factor structure shown in Table 13.

Table 12
Principle Components Analysis for Proximal SA Measure

Item	Factor 1	Factor 2	Factor 3
Break 1, enemies threatening	**0.70**	-0.11	0.10
Break 1, current weapons	0.12	0.20	**0.71**
Break 1, targets in range	**0.62**	0.43	0.15
Break 2, current speed	-0.10	**0.90**	-0.03
Break 2, current heading (direction)	-0.13	-0.21	**0.71**
Break 2, enemies threatening	**0.70**	-0.08	-0.27
Eigenvalue	1.44	1.14	1.04
Percent variance	23.92%	18.97%	17.36%

Note. Highest factor loading for each item is shown in bold.

Table 13
Principle Components Analysis for Proximal SA Measure, Without Cued Item

Item	Factor 1	Factor 2	Factor 3
Break 1, enemies threatening	**0.83**	-0.22	-0.08
Break 1, current weapons	0.25	0.14	**0.66**
Break 1, targets in range	**0.65**	0.38	0.15
Break 2, current speed	-0.03	**0.92**	-0.03
Break 2, current heading (direction)	-0.20	-0.15	**0.79**
Eigenvalue	1.27	1.07	1.03
Percent variance	25.46%	21.36%	20.55%

Note. Highest factor loading for each item is shown in bold.

These analyses lend little clarity as to which item or items are best representative of the variable of interest, SA. Of these three factors, only one shows a significant correlation with objective performance (hits). The item that loaded highest on factor 2 correlated with performance, $r = -.17$, $p < .05$. Recall that the ratings on the proximal SA measure are difference scores and that smaller numbers are indicative of greater agreement. Thus, this result is in the expected direction. However, it seems inadvisable to use this measure in model testing.

SART

SART was measured by each participant, her or his partner and the experimenter. Principle components analyses supported separating those SART items that evaluated the game from those that evaluated the individual. Principle components analyses showed each of these to be unitary scales; thus, the values reported in Tables 14 and 15 represent unrotated factor solutions. In the interest of space, results from more than one analysis are reported in each table. Table 14 displays the unitary factor solution for the self ratings of the game items and the unitary factor solution for the experimenter ratings of the game items. Similarly, Table 15 displays the results of three separate principle components analyses; one analysis based on self ratings, one based on partner ratings and one based on experimenter ratings. Because the different scales represent different raters rather than different dimensions, it is appropriate and in fact desirable for items to have high factor loadings on multiple measures. That is, the high factor loadings were obtained for the same item on both the self rating and the experimenter rating were evidence of convergent validity, rather than of cross-loading.

Internal consistency reliability estimates were satisfactory for all subscales except self rating of the game. As a preliminary test of predictive validity, each of these subscales was also correlated with objective performance (hits). Internal consistency reliability estimates and the correlation with objective performance for each of the subscales are shown in Table 16.

Table 14
Principle Components Analysis for Situation Awareness Rating Technique (SART): Game Ratings

Item	Self rating	Experimenter Rating
Instability of situation	.76	.94
Complexity of situation	.68	.96
Variability of situation	.79	.94
Eigenvalue	1.67	2.69
Percent variance	55.60%	89.70%

Table 15
Principle Components Analyses for Situation Awareness Rating Technique (SART): Individual Ratings

Item	Self rating	Partner Rating	Experimenter Rating
Arousal	.85	.79	.89
Concentration of attention	.81	.81	.93
Division of attention	.83	.77	.95
Spare mental capacity	.73	.76	.85
Eigenvalue	2.61	2.45	3.27
Percent variance	65.13%	61.12%	81.77%

Table 16
Internal Consistency Reliability Estimate and Correlation with Performance for Situation Awareness Rating Technique (SART) Subscales

Subscale of SART	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
Game: Self Rating	.60	-.18*
Game: Experimenter Rating	.94	.52*
Individual: Self Rating	.82	.32*
Individual: Partner Rating	.79	.18*
Individual: Experimenter Rating	.93	.59*

Note. $^*p < .05$

SALIENT

Analysis of the adapted form of SALIENT used in this research yielded very positive results. Like SART, ratings on SALIENT were gathered from each participant, his or her partner and the experimenter. Despite the length of the measure, the unrotated factor matrix for each rating method yielded factor loadings that were all above .50. Results from the three principle components analyses of the scale (one from each rating source) are reported together in Table 17. Similar to the results reported for SART, due to

Table 17
Principle Components Analyses for Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT)

Item	Self Rating	Partner Rating	Experimenter Rating
Monitored environment	.61	.50	.81
Demonstrated spatial awareness	.69	.52	.76
Reported problems	.70	.68	.75
Location problem source	.71	.74	.90
Knowledge of consequences	.79	.79	.90
Resolved discrepancies	.80	.71	.88
Noted deviations	.77	.74	.83
Recognized a need for action	.66	.64	.79
Anticipated consequences	.72	.68	.88
Informed other of actions taken	.73	.72	.85
Monitored actions	.78	.74	.87
Demonstrated knowledge of tasks	.71	.62	.76
Shared attention among tasks	.75	.72	.83
Monitored workload	.80	.78	.80
Shared workload	.75	.69	.74
Answered questions promptly	.64	.55	.67
Communicated important	.79	.74	.85
Confirmed information	.76	.68	.79
Challenged information	.71	.66	.60
Re-checked old information	.68	.67	.57
Provided information in advance	.75	.79	.83
Obtained information	.77	.72	.83
Complex relationships	.80	.74	.85
Briefed status frequently	.74	.72	.69
Eigenvalue	12.95	11.54	15.24
Percent variance	53.97%	48.06%	63.48%

the fact that the different columns in Table 17 represent different raters rather than different constructs, it is appropriate and desirable for the items to have high factor loadings on multiple scales. Table 18 displays the correlations between the SALIENT scales and objective performance, and the internal consistency reliability estimates for SALIENT.

Table 18
Internal Consistency Reliability Estimate and Correlation with Performance for Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT) Subscales

Subscale of SALIENT	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
SALIENT: Self Rating	.96	.30*
SALIENT: Partner Rating	.95	.14*
SALIENT: Experimenter Rating	.97	.48*

Note. $*p < .05$

The high factor loadings, high internal consistency reliability estimates and significant correlations with objective performance are strong evidence that SALIENT is reasonably consistent across raters, that it is a unitary measurement, and that it possesses good reliability. Additionally, preliminary evidence suggests that it has good predictive validity for objective performance. Given that the statistics for SALIENT are more favorable than those for SART, it seems advisable to use SALIENT as the metric for SA in model testing analyses.

Communication

Analysis of the measures of communication were less encouraging. Although there did not appear to be a method effect (estimation versus observation), communication was not a unitary measure. Where the SART and SALIENT measures used a single set of items collected multiple times from three different raters, the communication measure was comprised of a single set of items, each of with was collected only once, although two different methods were used. Thus, the values reported in Table 19 are the results of one principle components analysis that included all items.

Table 19
Principle Components Analysis for Communication Measures

Measure	Factor 1	Factor 2
Communication frequency (estimation)	**0.82**	0.52
Communication mundane (estimate)	**0.69**	0.55
Communication emergency (estimate)	**0.59**	0.57
Communication role (estimate)	0.16	**0.82**
Communication team (estimate)	**0.78**	0.26
Communication closed loop (estimate)	0.33	**0.76**
Communication info request (observation)	**0.59**	0.48
Communication raw data (observation)	**0.82**	0.06
Communication processed (observation)	0.24	**0.86**
Communication other (observation)	**0.72**	0.28
Eigenvalue	6.05	1.07
Percent variance	60.52%	10.66%

Note. Highest factor loading for each item is shown in bold.

Table 20
Internal Consistency Reliability Estimate and Correlation with Performance for Communication Measures

Measure	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
Communication: Factor 1	.73	.05
Communication: Clarifying	.70	.18*
Communication: Processed		.24*

Note. *$p < .05$

Several items that loaded primarily on Factor 1 of the communication measure were cross-loaded; that is, they had fairly high factor loadings on both factors. This was less the case for Factor 2; items that loaded primarily on Factor 2 had much smaller factor loadings on Factor 1. This suggests that Factor 2 is a "cleaner" factor.

Internal consistency reliability estimates were calculated and predictive validity was also tested. The internal consistency reliability estimate for the items that comprised Factor 1 was acceptable, but the internal consistency reliability estimate for the three items comprising Factor 2 was only .27. Removing the communication measure for processed information resulted in a much higher internal consistency reliability estimate for the remaining two items in Factor 2. Therefore, communication was broken out into three components: Factor 1, Factor 2 (without processed) and processed communication. Internal consistency reliability estimates and correlations with objective performance are presented in Table 20. Given that the communication Factor 1 was not significantly related to performance and that it had problems with cross-loading, it seems best to include only communication Factor 2 and the processed communication item in model testing.

Examination of the content of each item supports this decision also – those items that comprise Factor 1 have been shown in previous research to be unrelated to SA, while those items that remain (Factor 2 and processed communication) do have a theoretical relationship to SA, as outlined in the introduction. The two items that comprise communication Factor 2 can be characterized as "clarifying communication" in that they demonstrated an effort on the part of the participant to address role issues and to ensure that communication was accurately understood by both parties (closed-loop communication).

Role Checklist

The role checklist was used to calculate role conflict and ambiguity, using the methodology suggested by Ilgen and Hollenbeck (1991). This methodology used the same set of items but two different calculation methods to arrive at the role ambiguity and role conflict scores. Role ambiguity was measured by the degree of certainty about whether or not a particular item was part of the participant's role. Role conflict was calculated as a difference score between the participants' rating of her or his role, and the partner's rating of the participant's role.

Although the two scales are fundamentally different metrics from the perspective that one takes into account the ratings of the partner, one of the goals was to identify those items which contributed positively to both methods of calculations. That is, the intention was to create one role checklist that could be used for calculating both role ambiguity and role conflict, rather than a separate checklist for each purpose. Therefore, those items that loaded only on one scale or the other when submitted to principle components analysis were eliminated, leaving only those items that formed a unitary scale for both methods of calculation. Eight items were identified that comprised unitary measures of both role conflict and role ambiguity. Only those items were included for the remaining analyses.

The difference between the two scales using this method is in the calculation rather than in the items themselves. Values reported in Table 21 reflect the factor solutions for both methods of calculation independently. That is, Table 21 reports the principle components analysis for the eight items calculated to reflect role ambiguity and the principle components analysis of the same items calculated to reflect role conflict.

Therefore, it is appropriate for the factor loadings for a single item to be high on both dimensions. Table 22 displays the internal consistency reliability estimates and relationships with performance for both of these calculation approaches. As role conflict is not significantly correlated with performance, role ambiguity will be used in model testing.

Table 21
Principle Components Analysis for Role Ambiguity and Role Conflict

Item	Role Ambiguity	Role Conflict
Leader	.57	.51
Responsible for setting strategy	.75	.57
Moral support	.62	.47
Navigate – decide where to go	.70	.62
Weapons expert – decide which weapons to use	.71	.68
Responsible for making task assignments	.82	.72
Responsible for monitoring partner's activities	.71	.62
Responsible for keeping perspective	.73	.63
Eigenvalue	3.98	2.92
Percent variance	49.73%	36.55%

Table 22
Internal Consistency Reliability Estimate and Correlation with Performance for Role Ambiguity and Role Conflict

Scale	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
Role Ambiguity	.85	.14*
Role Conflict	.75	-.05

Note. *$p < .05$

Role Conflict and Ambiguity Measures

Factor analysis of role conflict and ambiguity using the Rizzo, House, and Lirtzman (1970) classic measures did not fall out as cleanly as expected. While the items assessing role conflict did emerge as a unitary scale, the items for role ambiguity did not.

Further analysis revealed that one item, "I have to work under vague directives or orders" fit better with the role conflict items than with the role ambiguity items. The internal consistency reliability estimate of the role ambiguity scale was also improved by deleting

Table 23
Principle Components Analysis for Rizzo et al. (1970) Measure of Role Ambiguity and Role Conflict

Item	Role Ambiguity	Role Conflict
I feel certain about how much authority I have.	**0.69**	0.05
I have clear, planned goals and objectives for my job.	**0.75**	0.14
I know what my responsibilities are.	**0.86**	0.08
I know exactly what is expected of me.	**0.79**	0.16
I work on unnecessary things.	0.06	**0.68**
I receive a assignment without the ability to complete it.	0.21	**0.76**
I work under incompatible guidelines.	0.07	**0.84**
Eigenvalue	2.45	1.79
Percent variance	61.16%	59.69%

Note. Highest factor loading for each item is shown in bold.

Table 24
Internal Consistency Reliability Estimate and Correlation with Performance for Rizzo et al.(1970) Measures of Role Ambiguity and Role Conflict

Scale	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
Role Ambiguity	.78	.14*
Role Conflict	.65	.12

Note. *$p < .05$

this item. Therefore, for the purposes of this study, analyses were conducted with only the first 4 items of the role ambiguity measure. The three role conflict items were left as an intact scale. Principle components analyses for these scales are shown in Table 23, and Table 24 displays the internal consistency reliability estimates and the correlations to performance. Similar to the results for the Ilgen and Hollenbeck (1991) approach to measuring role conflict and role ambiguity, only role ambiguity was significantly correlated to performance. Unlike the Ilgen and Hollenbeck (1991) approach, the difference between the two measures here lies in the item content rather than the calculation.

Task/Social Experience

Analysis of task and partner experience revealed that items fell out nicely into two factors, corresponding to the dimensions of task and social experience. Factor loadings for these measures are listed in Table 25. Table 26 displays the internal consistency reliability estimates and correlations with performance.

Table 25
Principle Components Analysis for Experience Items

Item	Task Experience	Social Experience
Computer game skill	**0.78**	0.22
Comanche experience	**0.72**	-0.27
Joystick experience	**0.79**	0.25
Team experience	-0.04	**0.83**
Familiarity with partner	0.15	**0.56**
Eigenvalue	1.85	1.12
Percent variance	37.06%	22.33%

Task experience relates significantly to performance, although social experience does not. An examination of social experience at the item level revealed that although the extent to which the team members knew each other correlated modestly with objective performance ($r = .12, p < .05$), prior experience working with in a team was not correlated to performance ($r = .03$, *n.s.*). In addition, the social experience items have an unacceptably low internal consistency reliability estimate; thus, it is inappropriate to include this measure in model testing.

Table 26
Internal Consistency Reliability Estimate and Correlation with Performance for Experience Measures

Scale	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
Task Experience	.65	.46*
Social Experience	.21	.11

Note. *$p < .05$

Performance

The performance measures included both objective components (i.e., the number of mission goals achieved and enemies destroyed, collectively called "hits") and subjective components (i.e., rating scales). Both self and partner ratings on the subjective measure of performance were included in the study. Although the performance rating scale included six items, both principle components analysis and reliability were improved by deleting the sixth item. The factor loadings for the performance measures are displayed in Table 27. Table 28 displays the internal consistency reliability estimate and correlation with objective performance for subjective performance measures. As with several other measures included in this study, the different scales reflect different raters rather than different constructs, and thus it is appropriate and desirable for items to have high factor loadings on both scales.

Having established the factor structures, internal consistency reliability estimates and rudimentary predictive validity for the measures included in the study, it was important to identify those that displayed the best measurement characteristics for incorporation in the model testing phase of the analysis. Clearly, it was important to include only those measures which possessed unitary factor structures and adequate

Table 27
Principle Components Analyses for Subjective Performance, Without Item 6

Item	Self Rating	Partner Rating
Perform job as preferred	.89	.85
Performance met expectations	.80	.83
Role and responsibility effectiveness	.91	.87
Overall performance rating	.89	.85
Personal view of effectiveness	.86	.80
Eigenvalue	3.81	3.52
Percent variance	76.12%	70.42%

Table 28
Internal Consistency Reliability Estimate and Correlation with Objective Performance for Subjective Performance Measures

Scale	Internal Consistency Reliability Estimate	Correlation with Performance (hits)
Performance: Self Rating	.92	.50*
Performance: Partner Rating	.89	.23*

Note. *$p < .05$

reliability. In addition, it was preferable to include measures that have some relationship with the ultimate variable of interest, performance. These criteria left us with the following measures for inclusion in the model testing phase of analysis: SALIENT (self, partner and experimenter ratings) for Individual SA, clarifying communication and the processed communication item for communication, role ambiguity (both role checklist and Rizzo et al. (1970) measures) and task experience for shared understanding, and all three performance measures (objective, and self and partner ratings) for performance.

Bivariate Hypotheses

Bivariate hypotheses were proposed in the introduction. These hypotheses described the expected relationships between latent traits addressed in this research. The correlations displayed in Table 29 provide an opportunity to evaluate these bivariate hypotheses, by examining the correlations between the measures that comprise each latent trait included in the bivariate hypotheses. Appendix G displays the descriptive statistics and correlation table for all measures described above; Table 29 displays the correlation table for the subset of scales that were included in model testing.

Table 29

Correlation Table for Measures Included in Model Testing

Measure	*N*	Mean	*SD*	1	2	3	4	5	6	7	8	9	10
1	272	4.27	1.73										
2	272	4.26	1.58	.39*									
3	272	3.12	1.67	.51*	.47*								
4	268	0.37	0.67	.40*	.43*	.68*							
5	272	4.80	9.45	.38*	.37*	.54*	.73*						
6	272	2.86	1.02	.30*	.18*	.19*	.15*	.15*					
7	271	3.74	0.86	.10	.00	.02	.02	-.05	-.08				
8	271	2.13	0.87	.36*	.25*	.38*	.18*	.16*	.09	.16*			
9	270	41.19	26.81	.30*	.14*	.48*	.18*	.24*	.14*	.12*	.46*		
10	271	3.18	1.01	.50*	.22*	.35*	.17*	.20*	.20*	.30*	.49*	.50*	
11	271	3.93	0.74	.20*	.36*	.20*	.14*	.14*	.19*	.08	.13*	.23*	.23*

Note. *$p < .05$. Measures are listed below.

1. SALIENT: Self Rating
2. SALIENT: Partner Rating
3. SALIENT: Experimenter Rating
4. Communication: Clarifying
5. Communication: Processed
6. Role Ambiguity
7. Rizzo et al.: Role Ambiguity
8. Task Experience
9. Objective Performance
10. Subjective Performance: Self
11. Subjective Performance: Partner

In the interest of eliminating as much random error as possible, only those measures which were determined to be sound enough for use in model testing were used in evaluating the bivariate hypotheses. Due to the fact that these correlations and the evaluation of the bivariate hypotheses are primarily for descriptive purposes, no correction for Type I error was made.

All nine (100%) of the bivariate relationships that comprise Hypothesis 1, individual SA is positively correlated with performance, were statistically significant. All six (100%) of the correlations comprising Hypothesis 2, individual SA is positively correlated with communication, were statistically significant. Hypothesis 3, communication is positively correlated with performance, was also supported by all six (100%) of the bivariate relationships involved.

Hypothesis 4, individual SA is positively correlated with shared understanding, was comprised of nine bivariate relationships and, although six were statistically significant, only three (33%) were in the expected direction. Role ambiguity is inversely related to shared understanding; that is, low role ambiguity is indicative of high shared understanding. Thus, the positive relationships between individual SA and the role ambiguity measures are in the opposite direction from the hypothesis. Similar to the findings for Hypothesis 4, Hypothesis 5 was only partially supported. Three significant correlations (33%) emerged supporting Hypothesis 5, although eight of nine relationships comprising Hypothesis 5, shared understanding is positively correlated with performance, achieved significance. That is, five of the eight positive relationships were actually in the opposite direction; role ambiguity was positively correlated with performance.

Finally, of the six bivariate relationships included in Hypothesis 6, communication is positively correlated with shared understanding, only two (33%) significant correlations emerged supporting the hypothesis. Like Hypotheses 4 and 5, however, two more significant correlations emerged in the opposite direction (communication positively correlated with role ambiguity). Thus, Hypotheses 1-3 received strong support, and Hypotheses 4-6 received mixed support.

Model Testing

Statistical Analyses

Scale scores with the best measurement characteristics identified above were included as the observed variables in testing each of the five models identified in the introduction. Models were tested using Structural Equation Modeling (SEM). SEM is a general statistical modeling technique, which allows examination of the relationships among multiple variables, which can be observed or unobserved, at the same time. One advantage of SEM is that it explicitly models measurement error associated with observed variables to arrive at unbiased estimates of unobserved variables (Ullman, 1996). In addition, SEM permits reciprocal causation, that is, non-recursive models (Bentler & Raykov, 2000).

SEM is a confirmatory technique; thus, analyses with SEM begin with an *a priori* model, which is evaluated against the covariance structures of the data to determine the fit of the model to the data (Ullman, 1996). The more closely the *a priori* model fits the data, the more plausible it is that the model is an accurate explanation for the relations among the measures. SEM yields both standardized and unstandardized regression estimates (path coefficients). For ease of understanding, standardized regression estimates are displayed on the figures. Significant paths are noted with an asterisk.

Model testing analyses were performed using the Analysis of Moment Structures (AMOS) structural equation modeling program (SmallWaters Corporation, 1999). Goodness of fit of the models was evaluated using several of the more widely-used indicators. The most commonly used indicator of closeness of model fit is chi-square (e.g., Kline, 1998), thus, chi-square was calculated for each of the models. Due to the fact that the goal in model testing is to have a model that fits the data, that is, a model which is not significantly different from the data, a nonsignificant chi-square is desired. That is, it is desirable to accept the null hypothesis that there is no difference between the data and the model (Ullman, 1996).

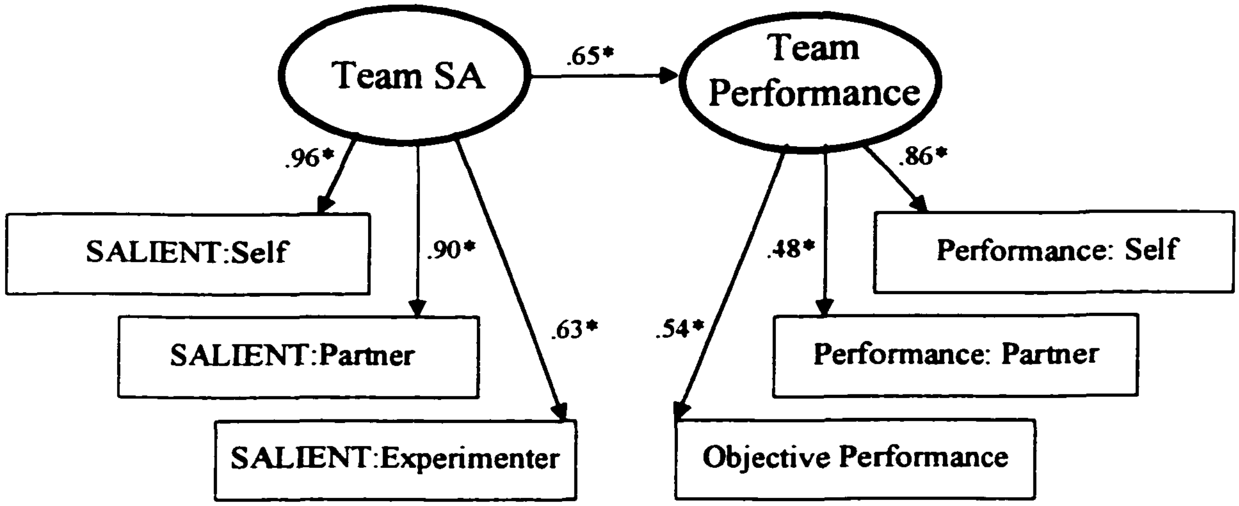
Since chi-square is affected by sample size, four other commonly used fit statistics were also included. Each of these take a comparative approach to estimating fit. That is, they compare the fit of the data to some theoretical baseline model, either completely uncorrelated (the independence model) or perfectly correlated (the saturated model; Arbuckle & Wothke, 1999). The Normed Fit Index (NFI; Bentler & Bonett, 1980) and Non-Normed Fit Index (Bentler & Bonett, 1980) are both commonly used indices, although they tend to underestimate fit in models with smaller samples, and the NNFI sometimes yields values outside the 0 – 1 range (Ullman, 1996). The Incremental Fit Index (IFI; Bollen, 1989) is another variation of the NFI and NNFI, and addresses the variability issues of the NNFI (Ullman, 1996). The last fit index used, the Comparative Fit Index (CFI; Bentler, 1990), uses a slightly different statistical approach than the NFI, NNFI and IFI, but it yields accurate estimations of fit, independent of sample size. For all four of these fit indices, values close to 1 (e.g., greater than .9) indicate a good fit (Arbuckle & Wothke, 1999).

Models were compared to determine which best fit the data using the Akaike Information Criterion (AIC; Akaike, 1987). The AIC is a composite measure, formed by a weighted sum of badness-of-fit and model complexity, where a constant *k* is applied to complexity as a relative penalty for badness-of-fit and complexity (Arbuckle, 1997). Thus, complex, poorly fitting models yield high AICs, and simple, well-fitted models yield low AICs. Although values close to zero are desirable, no agreed-upon cutoff has been established (Schumacker & Lomax, 1996). The statistic is inappropriate as a general evaluation of an individual model; rather it is designed to allow comparisons among models (Arbuckle, 1997). The statistic explicitly favors more parsimonious models.

This test is an appropriate method by which to compare the models presented here, in that it works with both hierarchical (i.e., nested) and non-hierarchical models (Kline, 1998). Although the test is imperfect, it is considered to be the best method to compare both hierarchical and non-hierarchical models for goodness of fit (Schumacker & Lomax, 1996).

Summation Model

Figure 9 displays the summation model with standardized regression coefficients for the paths. The summation model included only two latent traits, team SA and team performance, each of which was comprised of three observed variables. This model had excellent goodness-of-fit indicators, which are displayed in Table 30.



* $p < .05$

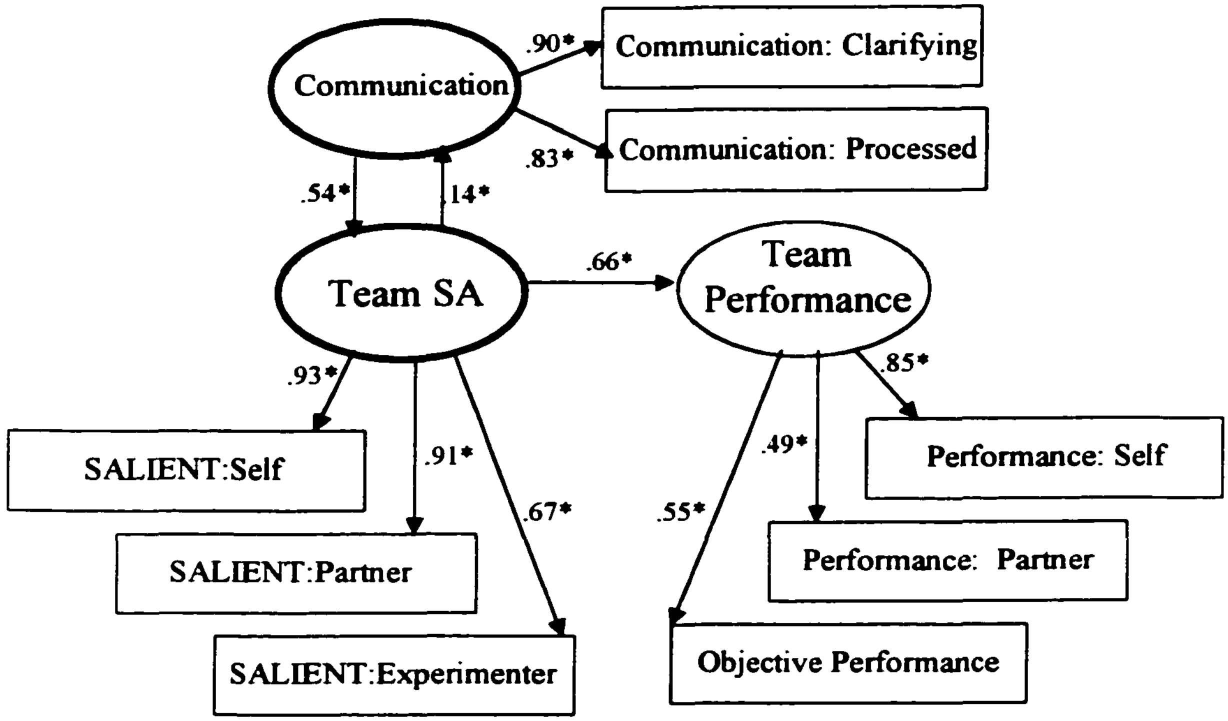
Figure 9. Standardized regression estimates for summation model

Table 30
Goodness of Fit Indices for Summation Model

Index	Value
χ^2	(8, N = 126) = 44.05, $p < .05$
Normed Fit Index	.98
Incremental Fit Index	.98
Non-Normed Fit Index	.96
Comparative Fit Index	.98
Akaike Information Criterion	82.05

Non-Recursive Communication Model

Figure 10 displays the non-recursive communication model with standardized regression coefficients for the paths. The non-recursive communication model included three latent traits: communication, team SA and team performance. Communication was comprised of two observed variables, and had a non-recursive (i.e., two-way) relationship with team SA. This model had good goodness-of-fit indicators, which are displayed in Table 31.



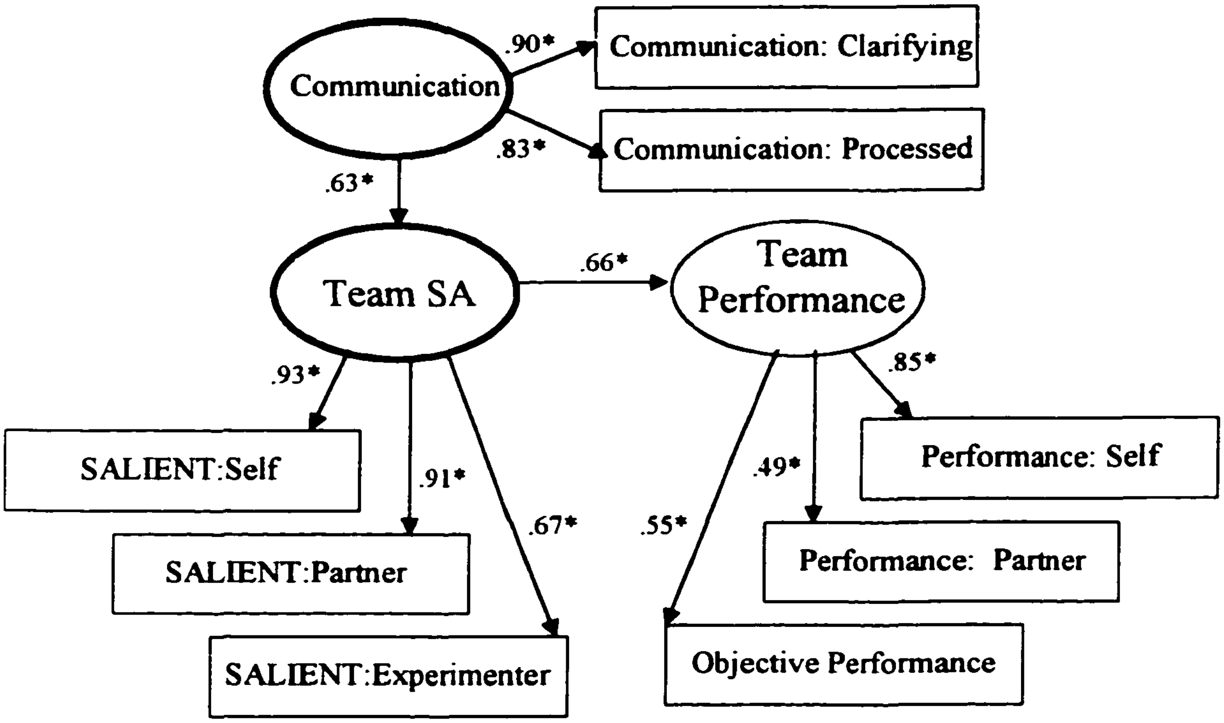
* $p < .05$

Figure 10. Standardized regression estimates for non-recursive communication model

Table 31
Goodness of Fit Indices for Non-Recursive Communication Model

Index	Value
χ^2	(18, N = 126) = 115.76, $p < .05$
Normed Fit Index	.95
Incremental Fit Index	.96
Non-Normed Fit Index	.92
Comparative Fit Index	.96
Akaike Information Criterion	167.76

Communication Model



* $p < .05$

Figure 11. Standardized regression estimates for communication model

Table 32
Goodness of Fit Indices for Communication Model

Index	Value
χ^2	(18, N = 126) = 115.76, $p < .05$
Normed Fit Index	.95
Incremental Fit Index	.96
Non-Normed Fit Index	.92
Comparative Fit Index	.96
Akaike Information Criterion	167.76

Figure 11 displays the communication model with the standardized regression coefficients for the paths. The communication model was identical to the non-recursive communication model, except that the communication model included a unidirectional path between communication and team SA, as opposed to the bi-directional relationship between those two variables in the non-recursive communication model. This model had good goodness-of-fit indicators, which are displayed in Table 32.

Shared Understanding Model

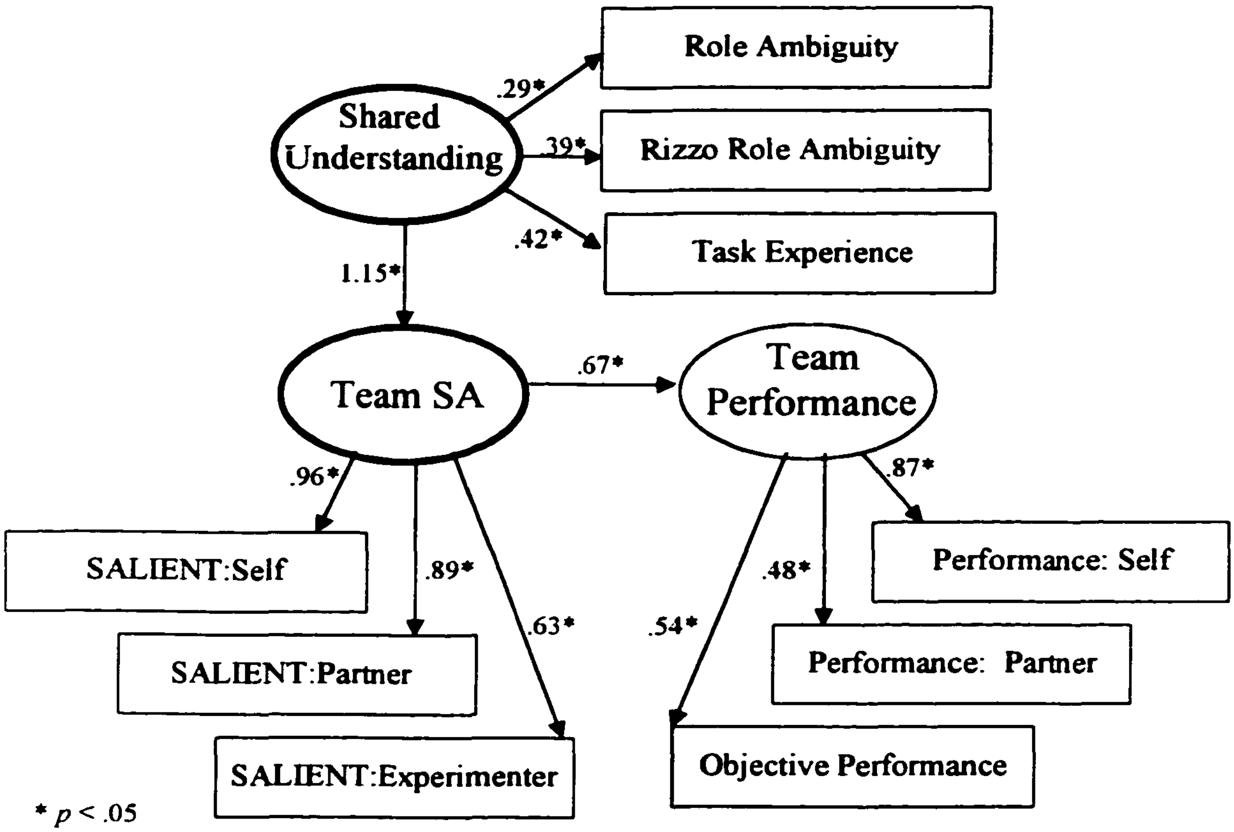


Figure 12. Standardized regression estimates for shared understanding model

Figure 12 displays the shared understanding model with standardized regression coefficients for the paths. The shared understanding model included three latent variables, team SA, team performance and shared understanding. Each of these was comprised of three observed variables. This model had excellent goodness-of-fit indicators, which are displayed in Table 33.

Table 33
Goodness of Fit Indices for Shared Understanding Model

Index	Value
χ^2	(25, N = 126) = 95.33, $p < .05$
Normed Fit Index	.97
Incremental Fit Index	.98
Non-Normed Fit Index	.96
Comparative Fit Index	.98
Akaike Information Criterion	153.33

Integrated Model

Figure 13 displays the integrated model with standardized regression coefficients for the paths. The integrated model includes all four of the latent traits from the previous models. The integrated model includes the unidirectional relationship between communication and team SA from the communication model, rather than the bidirectional relationship between these two variables tested in the non-recursive communication model. The integrated model had good goodness-of-fit indicators, which are displayed in Table 34.

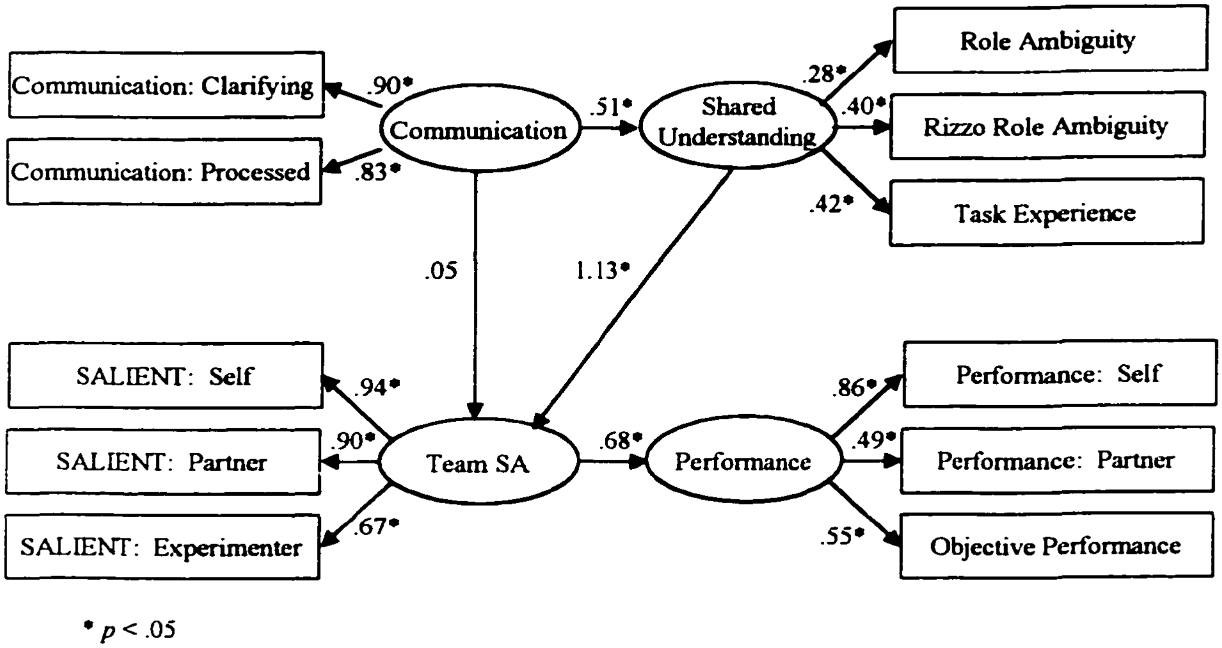


Figure 13. Standardized regression estimates for integrated model

Table 34
Goodness of Fit Indices for Integrated Model

Index	Value
χ^2	(40, N = 126) = 176.25, $p < .05$
Normed Fit Index	.95
Incremental Fit Index	.96
Non-Normed Fit Index	.94
Comparative Fit Index	.96
Akaike Information Criterion	250.25

Model Comparisons

Although the chi-square value for each of the models tested was significant, the other fit indices revealed that the each of the models fit the data well. Table 35 displays the fit indices for all the models. The AIC for the summation model is smallest; thus, it represents the best fit of the data.

Table 35
Fit Indices for Models

Models Compared	Chi-square	NFI	IFI	NNFI	CFI	AIC
Summation	(8, N = 126) = 44.05, $p < .05$	.98	.98	.96	.98	82.05
Non-recursive Communication	(18, N = 126) = 115.76, $p < .05$	.95	.96	.92	.96	167.76
Communication	(18, N = 126) = 115.76, $p < .05$	.95	.96	.92	.96	167.76
Shared Understanding	(25, N = 126) = 95.33, $p < .05$	.97	.98	.96	.98	153.33
Integrated	(40, N = 126) = 176.25, $p < .05$	.95	.96	.94	.96	250.25

DISCUSSION AND CONCLUSIONS

Two primary goals were identified for the present research. The first goal was validation of metrics for use in SA research. The primary goal, identification of a model of team SA, was dependent upon the first. Results of this research were very encouraging in terms of both of these goals. Implications for future research are discussed throughout this section.

Measurement Validation

Measures of SA

Three approaches to measuring SA were identified in the introduction: proximal meso or middle and distal. All three of these were operationalized and tested. A measure was constructed to assess proximal SA, two previously described measures, SART and SALIENT were included to test meso SA, and performance measures were included to quantify distal SA, although distal SA measures are more inferential than direct measures of SA. The proximal SA measures, SART and SALIENT are discussed below.

Proximal SA Measure

A probe-type measure of proximal SA was developed. Although this approach to measuring SA was the most common one found in the literature, upon statistical analysis, it proved disappointing. It was not unitary, and only one factor, comprised of a single item, correlated significantly with performance. Although there are examples in the literature of proximal SA techniques being effectively used with a composite measure (e.g., Chaparro et al.,1999), Endsley herself has demonstrated that the technique as she practices it is not a unitary measure (e.g., Endsley, 1998). Despite the fact that previous research has suggested that use of a freeze or probe technique does not intrude on SA

(Endsley, 1995b), a t-test revealed a significant cueing effect for the one item that was included in the probe question on the first and second breaks.

There are several possible explanations for the disappointing results for the proximal SA measure. First, it is possible that the items included in the breaks were not central enough to the task to be good indicators of SA. Second, it is possible that the participants were attending to more the naturalistic, context-rich environmental cues rather than attending to the simulated aircraft's displays. That is, it is possible that the participants were more focused on how many targets they could see out their windshield as opposed to attending to their cockpit display of mission goals within range. Given the good visual conditions of the simulation, it was possible for participants to fly using visual flight rules rather than attending to instrumentation. Participants' novice status may have contributed to a tendency to attend to naturalistic cues rather than the instrumentation. Much SA research (e.g., Bolstad & Endsley, 1999) uses impoverished displays such as a single simulated radar screen, rather than the full-graphics display provided in the *Werewolf vs. Comanche* simulation.

Third, it is possible that the novice participants were not attending to the cues in any structured way, or that they simply had poor SA. It is also possible that the dimensions captured on the measure are truly independent, and that it was inappropriate to attempt to create a composite measure from them. However, item level analyses reveal that only one of the items included in this measure achieved a statistically significant correlation with objective performance ("what is your current speed", $r = -.17$, $p \leq .05$), although this correlation was in the expected direction.

These criticisms could be overcome through the use of a different simulation or different items. It may be that probe-type measures of proximal SA are indeed useful indicators of SA when the data that they capture are most salient to the participants. If participants have only the radar to indicate the location of targets, it is possible that their SA for target location will be more accurate than if they can also "see" targets out of their windshields. That is, the additional visual information may be acting as a distractor. Thus, although the proximal SA measure was disappointing in this research, the approach may still be useful in other research.

The disappointing results related to proximal SA may have had less to do with the items comprising measure and more to do with the approach to measuring proximal SA. That is, the items themselves may not have been problematic; rather, the probe technique may be the culprit. Durso, Hackworth, Truitt, Crutchfield, Nikolic and Manning (1998; cited in Durso & Gronlund, 1999) suggested that retaining information about the situation, such as speed or current weapons setting, in memory may actually be indicative of poor resource use. That is, if the information is available in the environment, it may be indicative of good SA to know where to find relevant information, rather than attempting to maintain the information in working memory. Thus, different methods for assessing proximal SA should be tested in future research.

SART

SART yielded acceptable validation statistics. The measure is unique in that it assesses both the complexity of the game and the individual's SA. The two distinct sections of the scale were both unitary, all versions of the scale correlated significantly with performance, and the internal consistency reliability estimates for the versions of the scale were generally acceptable, although the internal consistency reliability estimate for the self rating of game complexity was borderline at .60.

Although in general SALIENT appears to be a superior method to SART for the assessment of SA, SART still holds promise as a means to answer research questions that SALIENT cannot address. For example, research investigating the interactions among situation complexity, SA and performance could be quite interesting. The brevity of the measure is also an advantage.

SALIENT

The adaptation of Muniz and her colleagues' (1997) SALIENT yielded the best validation statistics. Although Muniz et al. designed SALIENT as a sophisticated behavioral observation measure for use with expert pilots, the content performed well using a rating scale approach to assess each content dimension. This adaptation both simplified the use of the measure and made it more appropriate for novice participants. Research comparing the results of the two approaches (behavioral observation checklist and rating scale) to the measure is required to determine the relative merits of each.

Principle components analysis revealed that the 27-item scale was unitary using each of the three (self, partner and experimenter) rating techniques. Furthermore, all three rating methods yielded high internal consistency reliability estimates, and were significantly correlated with objective performance. This represents a significant advancement in that these results were achieved with the easily-generalizable adaptation of SALIENT that was developed for this study.

It must be explicitly acknowledged, however, that it is possible that some of the strength of SALIENT may be inflated due to the fact that it is a retrospective measure.

During the simulation, participants received feedback on their performance in the form of a counter that showed the number of mission goals remaining for the team in each scenario, and the visual cue of seeing when targets were destroyed. Therefore, participants had some performance data when they completed the SALIENT measure. Although the measure focuses on behavior rather than general impressions, it is possible that the measure was somewhat biased by this. Future research should validate this measure in a simulation where participants are blind to their performance.

In addition, it is somewhat disconcerting that the five proposed dimensions that comprise SALIENT did not fall out as distinct factors in principle components analysis. This suggests one of two things. Either the scale is actually unitary and the dimensions are highly correlated with each other, or the participants (and experimenter) in the study lacked the sophistication to recognize the subtle variations that would create the various factors. This again begs future research.

Other Measures

Other measures were also investigated as a part of this research. Specifically, measures of communication, shared understanding and performance were included. Validation of these measures yielded mixed results.

Communication

One model of team SA focuses on communication, but it was unclear which specific aspects of communication were most central to team SA. Thus, two measurement approaches and several categories of communication were included in the study. A 10-minute communication sample was observed between the two breaks for the probe measures of proximal SA. In addition, after the experimental task ended the experimenter made estimations about the relative frequency of different communication types across the entire 20-minute task.

Two factors emerged from principle components analysis, although many of the items that primarily loaded on the first factor cross-loaded onto the second factor. The items included in factor 2, clarifying communication, were less subject to this cross-loading, and were also notable for their content: they were precisely the types of communication that might be expected to differentiate "good" teams from average or poor ones. The three types of communication that comprised this second factor were role-related communication, closed-loop communication and processed communication. The internal consistency reliability estimate was improved by separating the processed communication item from the first two. Thus, it should not be a surprise that these items were significantly related to performance, whereas the more generic communication items comprising Factor 1 were not.

This result confirmed Hollenbeck et al.'s (1995) post hoc finding that high performing teams processed information prior to communicating it more than low performing teams did. This result also confirmed McIntyre and Salas's (1995) contention that closed loop communication is particularly critical to good team performance, especially in high-complexity, time constrained settings. The inclusion of role-related communication speaks to the importance of developing shared understanding.

On the other hand, the literature has mixed findings on the relationship of general communication and SA. Major et al. (1998/1999) suggested that high levels of team SA should decrease the amount of communication required among team members. Artman (1999) found that fewer communications were related to better performance (presumably due to greater SA), but Schreiber et al. (1996) found the opposite, that those with greater overall communications had higher SA. It is possible that these discrepant results are an artifact of the different studies' operationalizations of communication.

Item-level analysis of the relationship between overall communication frequency and performance in this study revealed a nonsignificant correlation. At first blush, it seems possible that overall frequency may be more strongly related to familiarity between partners than to performance. However, even when the extent to which partners knew each other was controlled, the relationship between communication frequency and performance remained nonsignificant. This suggests that the mere frequency of communication is not a contributing factor to performance.

Shared Understanding

Shared understanding was assessed in two broad categories in this research: shared understanding regarding roles, and shared understanding developed through experience. Both types of shared understanding are discussed below.

Role conflict and role ambiguity. Of particular interest is the role conflict and ambiguity assessment made using the methodolgy recommended by Ilgen and Hollenbeck (1991). Although not as simple to calculate as a likert-type scaling approach, the methodology was shown to have merit. Of the 16 items included on the role checklist, 8 were identified that yielded unitary scales for both the role ambiguity calculation and the role conflict calculation. Although both calculation techniques yielded acceptable internal consistency reliability estimates, only role ambiguity was significantly related to performance. It is interesting to note that this same pattern emerged with the Rizzo et al.

(1970) measure of role conflict and ambiguity, offering some evidence of convergent validity between these two measures.

It is possible that the failure of role conflict to achieve a significant relationship with performance was due to the task used in this study; it was possible for team members to act fairly independently and still perform the task adequately. Thus, role conflict may not have been as great as impediment in this task than in other, more highly interdependent tasks, or those that require distributed expertise. Research further validating the use of this methodology should be welcomed in the field.

Experience. Shared understanding was also assessed through measures of social experience and task experience in this study. Stout, Cannon-Bowers and Salas (1996/1997) suggested that one route to high levels of team SA was through high levels of shared understanding. Two types of shared understanding were assessed: social and task. Although principle components analyses were promising for these measures, only task experience had acceptable internal consistency reliability estimates. Furthermore, only task experience correlated significantly with performance.

However, it is worth noting that several significant correlations with social experience did emerge. First, social experience was significantly positive related to self ratings on both SART and SALIENT, suggesting that individuals with more social experience on teams and/or with their partner were more confident in their SA. Second, social experience was significantly positively correlated to the communication measures. Interestingly, when the social experience is broken down to its component parts, experience with teams and partners' familiarity with one another, familiarity is significantly positively correlated to all three types of communication, but team experience is only positively correlated to processed communication ($r = .13, p < .05$).

Given that mixed results were achieved for this variable, it is probably premature to dismiss it as unimportant. Rather, further research should endeavor to clarify this relationship. It is possible that the items used to measure this variable were simply inadequate to capture the complexity of social experience; only two items were used, and these assessed fairly rudimentary social experience, which may not have been sufficient. Alternatively, it may be that the task was not sufficiently interdependent for a difference based on social experience to emerge. That is, due to the fact that it was possible for participants to play the game with some success without depending on their partners, social experience may not have been an important variable.

The experimenter's casual observations revealed that partners who appeared to know each other well generally performed either very well or very poorly, and often those who had outstanding performance at the task brought along a good friend to participate with them. Examination of the item level correlations supported this; the correlation between familiarity with partner and task experience was ($r = .14, p < .05$), and between familiarity with partner and objective performance was ($r = .12, p < .05$). However, this pattern was not repeated for more general team experience.

It may be that there was an interaction between task experience and how well participants knew each other. Participants who knew each other often appeared to share an enjoyment of tasks like the one used in the simulation or to share an unfamiliarity with them; thus both their task experience and social experience were high, as was their performance, or their social experience was high, and both their task experience and performance were low. Participants who participated with a stranger did not follow such a pattern; their task experience varied more randomly, and thus their aggregated task experience tended to regress toward the mean. This is another area that warrants further investigation. However, there is a possibility, however unlikely, that social experience is simply unimportant to team SA.

Finally, it must be acknowledged that the two items included in the social experience measure were conceptually distinct. Reviewing item-level results for this measure yielded a pattern of results similar, but not identical, to those found at the scale level. Specifically, both team experience and familiarity with the experimental partner correlated significantly with self ratings on SART, SALIENT, processed communication, and self-rating of performance, as did the social experience measure. However, several other measures that correlated significantly with social experience were related to only experience with partner (and not team experience): experimenter rating on SALIENT, communication Factor 1, clarifying communication, task experience, objective performance and partner rating of performance. Interestingly, role conflict (using the Ilgen and Hollenbeck, 1991, methodology) was related to team experience only at the item level. That is, the more experience team mates had with working in teams, the less they agreed about their respective roles. Scales that specifically assess social experience should be developed in future research.

Performance

Subjective and objective measures of performance were included in this research. Subjective measures of performance were significantly positively correlated to objective measures of performance, which is evidence of convergent validity between the measures.

Both self and partner ratings of subjective performance were unitary scales, and had high levels of internal consistency reliability. Thus, either approach could probably be used in future research. This represents an advantage in that performance in SA contexts is not always easily quantifiable. For example, it may be difficult to quantify individual "performance" in fire fighting or disaster relief efforts. However, this research has demonstrated convergence between subjective performance ratings and objective measures of performance, thus clearing the way for subjective measures to be used when objective measures may be impractical.

Bivariate Hypotheses

Bivariate hypotheses were generally supported. Hypotheses 1-3, which address the relationship between individual SA and performance, between individual SA and communication and between communication and performance were all universally supported by the correlations. Hypotheses 4-6, which address individual SA and shared understanding, shared understanding and performance, and communication and shared understanding all had mixed support.

Further examination reveals that the mixed results for Hypotheses 4-6 were due to measures of role ambiguity. Specifically, where these measures should have been negatively related to the outcome measures and other study measures, in every instance the correlations were either positive or nonsignificant. However, it would be premature to take these results as evidence that roles are not important to effective team performance.

There are several possible reasons why the relationships between role ambiguity, as measured using the Ilgen and Hollenbeck (1991) methodology, and the other study measures were positive rather than negative. First, it is possible that teams who devoted effort to role negotiation had less time to devote to the task, thereby lowering their performance. This would yield a pattern of high role ambiguity (little effort spent on role negotiation) and high performance (more time to devote to the task). However, this possibility does not account for the positive relationship between role ambiguity and communication; if this were the causal mechanism, then high levels of role-related communication in particular should be related to low levels of role ambiguity, which was not the case. Instead, it is possible that this positive relationship between role ambiguity and role-related communication is due to the efforts of participants to resolve their role ambiguity. That is, participants may have been unclear about their roles and may have engaged in role-related communication in an effort to clarify.

A second possibility is that there is an overconfidence effect. That is, those individuals who were least competent at the task were nonetheless quite confident about their competence (Kruger & Dunning, 1999). If this were the case, then these individuals who had higher levels of role ambiguity would also have higher actual levels of performance, and those with very low levels of role ambiguity (those who felt very clear about their roles and very competent at the task, yet were actually incompetent) would have had lower objective performance, yielding the positive correlation that was observed. However, if this were the causal mechanism, then self ratings of performance would have been inflated for those individuals with poor performance and low role ambiguity, and this effect did not emerge.

The third possibility is that the novice participants in this simulation who were more flexible in their roles were more successful. Where members of mature teams presumably are fairly clear with regards to who will play what role under what circumstances, the participants in this study had only 20 minutes of working together on the task. It is likely that each participant in the more effective teams played several different roles at different times during the simulation, trying various role configurations to determine which were most effective. Since the roles that they played were not consistent across the entire period of game play, it was not accurate for them to express confidence that particular roles were highly descriptive of them. Less effective teams, on the other hand, may have been more clear with regards to their roles, to the detriment of their performance. That is, they may have become too focused on a particular role and missed opportunities to enhance overall team performance. This would account for the constellation of findings associated with this variable.

Bolstad and Endsley (1999) showed that teams that started in a condition designed to facilitate shared understanding initially had performance that was poorer than teams without the shared understanding information, but later outperformed teams in other conditions. The present study did not include an explicit element of time, but the findings with regards to role ambiguity are consonant with the findings of Bolstad and Endsley (1999). Future research should include the element of time in investigations of the relationship between roles, shared understanding, SA and performance.

Lending further support to this hypothesis about the curious positive relationship found between role ambiguity and performance is the positive relationship between communication (notably the role-related communication component) and role ambiguity. This suggests that teams who were actively negotiating roles (and therefore may still have some level of role ambiguity) were more effective.

The fourth possibility is that this is not an anomalous finding at all, but entirely appropriate. That is, to the extent that roles actually are "subjective, personal, dynamic and specified by a variety of social sources" (Ilgen & Hollenbeck, 1991, p. 174), it may be that a certain level of role ambiguity is reflective of good "role work" in a team, rather than a failure on the part of the team. Recall that, in contrast to jobs, which are static, roles are naturally and by definition fluid. Empirical investigation may be able to determine whether there is an "optimal" level of role ambiguity, and whether that level changes with time. Additionally, it may be possible to determine empirically if there is an optimal balance between jobs and roles (somewhere between Ilgen and Hollenbeck's 1991 "loose cannon" and bureaucratic job-role combinations; see Figure 3) for teams that must operate in contexts that require SA. Clearly, this is an area for future investigation.

All but one of the nonsignificant relationships within the proposed bivariate hypotheses involving shared understanding included the Rizzo et al. (1970) measure of role ambiguity. Of the 10 bivariate relationships between this measure and the other measures included in this analysis, only 3 were statistically significant. Notably, the three that did achieve significance were positive; that is, higher role ambiguity was related to higher performance. As these few significant findings were in the same direction as those including the Ilgen and Hollenbeck (1991) approach, this suggests that the failure to achieve significance may be due more to the measure than to the nature of the relationships between the variables. That is, this measure may simply be less sensitive than the measure using the Ilgen and Hollenbeck (1991) methodology. Alternatively, it may be that this measure is less subject to the overconfidence effect that may be causing the Ilgen and Hollenbeck (1991) measure to positively correlate. Again, further research using both methods for assessing role conflict and ambiguity is required.

Model Testing

The primary goal of this research was to identify a model of team SA. Several models have been described, directly or indirectly, in the literature. Specifically, a summation model, both recursive and non-recursive communication models, a shared understanding model and an integrated model were tested. Each of these yielded excellent fit statistics. In fact, the greatest difference between the best-fitting and worst-fitting model was only .04. Thus, although the AIC suggested that the summation model, the most parsimonious model tested, fit the data best, it is impossible to emphasize enough that these models all essentially fit the data about equally well.

It seems reasonable to suggest that the similar results found for the models is due to the fact that each model tested is built upon the basic summation model. Additional variables were added into the models to examine their relationship with team SA, but the foundational model remained intact in all models tested. It is interesting to note that the standardized regression coefficients for the relationship between team SA and performance (the two latent variables in the summation model) were relatively stable across models; the estimates ranged from .65 to .68.

Despite the fact that the AIC favors the summation model, it would be a mistake to dismiss the other models put forth as unimportant. The fit indices for all models were very high, and all models yielded similar patterns in the indices. In addition, it needs to be highlighted that this research was conducted with novice participants, who worked at a novel simulation task for relatively brief time periods. Thus, it is possible that, given different research parameters, very different results could emerge.

Summation Model

Several noteworthy results emerged from model testing. As mentioned above, the summation model was determined to be the best fit to the data by the AIC. This model also had the smallest chi-squared statistic. This makes clear the centrality of individual level SA, at least for novice operators and novel tasks.

It must be acknowldeged that the summation model had somewhat of an unfair advantage over the other models as far as the AIC is concerned. This statistic explicitly punishes complexity. Therefore, it is possible that the summation model nosed out the other models simply by virtue of its parsimony, as opposed to the closeness of its fit to the data. However, it should also be acknowledged that the summation model also had the smallest chi-squared statistic, and its other fit indices were in every case equal to or higher than the next highest model's indices (although, as mentioned above, these differences were negligible).

It is interesting to note that the summation model seems to suggest that teams requiring SA are most like McGrath et al.'s (1995) task forces, rather than crews or teams. That is, the summation model suggests that aggregated individual SA is the most important predictor for team performance. McGrath et al.'s task forces are characterized by the primacy of tasks, followed by members, with tools being the least central (although still important) element. Furthermore, in a task force, according to McGrath et al., the two most critical intersections are division of labor and job structure, with role network being the least important of the intersection element. If in fact the summation model is supported in future research as most important, this theory suggests that role networks are less important to team (task force) performance than are the division of labor and job structure. This may account for the disappointing findings around role conflict and role ambiguity.

Communication Models

Both communication models received support. Examination of the non-recursive communication model reveals that the reciprocal relationship between communication and team SA is considerably stronger from communication to team SA than the other way around. Removing the feedback loop from team SA back to communication did not have any effect on the fit statistics, but it did increase the strength of the regression coefficient between communication and team SA to .63 from .54. Although communication was demonstrated to be important to team SA, no support was found for Shrestha et al.'s (1995) hypothesis that communication acts as a moderating variable. Instead, as was demonstrated by the integrated model, the relationship between communication and team SA is mediated by shared understanding.

Shared Understanding Model

Examination of the results from the shared understanding model revealed a strong regression coefficient between shared understanding and team SA. Again, results supported the validity of this model. It should be duly noted, however, that shared understanding in this model was operationalized as task experience and the two measures of role ambiguity. Thus, the results regarding shared understanding should be replicated before any substantive claims about the centrality of shared understanding to team SA are made.

The central premise put forth by Stout et al. (1996/1997) was that multiple routes were available by which high levels of team SA could be achieved. Specifically, team SA could be achieved either through high levels of shared understanding or through high levels of individual SA. Our team level analysis confirmed that team SA was indeed built from both shared understanding and individual SA. Stout and her colleagues (1996/1997) in essence suggested that the two components could compensate for one another. It may be that some minimum level of one or both are required for adequate team SA. Future research should examine the extent to which these two factors are complementary.

The relatively brief duration of the simulation (20 minutes of data collection during game play) also posed a disadvantage as far as shared understanding. The very nature of shared understanding suggests that it must evolve over time. Participants in this research had little opportunity for shared understanding between them to evolve. Recent research by Bolstad and Endsley (1999) suggests that the beneficial effects of shared understanding become more pronounced over time. The present study did not include an element of time in the analyses. This represents another clear need for future research.

Integrated Model

The strong relationship between shared understanding and team SA was stable when communication was added in the integrated model. However, the significant regression coefficient between communication and team SA seen in the two communication models evaporated when the integrated model was tested. Instead, a .51 regression coefficient emerged between communication and shared understanding, and the relationship between shared understanding and team SA remained strong. The relationship between communication and team SA dropped to nonsignificance.

Thus, although the complexity of the integrated model is a disadvantage for it statistically, interesting relationships that would not otherwise have emerged were revealed through testing this model. Most interesting was the relationship between communication and shared understanding, which had a higher regression coefficient than any of the direct indicators of shared understanding, and the fact that the relationship between communication and team SA appears to be mediated by shared understanding. This suggests that it is important to re-think the role of communication in SA, as well as the importance of shared understanding.

It seems intuitively reasonable that the primary importance of communication for SA is in developing and maintaining shared understanding. Certainly the regression coefficients for the integrated model support this. However, as mentioned above, shared understanding as it related to team SA in this research was a product of task experience and two measures of role ambiguity. This is clearly an improvished measurement of shared understanding. Future research should take a rigorous mental models or situation models approach and test the relationships among shared understanding, communication and team SA.

This relationship between communication and shared understanding may also have been inflated by the content of the communication measure. This latent variable was comprised of two observed variables. Only three items contributed to this latent variable, and one of those three was role-related communication. Thus, it is possible that shared understanding (comprised largely of role ambiguity) was strongly predicted by communication because communication was built, to some extent, out of role-related communication. This is clearly an area for future research to investigate and clarify.

It is possible that the relationship between communication, shared understanding and team SA was an artifact of the impromptu teams and novel task used in this research. That is, participants did not arrive with shared understanding of their roles and the task at hand, but rather had to build their understanding during the simulation period. It is possible that experience or training could substantially alter this constellation of relationships. Research examining the roles of communication and shared understanding in developing team SA in experienced operators may yield different results.

Even given the possibility that the intriguing findings in the integrated model are an artifact of the novice participant population, the results are important. The central role of shared understanding, and the substantial contribution of communication to that shared understanding, suggest that training for team SA should attend carefully to the shared understanding being developed by trainees. A significant component of this attention should be communication monitoring.

Summary

This research made several important contributions to the body of team SA research. First, it empirically identified the strengths and weaknesses of measures of SA. SALIENT, the best measure identified here, has good measurement characteristics, whether used for self rating, rating by a partner or experimenter rating. It is unitary, has high levels of coefficient alpha reliability and correlates nicely with performance. Thus, this research found that measuring SA at the meso level was most appropriate.

Second, the research validated the methodology proposed by Ilgen and Hollenbeck (1991) for assessing role conflict and role ambiguity. Eight items were identified that were unitary across calculation methods, and had acceptable levels of alpha reliability. Although only role ambiguity was related to performance in this research, it is premature to conclude that this method is not a good one for calculating role conflict. The failure of the standard measure of role conflict (Rizzo, House & Lirtzman, 1970) to achieve a significant relationship with performance may suggest that role conflict was simply not an important element in this task, rather than suggesting that there is something inherently wrong with the measure. Alternatively, it is possible that the time period was simply to short to establish clear role sets.

Third, this research confirmed that each of several models of team SA has merit. While the summation model yielded the best fit to the data, the difference in fit statistics is essentially negligible. It is premature, therefore, to dismiss the contributions made by the other models. Of particular interest is the mediating role that shared understanding plays between communication and team SA.

Study Limitations

Novice Participants

This study had a few major limitations. First, its use of novice participants, while appropriate for the goals of the study, significantly limits generalizability. First, novices are typically not put in settings where high levels of SA are required. Typically, firefighters, aircraft pilots and anesthesiologists are extremely knowledgeable and have had tremendous amounts of skill practice prior to being expected to perform "for real." It is possible that novice behavior and performance in any SA simulation or task is categorically different from expert behavior or performance in that simulation or task. To the extent that this is true, any research that utilizes novices lacks external validity.

There is evidence that experts process information differently than novices. For example, experts often have the benefit of automatic processing of information (e.g., Schneider & Shiffrin, 1977). Experts also have been shown to recognize cues more completely and more quickly than novices (Means, Salas, Crandall & Jacobs, 1993). Finally, has been well known for a long time that experts are able to "chunk" related information in working memory far better and more accurately than novices (e.g., Chase & Simon, 1973; Miller, 1956) Durso and Gronlund (1999) suggested that experts may be able to utilize long-term working memory (Ericsson & Kintsch, 1995) to efficiently retrieve information from long term memory, thereby lessening the load on working memory. Novices' inexperience makes them unable to take advantage of this mechanism. From an SA perspective, all of this suggests that perhaps expert operators can leap from Level 1 SA to Level 2 SA far more easily than can a novice.

This in itself is not sufficient reason to discount the research, however. First, it should be acknowledged that all experts were once novices themselves, and studying novices may offer clues to means by which to accelerate the learning process. Second, it should be acknowledged that, while the high-risk, high-complexity contexts that are typically studied in SA research are generally the domain of expert operators, SA is important, although possibly less critical, in dozens of more mundane tasks that people engage in every day; walking across campus, supervising children, playing basketball, and even mowing the lawn all require SA.

Proximal SA Measure

The use of only one proximal SA measure was a limitation of this research. The disappointing results of this measure precluded any testing of the relative importance of proximal SA as a predictive mechanism or a limiting factor on overall levels of SA. Given that proximal SA is, at least in theory, an essential foundation for SA at levels 2 and 3, exploration of level 1 (proximal) SA seems a necessary component to understanding the mechanisms by which SA is built and maintained.

Parsing out the importance of different levels of SA on ultimate performance may be more than a theoretical exercise. Durso and Gronlund (1999) suggested that the levels of SA may be inextricably intertwined. It has been long established (e.g., Meyer & Schvaneveldt, 1971) that recognition of a pattern automatically activates other related components that were not perceived. That is, the pattern recognition literature suggests that the leap from Level 1 SA to Level 2 SA to Level 3 SA may be more automatic and less distinct than SA researchers have suggested.

The naturalistic decision making literature suggests a similar linkage between perception (Level 1 SA) and the other 2 levels of SA, at least for expert operators. According to Klein (1989), recognition (Level 1 SA) and deciding what to do about it (planning – Level 3 SA) are actually inseparable parts of the same process, at least for experts. Research comparing experts and novices in other domains (e.g., Wright, Pleasants & Gomez-Meza, 1990) suggests that experts are able to anticipate future events very accurately on the basis of advance cues currently available. Therefore, rather than arriving at a decision as a result of a laborious process of perception, integration, evaluation and planning, experts may recognize the future and the appropriate course of action as they are perceiving the present. These recognition-primed decisions are based on experience with a variety of situations, and retrieval of the appropriate course of action, according to this theory, follows fairly automatically from recognition of the situation itself. Empirical evidence in the naturalistic decision making literature supports this: Kaempf, Lkein, Thordsen and Wolf (1996) studied expert command-and-control officers, and determined that 87% of decisions conformed to this pattern.

Given the extent to which the three levels of SA play a foundational role in SA theory, empirical testing to determine the extent to which these levels of SA are actually distinct components is sorely needed.

Operationalization of Social Experience

The operationalization of social experience was the weakest operationalization in the study. Specifically, social experience was operationalized as two items, team experience and familiarity with partner, that, while both forms of social experience, one general the other specific, are not necessarily related to each other in a meaningful or reliable way. Indeed, the correlation between the two items barely achieved statistical significance ($r = .12$, $p < .05$). Although similar correlations emerged for the two items with a variety of measures included in this research, notably self ratings of performance, SART, and SALIENT, and for processed communication, the patterns were by no means identical. Interestingly, and contrary to expectation, generic team experience was positively correlated to role conflict (calculated using the Ilgen and Hollenbeck, 1991, methodology) but specific team experience was not (generic, $r = .20$, $p < .05$; specific, $r = -.06$, *n.s.*).

Future research, however, should carefully operationalize social experience, attending both to general social experience (such as experience in teams) and specific social experience (such as familiarity with team members). A solid scientific understanding of the influence of both of these factors on team SA could inform decisions about training for team SA as well as about team composition.

If generic social experience is found to be important to maintaining team SA, then mechanisms by which to provide trainees with the requisite social experience should be added to curricula. Such experience could come in a variety of forms. However, if generic social experience is found to be inconsequential to team SA, then perhaps training efforts should focus more toward the technical aspects of the task at hand. This research found a significant correlation between generic social experience and processed communication. Research identifying what generic team skills can be learned and applied in a team SA setting may also contribute to improved team SA performance.

Findings related to specific social experience may also assist in improving team performance. If it is determined that specific social experience is important to team SA and ultimately to success, then perhaps teams should be viewed less as an assembly of interchangable components and more as an intact whole. Additionally, perhaps it would be wise to provide additional opportunities for specific social experience where feasible. Clearly, in a combat setting, it will not always be possible to send a team of whatever operators one can scrape together off for a week at a ropes course to "get to know one another," but it may be possible to foresee likely team constellations, and provide the future team members with opportunities to work together in advance of their actual need to do so.

If one accepts the premise that specific social experience is important because the team members know what to expect of one another, one is essentially acknowledging the importance of roles. Thus, it can be said that specific social experience is built on a set of role episodes among the team members. Fink and Major (1999) suggested that team SA can be improved by addressing roles. Specifically, they suggest that role negotiations should take place before entering the SA context, and that a climate of openness and trust be fostered within the team, to facilitate on-the-fly role negotiations as necessary. It may be possible to train operators to engage in more efficient, explicit role negotiations, thereby accelerating the familiarization process.

Data Recycling

The fact that this research used the same dataset for measure validation and for model testing is mildly problematic, in that any idiosyncrasies of the dataset would have affected both sets of analyses. Research designed to replicate and cross-validate these findings is necessary and would be welcomed.

Team Task

The primary limitation of this research, however, was the team task. Although the task has been shown as appropriate for team SA research (Jentsch et al.,1997) the task itself did not require high levels of interdependence to function, and it was not necessary for team members to view themselves collectively. McIntyre and Salas (1995) suggest that an identity as a team and levels of interdependence are both important to enabling good teamwork behaviors. This research did not systematically ensure that either of those conditions existed. Although this research suggested that command and control teams (Klimoski & Jones, 1995) were most likely to require SA, the marginal interdependence meant that this task did not simulate a command and control team particularly well. Partners in this research had only outcome interdependence, rather than process interdependence. Partners were able to work together to achieve very high performance, but it was also possible for them to work fairly independently. It would have been possible for one partner to have continued to play and enjoy whatever level of success she or he was capable of, even if her or his partner was asleep at the controls. Clearly, this is not the case in most team SA applications.

It is quite possible that the results of the model comparisons would be quite different in a more highly interdependent task. This is not to suggest that *Werewolf v. Comanche* should be abandoned as a testing platform, but rather to suggest that future researchers exercise more creativity in carefully structuring interdependent tasks. *Werewolf v. Comanche* includes a variety of display settings, and interdependence could be fostered by giving team members different displays, such that none had all the display information necessary to successfully play the game. That is, it may be possible to configure the game such that team members have both process and outcome interdependence. Such configurations should be tested in future SA research using this platform.

Conclusions

Originally, this research set out to identify the extent to which team SA was an important factor in performance, and the mechanisms by which team SA was built. Secondary to that goal, the research also set out to identify the best available measure of SA.

Both of the goals of this research have been achieved. Team SA has been shown to be an important predictor for team performance, even with novice participants and a novel task. Both of the contributing factors hypothesized to be important to team SA, communication and shared understanding, were shown to be significantly related to team

SA. Furthermore, testing of the integrated model demonstrated an interrelationship between these two elements; shared understanding was shown to mediate the relationship between communication and team SA.

The secondary goal of this research, to identify the best available measure of SA, was also achieved. The modified version of SALIENT used in this research was shown to be reliable and valid. In addition, the measure is behaviorally-based, which lends it face validity, and it is simple to use. As part of the research process, however, another popular measure of SA, SART was also shown to be an acceptable method by which to capture SA, and a new methodology for assessing role conflict and ambiguity was tested. SART is also simple to use and face valid, although it lacks the behavioral focus that makes SALIENT so appealing. The Ilgen and Hollenbeck (1991) approach to capturing role ambiguity and role conflict was demonstrated to be feasible and reliable, and preliminary evidence of validity was established.

Thus, this research accomplished the goals originally set for it. Although considerable research is still required in this field, this research has established a model of team SA and a method by which to investigate it.

REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. A. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85-104.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-322.

Arbuckle, J. L. (1997). *AMOS User's Guide Version 3.6.* Chicago: SmallWaters Corporation.

Arbuckle, J. L, & Wothke, W. (1999). *AMOS 4.0 User's Guide.* Chicago: SmallWaters Corporation.

Artman, H. (1999). Situation awareness and co-operation within and between hierarchical units in dynamic decision making. *Ergonomics, 42*, 1404-1417.

Atkinson, R. C. (1963). A variable sensitivity theory of signal detection. *Psychological Review, 70*, 62-77.

Baker, D. P, Stout, R. J., Salas, E. Fowlkes, J. & Cannon-Bowers, J. A. (1998). *The Utility of Interviews and Guided Verbal Reports for Knowledge Elicitation.* Unpublished report.

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance strucutres. *Psychological Bulletin, 88*, 588-606.

Bentler, P. M. & Raykov, T. (2000). On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology, 85*, 135-131.

Biddle, B. J. (1979). *Role theory: Expectations, identities and behaviors.* New York: Academic Press.

Billings, C. E. (1994). Situation awareness in complex systems: A commentary. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 321-325). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Boeing Commercial Airplane Group (1998). *Statistical summary of commercial jet airplane accidents: Worldwide operations 1959-1997.* Seattle, WA: Airplane Safety Engineering, Boeing Commercial Airplane Group.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods and Research, 15,* 375-384.

Bolman, L. (1979). Aviation accidents and the theory of the situation. *Resource management on the flight deck.* Moffett Field, CA: NASA Ames Research Center (NASA Conference Publication 2120), 31-58.

Bolstad, C. A., & Endsley, M. R. (1999). Shared mental models and shared displays: An empirical evaluation of team performance. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 213-217). Santa Monica, CA: Human Factors and Ergonomics Society.

Breaugh, J. A., & Colihan, J. P. (1994). Measuring facets of job ambiguity: Construct validity evidence. *Journal of Applied Psychology, 79,* 191-202.

Brickman, B. J., Hettinger, L. J., Stautberg, D. K., Haas, M. W., Vidulich, M. A., & Shaw, R. L. (1999). The Global Implicit Measurement of situation awareness: Implications for design and adaptive interface technologies. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current Research and Trends* (pp. 160-164). Mahwah, NJ: Lawrence Erlbaum.

Carretta, T. R., Perry, D. C., & Ree, M. J. (1994). The ubiquitous three in the prediction of situational awareness: Round up the usual suspects. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 125-137). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Chaparro, A., Groff, L., Tabor, K., Sifrit, K., & Gugerty, L. J. (1999). Maintaining situational awareness: The role of visual attention. *Proceedings of the Human Factors and Ergonomics Society 43rdh Annual Meeting* (pp. 1343-1347). Santa Monica, CA: Human Factors and Ergonomics Society.

Chase, W. G., & Simon, H. A (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Companion, M. A. (1994). Situational awareness in emergency management systems: An overview. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 282-289). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Comrey, A. L. (1973). *A first course in factor analysis.* New York: Academic Press.

Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1981). *The experience of work.* London: Academic Press.

Dansereau, F., Graen, G., & Haga, W.J. (1975). A vertical dyad approach to leadership within formal organizations. *Organizational Behavior and Human Performance, 13*, 46-78.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay & M. T. H Chi (Eds.), *Handbook of applied cognition* (pp. 283-314). New York: John Wiley & Sons.

Earley, P. C., & Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal, 43*, 26-49.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors and Ergonomics Society – 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1989). A methodology for the objective measurement of pilot situation awareness. In *Situation awareness in aerospace operations* (AGARD-CP-478, pp. 1.1-1.9). Copenhagen, Denmark: NATO-Advisory Group for Aerospace Research and Development.

Endsley, M. R. (1994). Addendum – Situation Awareness: Some reflections and comments. In R. D. Gilson, D. J. Garland & J. M Koonce (Eds.), *Situational awareness in complex systems* (pp. 315-317). Daytona, FL: Embry-Riddle Aeronautical University Press.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*, 65-84.

Endsley, M. R. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics*

Society 42nd Annual Meeting (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (in press). Situation awareness in aircraft maintenance teams. *International Journal of Industrial Ergonomics.*

Endsley, M. R., & Rogers, M. D. (1996). Attention distribution and situation awareness in air traffic control. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting,* (pp. 82-85). Santa Monica, CA: Human Factors and Ergonomics Society.

Ericsson, F. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

Farley, T. C., Hansman, R. J., Endsley, M. R., & Amonlirdviman, K., (1999). *The Effect of Shared Information on Pilot-Controller Situation Awareness and Re-route Negotiation.* Paper presented at the 10^{th} International Symposium on Aviation Psychology, Columbus, OH.

Fink, A. A., & Major, D. A. (1999). A role theory approach to strategizing for team situation awareness. In M. Scerbo & M. Mouloua (Eds.) *Automation technology and human performance: Current research and trends* (pp. 171-175). Mahwah, NJ: Lawrence Erlbaum Associates.

Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors, 37*, 149-157.

Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist, 39*(8), 885-893.

Fracker, M. L. (1989). Attention gradients in situation awareness. In *Situation awareness in aerospace operations* (AGARD-CP-478, pp. 6.1-6.10). Copenhagen, Denmark: NATO-Advisory Group for Aerospace Research and Development.

Gaba, D. M., & Howard, S. K. (1995). Situation awareness in anesthesiology. *Human Factors, 37*, 20-31.

Garland, D. J., Endsley, M.R., Andre, A. D., Hancock, P. A., Selcon, S. J., & Vidulich, M. A. (1996). Assessment and measurement of situation awareness. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting,* (pp. 1170-1173). Santa Monica, CA: Human Factors and Ergonomics Society.

Garonzik, R., Brockner, J., & Siegel, P. A. (2000). Identifying international assignees at risk for premature departure: The interactive effect of outcome favorability and procedural fairness. *Journal of Applied Psychology, 85*, 13-20.

Gilson, R. A. (1995). Special issue preface. *Human Factors, 37*, 3-4.

Golden, C. J., Sawicki, R. F., & Franzen, M. D. (1990). Test contruction. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (2nd ed.). New York: Pergamon Press.

Graen, G. (1976). Role making processes within complex organizations. In M.D. Dunnette (Ed.), *The handbook of industrial and organizational psychology* (pp. 1201-1245). Chicago: Rand McNally.

Graen, G. & Scandura, T.A. (1987). Toward a psychology of dyadic organizing. In L.L. Cummings & B.M. Staw (Eds.), *Research in organizational behavior* (Vol. 9, pp. 175-208). Greenwich, CT: JAI Press.

Gugerty, L. T., & Tirre, W. C. (1996). Situation awareness: a validation study and investigation of individual differences. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 1170-1173). Santa Monica, CA: Human Factors and Ergonomics Society.

Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Colquitt, J. A., & Hedlund, J. (1998). Extending the multilevel theory of team decision making: Effects of feedback and experience in hierarchical teams. *Academy of Management Journal, 41*(3), 269-282.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*, 292-316.

Hopkin, V. D. (1994). Situational awareness in air traffic control. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 171-197). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Ilgen, D. R., & Hollenbeck, J. R. (1991). The structure of work: Job design and roles. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, (second ed. vol. 2, pp.165-207). Palo Alto, CA: Consulting Psychologists Press.

Jackson, S. E., & Schuler R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Decision Processes, 36*, 16-78.

Jentsch, F., Barnett, J., Bowers, C., Hicks, J., & Sierra, C. (1997). *An evaluation of commercially available simulation software for team situation awareness research.*

(NAWCTSD Report, No. 97-XXX). Orlando, FL: Naval Air Warfare Center Training Systems Division.

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38*, 220-231.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Occupational stress: Studies in role conflict and ambiguity*. New York: Wiley.

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

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (vol 5; pp. 47-92). Greenwich, CT: JAI.

Klimoski, R. & Jones, R. G. (1995). Staffing for effective group decision making: Key issues in matching people and teams. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 291-332). San Francisco, CA: Jossey Bass.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121-1134.

Major, D. A., & Fink, A. A. (1998). *Linking situation awareness to hierarchical team decision making theory.* Poster presented at the Twenty-fourth International Congress of Applied Psychology, San Francisco, CA.

Major, D. A., Fink, A. A., & Stout, R. J. (1998/1999). The development of shared understanding: Exploring team situation awareness through role theory. *Training Research Journal, 3,* 27-55.

McGrath, J. E., Berdahl, J. L, & Arrow, H. (1995). Traits, expectations, culture and clout: The dynamics of diversity in work groups. In S. E. Jackson & M. N. Ruderman (Eds.), *Diversity in work teams* (pp. 17-46). Washington, DC: American Psychological Association.

McIntyre, R.M., & Salas, E. (1995). Measuring and managing for team performance: Lessons from complex environments. In R.A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9-45). San Francisco: Jossey-Bass.

Means, B., Salas, E., Crandall, B., & Jacobs, T. O. (1993). Training decision makers for the real world. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs or words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

Miller, G. A. (1956). The magic number seven, plus or minus two: some limits on our capacity for processing information. *Journal of Experimental Psychology, 90,* 227-234.

Mogford, R. H. (1994). Mental models and situation awareness in air traffic control. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in* *complex systems* (pp. 199-207). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Morgan, B. B., Jr., Glickman, A. S., Woodard, E. A., Blaiwes, A., & Salas, E. (1986). *Measurement of team behaviors in a Navy environment* (NTSC Report, No. 86-014). Orlando, FL: Naval Training Systems Center.

Muniz, E. J., Stout, R. J., Bowers, C. A & Salas, E. (1997). *A Methodology for Measuring Team Situational Awareness.* Paper presented at the 18th Annual Industrial/Organizational Psychology-Organizational Behavior Graduate Student Conference, Roanoke, VA.

Muniz, E. J., Stout, R. J., Bowers, C. A & Salas, E. (1998). *A Methodology for Measuring Team Situational Awareness: Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT).* Orlando, FL: Naval Air Warfare Center Training Systems Division.

NovaLogic, (1995). Werewolf v. Comanche [computer software]. Calabasas, CA: Author.

Pew, R. W. (1994). An introduction to the concept of situation awareness. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 17-23). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Rantanen, E. (1994). The role of dynamic memory in air traffic controllers; situation awareness. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 209-215). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Richardson, G. (1990). *Feedback thought in social science and systems theory.* Philadelphia: University of Pennsylvania Press.

Rizzo, J. R., House, R. J., & Lirtzman, S. I. (1970). Role conflict and ambiguity in complex organizations. *Aministrative Science Quarterly, 15,* 150-163.

Rodgers, M .D., & Duke, D. A. (1994). SATORI: Situation Assessment Through Re-Creation Of Incidents. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 217-225). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology, 1,* 45-57.

Schenk, T. L. (1994). Situation awareness in emergency services incident command. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 291-293). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1-66.

Schrieber, B. T., Bell, H. N., & Raspotnik, W. B. (1998). Investigating communication and situation awareness. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 21-25). Santa Monica, CA: Human Factors and Ergonomics Society.

Schreiber, B. T., Lee, K. M., Raspotnik, W. B., & Hubbard, D. C. (1996). Examining the relationship between communication and situation awareness (SA) in the tactical air-to-air environment. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (p. 1253) Santa Monica, CA: Human Factors and Ergonomics Society.

Schuler, R. S., Aldag, R. J., & Brief, A. P. (1977). Role conflict and ambiguity: A scale analysis. *Organization Behavior and Human Performance, 20*, 111-128.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.

Seers, A. (1989). Team-member exchange quality: A new construct for role-making research. *Organizational Behavior and Human Decision Processes, 43,* 118-135.

Selcon, S. J. & Taylor, R. M. (1989). Evaluation of the Situation Awareness Rating Technique (SART) as a tool for aircrew systems design. In *Situation awareness in aerospace operations* (AGARD-CP-478, pp. 5.1-5.8). Copenhagen, Denmark: NATO-Advisory Group for Aerospace Research and Development.

Shrestha, L. B., Prince, C., Baker, D. P., & Salas, E. (1995). Understanding situation awareness: Concepts, methods and training. *Human/Technology Interaction in Complex Systems, 7*, 45-83.

Simon, H. A. (1960). *The new science of management decision*. Englewood Cliffs, NJ: Prentice-Hall.

SmallWaters Corporation (1999). AMOS 4.0 [computer program]. Chicago: Author.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors, 37*(1), 137-148.

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1996/1997). The role of shared mental models in developing team situational awareness: Implications for training. *Training Research Journal, 2*, 85-116.

Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics.* New York: Harper & Row.

Taylor, R. M. (1989). Situation Awareness Rating Scale (SART): The development of a tool for aircrew systems design. In *Situation awareness in aerospace operations* (AGARD-CP-478, pp. 3.1-3.17). Copenhagen, Denmark: NATO-Advisory Group for Aerospace Research and Development.

Tenney, Y. J., Adams, M. J., Pew, R. W., Huggins, A. W. F., & Rogers, W. H. (1992). *A Principled Approach to the Measurement of Situation Awareness in Commercial Aviation* (NASA Contractor Report 4451; Contract NAS1-18788). Hampton, VA: Langley Research Center.

Tracy, L., & Johnson, T. W. (1981). What do the role conflict and role ambiguity scales measure? *Journal of Applied Psychology, 66*, 464-469.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (third ed., pp. 709-811). New York: Harper Collins.

Van Sell, M., Brief, A. P., & Schuler, R. S. (1981). Role conflict and ambiguity: Integration of the literature and directions for future research. *Human Relations, 34*, 43-71.

Venturio, M., Hamilton, W. L., & Dvorchak, S. R. (1989). Performance-based measures of merit for tactical situation awareness. *In Situation awareness in aerospace operations* (AGARD-CP-478, pp. 4.1-4.5). Copenhagen, Denmark: NATO-Advisory Group for Aerospace Research and Development.

Wayne, S. J., & Liden, R. C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal, 38*, 232-260.

Wellens, A. R. (1993). Group situation awareness and distributed decision making: From military to civilian applications. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 267-291). Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C. D. (1999). Cognitive factors in aviation. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay & M. T. H Chi (Eds.), *Handbook of applied cognition* (pp. 247-282). New York: John Wiley & Sons.

Wright, D. L., Pleasants, F. & Gomez-Meza, M. (1990). Use of advanced visual cue sources in volleyball. *Journal of Sport & Exercise Psychology, 12*, 406-414.

Appendix A

Situation Awareness Rating Scale
(SARS)
Caretta, Perry & Ree, 1994

General Traits
- Discipline
- Decisiveness
- Tactical knowledge
- Time-sharing ability
- Reasoning ability
- Spatial ability
- Flight Management

Tactical Game Plan
- Developing plan
- Executing plan
- Adjusting plan on-the-fly

System Operation
- Radar
- TEWS
- Overall weapons system proficiency

Communication
- Quality (brevity, accuracy, timliness, completeness)
- Ability to effectively use comm info

Information interpretation
- Interpreting VSD
- Interpreting RWR
- Ability to effectively use AWACS/GCI
- Integrating overall information (cockpit displays, wingman comm, controller comm)
- Radar sorting
- Analyzing engagement geometry
- Threat prioritization

Tactical Employment-BVR Weapons
- Targeting decisions
- Fire-point selection

Tactical employment – Visual Maneuvering
- Maintain track of bogeys/friendlies
- Threat evaluation
- Weapons employment

Tactical employment – general

Assessing offensiveness/defensiveness
- Lookout (VSD interpretation, RWR monitoring, visual lookout)
- Defensive reaction (chaff, flares, maneuvering, etc)
- Mutual support

Appendix B

Items from Situation Awareness Rating Technique
(SART)
Taylor, 1989

Think about the game you have just played. Rate the game for:
1. Instability of situation: Likeliness to change suddenly
2. Complexity of situation: degree of complication
3. Variability of situation: number of variable/factors changing

Think about the partner that you have just played this game with. Rate your partner on:
1. Arousal: degree of alertness: readiness for activity
2. Concentration of attention: degree to which thought was brought to bear
3. Division of attention: distribution/spread of focus of attention
4. Spare mental capacity: mental ability available for new information

Appendix C

Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIENT)

Muniz, Stout, Bowers & Salas, 1997

Please rate the extent to which your partner did each of the following things:

1. Monitored environment for changes, trends, abnormal conditions
2. Demonstrated awareness of where he/she was
3. Reported problems
4. Located potential sources of problem
5. Demonstrated knowledge of problem consequences
6. Resolved discrepancies
7. Noted deviations
8. Recognized a need for action
9. Anticipated consequences of action and decisions
10. Informed other of actions taken
11. Monitored actions (self & others)
12. Demonstrated knowledge of tasks
13. Exhibited skilled time sharing attention among tasks
14. Monitored workload (self & others)
15. Shared workload within station
16. Answered questions promptly
17. Communicated important information
18. Confirmed information when possible
19. Challenged information when doubtful
20. Re-checked old information
21. Provided information in advance
22. Obtained information of what is happening
23. Demonstrated understanding of complex relationships
24. Briefed status frequently

Appendix D

Communication Evaluation

Definitions:

Communication Frequency:

Rate the frequency of communication for each team member while performing the simulation. Rate a 10 for constant communication (the team member is talking 100% of the time), a 5 for frequent communication (the team member is talking 50% of the time) and a 0 for no communication (the team member says nothing. Do the same for both team members. Scores need not add up to 10.

On-Task Communication: Mundane

Rate the amount of communication for each player that is on-task but mundane - issues such as speed, altitude, weapon status. Use the percentage rating scaling

On-Task Communication: Emergency

Rate the amount of communication for each player that is on-task and in response to emergencies, such as attacks. Use the percentage rating scaling

Role Related Communication

Communication by the team members that indicate continued role negotiation or refer back to established roles. Use the percentage rating scaling.

Off-Task Communication: Team related (rapport building, support statements etc)

These are communications between team members that are essentially social in nature, but are not role negotiation-type statements. These might be congratulations, or other rapport-building inter-member communications. Use the percentage rating scale.

Closed loop Communication

Communication initiated by the team member which follows the following three step pattern: communication is initiated, receiver accepts the communication and provides feedback sender confirms (closing the loop). Use the percentage rating scale.

Communication Scales:

Communication Frequency

Player A											Player B										
0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10

On-Task Communication: Mundane

Player A											Player B										
0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10

On-Task Communication: Emergency

Player A											Player B										
0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10

Role Related Communication

Player A											Player B										
0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10

Off-Task Communication: Team related (rapport building, support statements etc)

Player A											Player B										
0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10

Closed loop Communication

Player A											Player B										
0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10

Appendix E

Communication Score Sheet

Team No.: _________

Player A

Information Request																								
Situation Fact: **Raw Data**																								
Situation Fact: **Processed**																								
Other Communication																								

Player B

Information Request																								
Situation Fact: **Raw Data**																								
Situation Fact: **Processed**																								
Other Communication																								

Appendix F

Participant Training Materials

Project Fly

Introduction

Thank you for participating in Project Fly.

Project Fly uses a team computer game, a helicopter flight simulator. We will teach you and your partner how to use the flight simulator. It is not necessary that you be experienced at computer games to participate in this project. You will have an opportunity to practice both the skills and the game.

Before we begin, take a moment to answer the pre-game questions. Then the experimenter will complete a brief demonstration mission. Please tell the experimenter when you are ready to see the demonstration mission.

If you have any questions, please ask the experimenter.

Project Fly

The Game

The game you will be playing in Project Fly is a helicopter flight simulator. You and your partner will play together. That is, even though you and your partner are sitting at different computers, you will be teammates in the same scenario. Each of you will play from a separate computer, but the computers are networked together. You are two separate helicopters on the same mission, with the same terrain, same targets and same threats. When your partner destroys a threat, it is destroyed for both of you.

While you are playing the game, the experimenter will pause it occasionally, and ask you to answer a few questions about how the game is going at that point. The game will pick back up where it left off.

Project Fly

Playing the Game -

The experimenter will walk you through the game to help you learn how it is played. The experimenter will also setup the computers to run in 2-player mode and will select the pre-arranged missions at the appropriate times.

It is important to remember that this flight simulator is a helicopter rather than an airplane, and they fly differently - you must control the collective for lift and the cyclic for direction. Fortunately, COMANCHE's computer system automatically handles throttle.

Basically, you will use the joystick to control the cyclic (the direction you fly) and the "+" and "-" keys on the 10-key keypad to control the collective, or lift. Press the higher key (-) to go up, and the lower key (+) to go down. You need to press the higher "-" key to take off at the beginning of the simulation.

You will also need to become familiar with the weapons options available to you. You must use the keyboard to select weapons. The experimenter will assist you through another training mission so that you can become familiar with the weapons systems that you can use.

You have seven weapons options available on COMANCHE: 20 mm cannon, 70mm rockets, laser guided missiles, (hellfire), air-to-air missiles (stingers), artillery, wingman, two 70mm rockets at once. The artillery and wingman options are not available on every mission, and when available only require that you lock onto the target. The coordinates of your target are transmitted to your artillery or wingman support. Your currently selected weapon is displayed on the screen. These weapons options can be selected using the keys of the bottom keyboard row:

Z = 20 mm cannon
X = 70mm rockets
C = laser guided missiles (hellfire)
V = air-to-air missiles(stingers)
B = artillery
N = wingman
M = two 70mm rockets at once (salvo)

First, you and your partner will fly a training mission to become more comfortable with the simulation. Don't worry too much about your performance or strategy just yet - focus on learning how to use the simulation. The experimenter will help you. <Play a mission - training goal is aircraft control. Note that a standardized cockpit will be used for all players>

Appendix G

Descriptive Statistics for Measures

Measure	*N*	Mean	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	258	2.95	3.65																						
2	238	18.86	23.88	.06																					
3	267	1.07	0.90	.09	.00																				
4	271	5.03	1.11	-.10	-.04	.01																			
5	272	4.37	1.19	-.15*	-.04	-.04	-.08																		
6	271	5.16	1.19	-.02	-.11	-.01	.09	.19*																	
7	271	5.41	1.04	.02	-.08	.10	.01	.12*	.19*																
8	272	4.25	1.51	-.07	-.04	.01	-.10	.69*	.16*	.16*															
9	272	4.27	1.73	-.01	-.07	-.08	.00	.32*	.52*	.15*	.37*														
10	272	4.26	1.58	.01	-.01	.03	.04	.27*	.19*	.40*	.31*	.39*													
11	272	3.12	1.67	-.12*	-.09	-.04	-.05	.62*	.19*	.16*	.79*	.51*	.47*												
12	272	2.59	2.78	-.01	.05	-.02	.06	.24*	.02	.04	.33*	.38*	.43*	.52*											
13	268	0.37	0.67	-.13*	.01	-.06	.06	.36*	.06	.10	.43*	.40*	.43*	.68*	.58*										
14	272	4.80	9.45	-.07	.05	-.10	.00	.37*	.09	.11	.39*	.38*	.37*	.54*	.53*	.73*									
15	272	2.86	1.02	.06	.17*	-.05	.03	.14*	.04	.00	.19*	.30*	.18*	.19*	.25*	.15*	.15*								
16	272	2.38	1.19	.01	-.03	.00	-.03	-.02	.12*	-.02	-.13*	.10	-.04	-.07	-.11	-.13*	-.17*	-.17*							
17	271	3.74	0.86	-.06	-.06	.04	-.02	.04	.24*	.07	.05	.10	.00	.02	-.12*	.02	-.05	-.08	.20*						
18	271	3.69	0.92	.06	-.03	.01	.02	.05	.26*	.09	.08	.38*	.24*	.12*	.08	.13*	.02	.19*	.11	.29*					
19	271	2.13	0.87	-.05	-.09	.06	-.12*	.28*	.34*	.20*	.37*	.36*	.25*	.38*	.09	.18*	.16*	.09	.11	.16*	.13*				
20	271	3.58	1.04	-.03	-.07	-.07	.03	.06	.25*	-.04	.08	.26*	.10	.17*	.19*	.23*	.19*	.06	.06	.01	.05	.16*			
21	270	41.19	26.81	-.09	-.17*	-.11	-.18*	.52*	.32*	.18*	.59*	.30*	.14*	.48*	.05	.18*	.24*	.14*	-.04	.12*	.14*	.46*	.11*		
22	271	3.18	1.01	-.05	-.06	.00	-.15*	.30*	.49*	.27*	.38*	.50*	.22*	.35*	.15*	.17*	.20*	.20*	.07	.30*	.33*	.49*	.22*	.50*	
23	271	3.93	0.74	.06	-.17*	.04	.01	.21*	.20*	.47*	.19*	.20*	.36*	.20*	.13*	.14*	.14*	.19*	-.03	.08	.00	.13*	.14*	.23*	.23*

Note. *p < .05. Measures are listed below.

1. Proximal SA: Factor 1
2. Proximal SA: Factor 2
3. Proximal SA: Factor 3
4. SART: Game: Self Rating
5. SART: Game: Experimenter Rating
6. SART: Individual: Self Rating
7. SART: Individual: Partner Rating
8. SART: Individual: Experimenter Rating
9. SALIENT: Self Rating
10. SALIENT: Partner Rating
11. SALIENT: Experimenter Rating
12. Communication: Factor 1
13. Communication: Clarifying
14. Communication: Processed
15. Role Ambiguity
16. Role Conflict
17. Rizzo et al.: Role Ambiguity
18. Rizzo et al.: Role Conflict
19. Task Experience
20. Social Experience
21. Objective Performance
22. Subjective Performance: Self
23. Subjective Performance: Partner

Appendix H

Permission Letters

Consulting Psychologists Press, Inc.

Alex Fink
1300 Melrose Pkwy
Norfolk . VA 23508

PERMISSION AGREEMENT TO INCLUDE SAMPLE ITEMS IN A RESEARCH PUBLICATION

Agreement Issued: March 28, 2000
Customer Number.
Permission Code: 7456DL
Permission Number. 12561

In response to your request of **March 28, 2000, upon concurrent receipt by Consulting Psychologists Press, Inc., of this signed** Permission **Agreement and payment of the Permission Fee, permission is hereby granted** to you to **include a copy from the Handbook Of Industrial & Organizational Psychology (HIOP) from page** 175 **in your Dissertation entitled** "*Developing A Model Of Team Situation Awareness*". **The copy may remain in your Dissertation for microfilming and individual copies may be distributed upon demand.** This Permission Agreement **shall automatically terminate upon violation of this** permission Agreement **including, but not limited to, failure to pay the Permission Fee of $45.00 + reproduction fee of $30.00 = $75.00** or by **failure to sign and return this Permission Agreement within 45 days fr**[redacted]

The permission granted hereunder is limited to this one-time use only.
The permission granted hereunder is specifically limited as specified in this agreement.

This Permission Agreement **shall be subject to the following conditions.**

(a) Any material reproduced **must be used in accordance with the guidelines of the American Psychological** Association.

(b) Any **material reproduced must contain the following** credit lines:

"Reproduced by special **permission** of the Publisher, Consulting Psychologists Press, Inc., Palo Alto, CA 94303 from **Handbook Of Industrial & Organizational Psychology** by Daniel R. Ilgen and John R. Hollenbeck. Copyright 1991 by Consulting Psychologists Press, Inc. All rights reserved. Further reproduction is prohibited without the Publisher's written consent."

(c) None of the materials may be sold or used for purposes other than those mentioned above, including, but not limited to, any commercial or for-profit use. Commercial and/or for-profit use of the **(HIOP)** and/or any **modification of the (HIOP) is specifically excluded from the permission granted herein**

(d) CPP subscribes to the general principles of test use as set forth in the *Standards for Educational and Psychological Testing* Copyright 1985 by the American Psychological Association. The customer's/user's attention is drawn to the following statements.

"The test user, in selecting or interpreting a test, should know the purposes of the testing and the probable consequences. The user should know the procedures necessary to facilitate effectiveness and to reduce bias in test use. Although the test developer and publisher should provide information on the strengths and weaknesses of the test, the ultimate responsibility for appropriate test use lies with the test user. The user should become knowledgeable about the test and its appropriate use and also communicate this information, as appropriate, to others

6.1 Test users should evaluate the available written documentation on the validity and reliability of tests for the specific use intended

3803 E. Bayshore Road P.O. Box 10096 Palo Alto California 94303 Tel (650) 969-8901 Fax (650) 969-8608

MAY 10 '00 08:33AM CAREER DEVELOPMENT
May 08 00 12:57p Marketing & Licensing 650-623-9273 P.3/4 p.3

6.3 When a test is to be used for a purpose for which it has not been validated, or for which there is no supported claim for validity, the user is responsible for providing evidence of validity

6.5 Test users should be alert to probable unintended consequences of test use and should attempt to avoid actions that have unintended negative consequences."

CPP shall not be responsible for the use or misuse of the materials or services licensed under this permission contract. The customer/user assumes all responsibility for use or misuse of the same. Unless expressly agreed to in writing by CPP, all materials and services are licensed without warranty, express or implied, including the implied warranties of merchantability and fitness for a particular purpose. Refund of contract fees at CPP's sole option is the sole and exclusive remedy and is in lieu of actual, consequential, or incidental damages for use or misuse of CPP materials and services and in no event shall CPP liability exceed the contract fees of license of said materials and services.

(e) Alex Fink agrees that the (HIOP) as modified under this Agreement is a derivative work of the (HIOP) and hereby assigns all right, title, and interest in any such derivative work created under this Permission Agreement in perpetuity to Consulting Psychologists Press, Inc (CPP) or as directed by CPP, immediately upon completion and without further consideration.

CONSULTING PSYCHOLOGISTS PRESS, INC

By ______________________
Authorized Representative

Date 5/10/00

I AGREE TO THE ABOVE CONDITIONS

By ______________________
Alex Fink

Date 9 May, 2000

To: Karen Thomas, APA Permissions Office, (202) 336 5549
From: Alexis Fink
Date: February 11, 2000
Re: Permission to reproduce a figure

Dear Ms. Thomas:

Per your request, here is my official request for permission to reproduce an APA copyrighted figure. The figure I wish to reproduce is Figure 4, Page 26 from:

McGrath, J. E., Berdahl, J. L, & Arrow, H (1995). Traits, expectations, culture and clout: The dynamics of diversity in work groups. In S. E. Jackson & M. N. Ruderman (Eds.), Diversity in work teams (pp. 17-46). Washington, DC: American Psychological Association.

I discuss McGrath et al.'s work in the theory development section of my dissertation, which I am scheduled to defend in May, 2000. Figure 4 of the above-referenced chapter is a concise visual representation of the theory that McGrath et al. describe. For this reason, I would like to include a reproduction of the figure in my dissertation. I have spoken with Dr. McGrath, and he has given me permission to reproduce the figure. I have also made his co-authors aware of my conversation with Dr. McGrath.

If you need further information from me, I can be reached via e-mail at: alexisfink@erols.com. Additionally, I can be reached via telephone during business hours at (973) 426 2170. Street mail can be addressed to me at 1333 Chetwynd Ave, Plainfield, NJ 07060.

I appreciate your time and attention to this request. I look forward to your favorable reply.

Sincerely,

Alexis A. Fink

...granted without fee for non-exclusive, one time ... material for the purposes stated in ... editions ... from ... permission. ... material ... citation and the following notice: ... 1995 ... American Psychological Association. ... (or adapted) by permission.

Karen A. Thomas 3-20-00
Permissions Office
American Psychological Association

FEB 11 '00 12:13PM CAREER DEVELOPMENT P.1/1

FEB 23 '00 10:00AM CAREER DEVELOPMENT P.1/1

TO: Karen Thomas, APA Permissions Office
FROM: Alexis Fink, ph 973 426 2170, fax 973 426 6832
DATE: Feb 23, 2000
RE: Permission to reproduce a figure

Ms. Thomas:

I request permission to reproduce, in my dissertation, figure 4, page 299, from the following article that appeared in an APA journal:

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J, Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. Journal of Applied Psychology, 80, 292-316.

I have author permission to reproduce the figure (one of the authors is my dissertation committee chair, and it is at her request that the figure appear).

Thank you for your time and attention to this request. If you require further information from me, please do not hesitate to contact me. My telephone number is 973 426 2170, my fax is 973 426 6832, and my e-mail is alexisfink@erols.com.

Regards,

Alexis A. Fink

[illegible] granted without fee for non-exclusive, one time [illegible] APA-copyrighted material for the purposes stated in [illegible] request. [illegible] does not apply to future editions [illegible] format different from [illegible] requested permission. Reprinted material [illegible] graphic citation and the following notice: Copyright 1995 by the American Psychological Association. Reprinted (or Adapted) by permission.

Karen A Thomas 3-31-00
Permissions Office
American Psychological Association

VITA

ALEXIS ANNE FINK

Alexis received her B.A. in psychology from DePauw University in December of 1993, graduating magna cum laude and Phi Beta Kappa. She received her M.S. in psychology from Old Dominion University in May of 1996, and her Ph.D. from Old Dominion University in 2000. She has had a wide sampling of professional experiences which includes work in overseas, an internship in the Office of the Vice President of the United States of America, and work in the electronics, financial services and chemicals industries, in addition to more traditional academic activities of research and teaching.

The address for the Department of Psychology at Old Dominion University is Mills Godwin Building, Room 250, Norfolk, Virginia 23529.