Johnston, J., **Fiore, S.M.,** Paris, C., & Smith, C. A. P. (in press). Application of Cognitive Load Theory to Developing a Measure of Team Decision Efficiency. *Military Psychology*.

Application of Cognitive Load Theory to Developing a Measure of Team Decision Efficiency

Joan H. Johnston

Naval Air Warfare Center Training Systems Division

Stephen M. Fiore

University of Central Florida

Carol Paris

Naval Air Warfare Center Training Systems Division

and

C.A.P. Smith

Colorado State University

Author Notes:

This project was funded by the Office of Naval Research. The views, opinions, and findings contained in this article are the authors and should not be construed as official or as reflecting the views of the Department of Defense. This paper is intended to be approved for public release and unlimited distribution. Requests for reprints should be sent to Joan Johnston, NAVAIR Orlando Training Systems Division, AIR 4651, 12350 Research Parkway, Orlando, FL, 32826. E-mail: joan.johnston@navy.mil.

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **2002**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Application of Cognitive Load Theory to Developing a Measure of Team Decision Efficiency**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Air Warfare Center Training Systems Division**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**31**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

ABSTRACT

Improving human systems integration through technologically advanced training and performance aids has become increasingly important to military transformation. Measures of improved cognitive and coordination processes arising from the employment of transformational tools are necessary to guide the refinement and future development of such technologies. In this paper we describe a Cognitive Load Theory approach to developing a combinatory measure of individual workload and team performance following an experimental intervention involving training and a Decision Support System. We discuss how indicators of what we term *team decision efficiency* can improve assessing the effectiveness of transformational processes and technologies.

Application of Cognitive Load Theory to Developing a Measure of Team Decision Efficiency

Improving human performance through advanced training and decision aids is a major objective of military transformation advocates. However, advances are needed in diagnostic measures of cognitive and team coordination processes to better guide the design and development of efficient transformational technologies. The Tactical Decision Making Under Stress (TADMUS) program, sponsored by the Office of Naval Research, successfully demonstrated that effective team training and aiding through a Decision Support System, based on cognitive and team task analyses, resulted in better performance, and with less individual mental effort exerted (for a discussion of this related research, see Cannon-Bowers & Salas, 1998). The final TADMUS experiment tested the combined effect of training and decision support, and a recent analysis showed that decision-making was improved through these interventions (Smith, Johnston, & Paris, 2003). In this paper we build upon this body of research so as to advance diagnostic measures for assessing human systems integration efficiencies (cf. Fiore, Cuevas, Scielzo, & Salas, 2002; Scielzo, Fiore, Cuevas, & Salas, 2004). Towards that goal, we describe and test a measure named the Team Decision Efficiency (TDE) score derived from Cognitive Load Theory and explored within TADMUS experimentation.

The Team Decision Efficiency measure is part of a theoretical framework developed in the area of team cognition (Salas & Fiore, 2004) to understand process and performance at the inter- and intra-individual level (for a full discussion see Fiore, Johnston, Paris, & Smith, in press). The overarching goal of the framework put forth by Fiore et al. is to aid in theory development by hypothesizing innovative strategies to assess human systems integration. This framework describes how measures of team performance can be simultaneously used in combinatory analyses with subjectively derived measures at the individual level to examine the impact of technology-based aids on team process and performance. By assessing subjective processes in ways analogous to those put forth in the instructional sciences, Fiore et al. (in press) argued that we can have a window into the manner in which processes at the level of the individual interact with, and alter, processes at the team level. In this paper we focus on the component of that framework derived from Cognitive Load Theory (CLT) and test a measure derived from that approach – the Team Decision Efficiency score. Although CLT has been used for a number of years in instructional systems design research, its application to team decision making represents a unique contribution to human-systems integration in general and team cognition studies in particular.

Cognitive Load Theory

Cognitive Load Theory (CLT) has been the focus of the educational and instructional sciences for over a decade (Chandler & Sweller, 1991; Sweller & Chandler, 1994; Sweller, Chandler, Tierner, & Cooper, 1990) and its utility continues to grow (see Paas, Renkl, & Sweller, 2003; Paas, Tuovinen, Tabbers, & Van Gerven, 2003). CLT articulates how cognitive processes in working memory interact with long term memory and learning content and performance. Sweller (1994) defines learning as schema acquisition and the transfer of learned procedures as one moves from controlled to automatic information processing. As knowledge is acquired it decreases the burden on working memory (see also Chandler & Sweller, 1996; Sweller, 1988; Van Merrienboer, & Paas, 1990). CLT posits that, depending upon the amount of knowledge already acquired within a given domain, learning and performance can be altered due to the load imposed by external factors (for a full discussion of CLT, see Sweller, 1999; Sweller & Chandler, 1994). Specifically, endogenous and exogenous factors are present when one interacts with an instructional environment. Endogenous factors are the long-term memory structures associated with a particular learning content and the working memory processes (see Baddeley, 1986; 1992a; 1992b) used when engaged in a learning activity (e.g., Chandler & Sweller, 1991).

Exogenous factors such as instructional system design or training content interact with these endogenous limitations in cognition. For example, without substantial, part task simulation-based training; the exogenous factors in learning military tasks (e.g., flying a high speed aircraft while operating command and control displays and coordinating with crewmembers and other external aircraft) are of sufficient quantity to overwhelm human information processing capacities. Further, the exogenous and endogenous factors can require either a high or low degree of interaction themselves. For instance, in instructional contexts, CLT characterizes the forms of cognitive load that result as intrinsic and extrinsic. Intrinsic load is high when learning content requires a substantial degree of interaction and involves a large number of cognitive elements. Moreover, when information is new for the learner (i.e., the long-term memory associated with the content is sparse), the intrinsic load is challenging. Extrinsic load is described as an additional (artificial) cognitive load imposed by poorly designed instruction and it is argued to hinder learning (e.g., Kalyuga, Chandler, & Sweller, 1999).

CLT advances the notion that analysis of instructional efficiency, which identifies the cognitive burden on the trainee in conjunction with performance, may increase the return on investment in developing training systems. Paas, Van Merrienboer, & Adam (1994) defined their instructional efficiency construct as the relationship between trainee subjective workload assessments and overall task performance. The “instructional efficiency score” is calculated using standardized scores of subjective assessments of mental effort and performance (Paas and Van Meerienboer, 1993). As one interacts with the learning environment, the burden on working memory should be subjectively assessed and simultaneously considered with learner performance because it reveals “important information about the cognitive consequences of instructional conditions that is not necessarily reflected by traditional performance-based measures” (Paas & Tuovinen, 2004, p. 134).

For example, within multimedia learning environments, cognitive resources are more efficiently used when animations are presented with a voiceover than with on-screen text (e.g., Kalyuga et al., 1999; Mayer, & Moreno, 1998; Mousavi, Low, & Sweller, 1995). The simultaneous presentation of animation and text (referred to as the principle of redundancy in theories of multimedia, see Mayer, 2001; Moreno & Mayer, 1999) may produce higher cognitive load due to overburdening the visuospatial-sketchpad in working memory. In contrast, tapping separate auditory and visual channels achieves greater instructional efficiency because the working memory burden is reduced -- referred to as the principle of temporal contiguity (see Mayer, 2001). Poorly designed instructional systems that violate the principle of temporal contiguity may produce low instructional efficiency scores due to increases in workload concomitant with decreases in performance.

This brief review of CLT was presented within the context of the Navy’s TADMUS effort because the conceptual underpinnings of that line of inquiry into training and decision support were based on analogous theories of cognition. In particular, the goal of the system improvements was to make decisions and team interaction requirements clearer and more transparent (cf. Marcus, Cooper, & Sweller, 1996). In the present study, using a variant of the

Paas et al. (1994) instructional efficiency scores, we developed the Team Decision Efficiency score in order to determine if it was possible to add a level of diagnosticity to efforts in human-systems integration. This score is based upon subjective assessments of workload combined with objective measures of team tactical performance. The specific derivation of the Team Decision Efficiency is described in detail in the Methods section, but, generally, it is derived in a manner similar to instructional efficiency. The primary difference is that we use team performance rather than individual performance in the formula. Thus, we label this team decision efficiency because it is a composite score derived from workload scores of individuals interacting within a team and the associated team performance scores. Following the theoretical framework put forth by Fiore et al. (in press), the overall hypothesis was that teams provided training and a Decision Support System (referred to hereafter as Training/DSS) would perform more efficiently than a control condition. As such, it is expected that compared to the control group, teams receiving Training/DSS would use more effective teamwork processes (information exchange, supporting behavior, initiative/ leadership, and communication), and this would lead to Team Decision Efficiency scores favoring the Training/DSS condition.

METHODS

Participants

Participants were 96 US Navy officers enrolled in an officer training program. Participants were primarily males (Male = 93, Female = 3), and participant rank was Lieutenant (0-3) with mean years of service at 9.6 years (SD = 3.8). Participants had served at least two tours on a ship and, in at least one of the tours, each had experience as a Division Head which is the equivalent of a first-level manager. Participants did not receive additional payments or course credits for their efforts. One participant did not complete the NASA-TLX inventory and was excluded from the analysis.

Design and Task

The research protocol described next is based on previous TADMUS team research (refer to Johnston, Poirier, & Smith-Jentsch, 1998 for details). The study was a quasi-experimental, between groups, post-test only design with two conditions (Training/DSS vs. Control) described in greater detail below. Each participant was assigned to a six-person team, with eight teams in each condition. Random assignment to condition was not possible, but efforts were made to ensure team composition was not biased based on a particular specialty (e.g., engineer versus combat systems officer). To act as a ship's air defense warfare team, participants were assigned to one of the following roles: Commanding Officer (CO), Tactical Action Officer (TAO), Air Defense Warfare Coordinator (ADWC), Tactical Information Coordinator (TIC), Identification Supervisor (IDS), and Electronic Warfare Supervisor (EWS). The reporting relationship among the team members was hierarchical, with the IDS reporting to the TIC, and the TIC reporting to the ADWC. The ADWC and EWS report directly to the TAO, and the TAO reports directly to the CO.

Teams performed their tactical decision making tasks on PC-Based watch-stations linked together through a local area network to form a distributed simulation training system named the Decision Making Evaluation Facility for Tactical Teams (DEFTT) (Johnston, Poirier, & Smith-Jentsch, 1998). Event-based simulator scenarios were time-tagged to identify specific expected team behaviors throughout. All information was unclassified. Headsets supported verbal communications among team members and role players, and role players read from a script in order to prevent any deviations from expected events. All participants had at least 48 hours of DEFTT experience prior to the experiment. Participation in the experiment was incorporated into the participant's training schedule.

The team task objective was to perform a ship's air defense warfare detect-to-engage sequence. Team members had to interact with their watch-stations and passed tactical information to each other to develop an accurate picture about potentially hostile and friendly aircraft and ships "radar tracks" as they appeared throughout each of four 30-minute scenarios. Teams had to report initial detection of a surface or air track, track type (commercial or tactical), and priorities for dealing with the most threatening contact. Although the simulated tracks did not react to watch-stander actions (i.e., they were not intelligent agents), as team members changed identification of specific tracks on the radar displays, that information changed across all watch-stander displays. If a threatening track met specific rules of engagement, the team had to report plans to obtain authority to prepare for the ship's self defense. When approved, the team had to execute actions based on their pre-planned responses in accordance with rules of engagement.

Control Condition. Teams in the control condition performed their watch-station tasks on the DEFTT system. The TAO and CO shared a single Command and Decision Display simulation watch-station configured specifically for them. The TIC, IDS, AAWC, and EWS each had a Command and Decision simulation watch-station. In addition, the EWS had an early warning system simulated watch-station. The research protocol for this condition was based on the typical combat training the officers received during their course curriculum.

Training/DSS Condition. Team members were assigned DEFTT watch-stations with the exception that the TAO and CO were each assigned a DSS (see Morrison, Kelly, Moore, &

Hutchins, 1998 for details). The DSS operates in a standalone mode, but was synchronized to run in tandem with DEFTT for this experiment. The TAO and CO received a 45-minute computer-based DSS tutorial that described display functions and allowed point and click practice. The DSS was designed based on the cognitive tasks underlying TAO and CO decision making processes, and then a set of supporting command and control displays were developed and tested (Morrison et al., 1998). The resulting display design on the PC monitor is organized into four general areas (refer to Smith, Johnston, and Paris 2004 for details). The upper left side shows the tactical radar symbols with enhanced shading to delineate areas of weapons engagement for potentially hostile tracks. The upper right side of the display (Track Summary, Track Profile, Response Manager) is oriented to present critical information about a single track (e.g., aircraft, ship, etc.) as efficiently as possible. The lower right side of the screen (Basis for Assessment and Comparison to Norms) presents historical track information in terms of its classification as friendly, neutral, or threat. Running from the lower left to the lower right of the display are the Track Priority and Alerts List that present a prioritized summary information related to the most critical tracks.

Computer-based training and videotape presentation were used to teach Decision Making and Teamwork Skills. The Decision Making Skills computer-based training (McCarthy, Johnston, & Paris, 1998) was adapted from critical thinking research (Cohen, Freeman, & Thompson, 1998), and other research on naturalistic decision making and training (Zsambok & Klein, 1997). It instructs participants to understand and develop decision-making strategies that they transfer to the scenario-based, team training environment. Next, participants were instructed by computer-based training and videotape on Teamwork Skills using Team Dimensional Training, and then practiced identifying specific combat information center (CIC) teamwork behaviors together in the classroom. Team Dimensional Training was developed and validated under previous TADMUS research and later refined under research for shipboard instructor training and support (Smith-Jentsch, Zeisig, Acton, & McPherson, 1998). Next, participants assembled in the DEFTT lab and an instructor trained them on how to conduct structured after action reviews using the Team Dimensional Training Debriefing Guide. Participants practiced Team Dimensional Training in the context of two DEFTT training scenarios. The participants were instructed on, and practiced using the DSS as a replay device following practice on the training scenarios to highlight critical events and support Team Dimensional Training discussions.

Dependent Measures

Air Defense Warfare Team Observation Measure. The Air Defense Warfare Team Observation Measure (ATOM) provides scores on four dimensions of teamwork behaviors: Supporting behavior, Leadership/Initiative, Information Exchange, and Communications (Johnston, Smith-Jentsch, & Cannon-Bowers, 1997). A trained rater, blind to conditions, used team communications transcripts and videotapes to assess team performance on 11 items. Each item is a five-point scale with anchors at each end. A rater assesses the extent to which a specific team behavior represented a “real weakness or strength for the team.” An acceptable internal consistency reliability (alpha) estimate of .79 was found.

Air Defense Warfare Team Performance Index. The Air Defense Warfare Team Performance Index (ATPI) is a paper-based measure of team task performance on the “detect-to-engage” (DTE) sequence (Johnston et al., 1997; Paris, Johnston, & Reeves, 2000). Subject Matter Experts (SMEs) established standards of DTE performance (timing and accuracy) for the most critical aircraft in each of the four post-test scenarios. Two trained raters, blind to conditions, used team communications transcripts to judge whether or not, and when, team members reported correct and incorrect DTE actions. Rater agreement ranged between 91 and 100 percent, with an average agreement of 97 percent. A third rater corrected the minor disagreements so that a single ATPI would exist for team task performance on each scenario.

Detection (DE) and Planning/Execution (PE) scores were developed as ATPI subscores to support diagnosis of team task performance based on the team decision making schema model by Paris et al. (2000). For the Detection (DE) subscore, teams were evaluated on their accuracy and timing in reporting initial detection of aircraft, aircraft type (commercial or tactical), and priorities for dealing with the most threatening aircraft. An On-time DE score is based upon the team's timely and accurate responses to all tactical aircraft, across the four scenarios. A Late DE score is based upon the team's accurate, but late responses to all tactical aircraft, across the four scenarios.

Planning and Execution actions represent the activities performed by the team after the DE sequence (e.g., warning, challenging, and covering the hostile aircraft with weapons). An On-time PE score is based upon the team's timely and accurate planned and executed actions for all tactical aircraft, across the four scenarios. A Late PE score is based upon the team's accurate, but late, planned and executed actions for all tactical aircraft, across the four scenarios.

Perceived Workload. In CLT, mental effort is typically assessed with a Likert scale and asks participants to rate perceived level of mental effort. For example, nine-point scales have anchors ranging from "very, very low mental effort" to "extremely high mental effort" (see Paas, 1992), and six-point scales have anchors ranging from "very easy" to "difficult" (see Marcus et al., 1996). Along analogous lines level of mental effort was important to diagnosing the effectiveness of the TADMUS Training/DSS intervention. In this experiment, a five-item Likert scale version of the NASA Task Load Index (TLX) asked participants to rate extent of perceived mental demand, physical demand, temporal demand, effort, and frustration[1] on scales labeled at each end with the anchors "low" and "high" (Hart & Staveland, 1988). An acceptable internal consistency reliability (alpha) estimate of .95 was found.

Team Decision Efficiency Score. The Team Decision Efficiency Score was calculated using an individual level metric – that is, individual levels of workload (NASA-TLX), combined with the overall team ATPI scores. Specifically, a given team had six separate workload scores, but one overall performance score was used within a team. This method was pursued for both practical and theoretical reasons. First, the ATPI is designed to capture team performance but there is no equivalent measure for team workload. Second, an argument can be made that this method allows a more precise form of diagnosis. In particular, because workload is an internal state, attempting to observe workload based on behaviors is much more problematic than doing so with team performance. Thus, from the standpoint of team cognition (e.g., Salas & Fiore, 2004), combining perceived individual workload with team performance allows us to capture how team processes may be related to individual processes.

Following CLT, the Team Decision Efficiency scores were derived by taking an individual team member's standardized TLX score and combining them with their team's respective standardized performance scores. Specifically, because "there is no direct method for mapping units of performance on units of mental effort, the measures are converted to standardized z-scores" (Paas & Tuovinen, 2004, p. 142). Kalyuga et al. (1999) utilized this

---

[1] Because one of the items used in the NASA-TLX is conceptually distinct from measures of workload traditionally used in CLT (Item 5 assessing *prediction* of performance), it was not included in the overall sums.

approach and adapted it to show how such scores can be represented as the perpendicular distance from a line representing a level of zero efficiency with the formula as $E = \frac{z_{perf} - z_{wrkl}}{\sqrt{2}}$. As described by Paas et al. (2003), the square root of two is used based upon the formula for calculating the distance of a point to a line (see also Kalyuga et al., 1999 for a full description of the formula's derivation). Because these are standardized scores this results in positive and negative values that hover around a mean of zero. Positive scores indicate relatively better performance in proportion to reported workload whereas negative scores indicate the opposite pattern (relative performance was less than relative workload).

We used the Kalyuga et al. formula for our analyses of the *Team Decision Efficiency* (TDE) scores and this was calculated by analyzing the data across the differing scenarios. Our interest was in viewing TDE across control and experimental groups, but dependent upon whom within a team was working with the DSS. To address the issue that only two of the six team members were actually utilizing the DSS, we created a variable within the teams so as to maintain the distinction between those with the DSS and those without it. Specifically, we divided each team into two sub-teams based upon their roles, that is, those roles within the team using the DSS in the Experimental Condition versus those roles not using the DSS. The variable we label "Role" has two factors, "Command" (the TAO and CO) versus "Support" (the EWS, IDS, TIC, and AAWC). The *Command* role in the Control condition had the same responsibilities as the Command role in the Experimental Condition, the difference was that they did not have the DSS to aid them. We reduced the data in this way so as to examine Team Decision Efficiency Scores emerging *within the teams* but based upon the more global "roles." This analysis allowed us to assess whether the DSS differentially impacted the roles within the teams, *as well as* the other team members. Note that, with this method, instead of 16 *teams* for analysis, we have a sample of 32 because the Command versus Support global roles represented an additional between participant factor. With respect to the formula's full derivation, because we were analyzing the data across the scenarios, we standardized the NASA-TLX values over all participants and scenarios (6 participants in 16 total teams over 4 scenarios)[2]. For the team performance scores, we standardized the relevant ATPI scores over all teams and scenarios (16 total teams over 4 scenarios). The derived *z*-scores were then used to calculate the TDE score based upon the formula described above. The mean TDE scores within the aforementioned team roles were then calculated and used for the subsequent analyses.

Procedure

Control Condition: Participants assembled and filled out informed consent forms. Information packets were provided that developed a context and rationale for the research, and then participants completed a questionnaire about their work experience. Based on these responses, members with the most ship CIC expertise were assigned as TAO and CO, and the remaining team members were assigned to the remaining watch-stations. Next, team members were trained on their respective DEFTT watch stations. First, a training administrator gave an introduction to CIC watch station responsibilities and functions, and then team members practiced operating the watch-stations. Next, team members participated in two 20-minute training scenarios to complete their familiarization with system functions, operations, and team interactions. Next, teams performed on each of the four 30-minute Arabian Gulf scenarios. Scenario order was counterbalanced. Prior to each scenario run team members conducted a quick pre-brief to familiarize themselves on important scenario background information (e.g., geopolitical situation, communications plan, identification matrix, and rules of engagement). At

---

[2] For two of the participants one TLX data point (of his/her four possible) was missing.

the end of each scenario session team members filled out the NASA TLX. Then, they used a Scenario Event Summary Sheet to guide their after action review of team performance. Following experiment completion participants were provided feedback on performance as a way to ensure they received training value for their efforts.

Training and DSS. The experimental condition involved participation over two days. The first day participants filled out informed consent forms, and then participated in the two and one half hour Decision Skills computer-based training. The second day team members completed the demographics questionnaire and, as in the control condition, were assigned to watch-stations based on experience. Next, the CO and TAO were trained in the use of the DSS while the other team members received DEFTT familiarization training. All team members then received the Team Dimensional Training computer-based training and videotape, practiced Team Dimensional Training in the DEFTT with two training scenarios, and employed the DSS during their after action review. At the end of training, teams were reminded they should use a Scenario Event Summary Sheet, DSS, and Team Dimensional Training Debriefing Guide to conduct their after action reviews. Following training the same protocol was used as in the control condition.

RESULTS

We conducted a preliminary analysis of the team process behaviors in order to document Team Dimensional Training intervention effectiveness. Our primary analysis concerned the simultaneous assessment of performance and workload using the Team Decision Efficiency scores. Thus, the preliminary analysis validated whether Team Dimensional Training was successful in supporting the learning and implementation of team process behaviors, and the primary analysis revealed Team Decision Efficiency through comparison of the Training/DSS and Controls conditions.

Team Process Behaviors. A 2-way between-subjects MANOVA was performed on the four dependent teamwork behavior variables (Supporting Behavior, Leadership/Initiative, Information Exchange, and Communications) with one independent variable (Training/DSS versus Control). Supporting our first hypothesis, results showed a significant effect of training on team performance behaviors, F (4, 11) = 4.74, $p < .02$. Associated univariate tests for the training factor revealed a significant main effect on Information Exchange, F (1, 14) = 15.77, $p < .01$, Communications, F (1, 14) = 10.43, $p < .01$, Leadership/Initiative, F (1, 14) = 6.31, $p < .03$, and marginally significant for Supporting Behaviors, F (1, 14) = 3.4, $p < .09$. Figure 1 illustrates the differing scores for these team process behaviors across condition.

Insert Figure 1 Here

Team Decision Efficiency. A 2 x 2 x (2 x 2) mixed-model, repeated measures ANOVA was run on the Team Decision Efficiency Scores with Condition (Training/DSS versus Control) and Role (Command/Support) as the between participant factors, and Decision Task (DE versus PE) and Timing (On Time versus Late) as the within participant factors. Estimated marginal means are reported below.

First, we find a significant interaction between Condition and Role F (1, 28) = 4.29, $p < .05$. Figure 2 shows the standardized Decision Efficiency Scores for the interaction between condition and role illustrating the larger difference between conditions for those in the "Command" role. Specifically, for those team members in the Command role, the Training/DSS condition produced positive efficiency (M = .199) scores while those in the Control condition within the “Command” role produced negative scores (M = -.315). In the Support role, the difference between the Training/DSS (M = -.049) and Control conditions was much less (M = .081). What this interaction suggests is that the DSS had an impact on workload/performance, but only for those roles utilizing the DSS. We next look at our within participant factors to examine whether the Team Decision Efficiency score varied dependent upon the nature of the decision task and the timing of those decisions.

Insert Figure 2 Here

For our second effect we find a significant interaction between Condition and Decision Task F (1, 28) = 4.73, p < .05. Figure 3 shows the standardized Decision Efficiency Scores for this interaction. Overall, the Training/DSS condition produced positive efficiency scores while the Control condition produced negative scores. But, on the PE scores the difference was greater between the Training/DSS (M = .115) and the Control (M = -.158) conditions. The difference for DE scores was substantially less between the Training/DSS (M = .035) and the Control (M = -.076) conditions. Thus, collapsed across on-time and late scores, while the Training/DSS had a small impact on the DE scores, there was a large difference across the PE scores, with the Control group showing a negative score and the Experimental group showing a positive score. This suggests that the teams with the Training/DSS were performing better on the PE decision processes, but this did not come at a cost of higher workload (i.e., they performed better while reporting relatively lower workload).

Insert Figure 3 Here

There was not a significant interaction between Condition and Timing F (1, 28) = 2.43, p < .15. Although this interaction was not significant, we see that, for the on-time scores, both the Training/DSS (M = .012) and the Control (M = -.054) conditions are near zero indicating relatively equal workload and performance. Figure 4 shows the standardized Decision Efficiency Scores between condition and timing of decision. The difference in the late scores was substantially larger between the Training/DSS (M = .138) and the Control (M = -.180) conditions. Thus, collapsed across DE and PE scores, while the Training/DSS condition had little impact for the on-time scores, there was a large difference across the late scores, with the control group showing a negative score, and the experimental group showing a positive score. Post-hoc analysis showed that this difference was significant, t(30) = 1.9, p < .05, one-tailed. This suggests that the Training/DSS teams were basically performing more deliberately (i.e., taking more time) but better, and this deliberation did not come at a cost of higher workload (i.e., they performed better while reporting relatively lower workload).

Insert Figure 4 Here

Last, we find a significant 3-way interaction between Condition, Timing, and Decision Task F (1, 28) = 4.16, p = .05. Figure 5 shows the standardized Decision Efficiency Scores for this interaction. Across the majority of the decisions, we see slightly positive or negative scores indicating relatively equal levels of workload and performance. Consistently, across these scores, we see the control group showing negative scores and the experimental group showing positive scores. Further, the largest difference across conditions was for the PE *late* scores, with these scores in favor of the Training/DSS. Thus, mirroring the prior interactions, this shows that the largest impact for the Training/DSS teams occurred in the late scores, but in this case, primarily for the PE decision processes.

Insert Figure 5 Here

DISCUSSION

In this paper the principle of instructional efficiency was expanded to encompass analyzing how training and DSS influence process and performance for tactical teams – what we termed the Team Decision Efficiency score. Following Fiore et al. (in press) we tested a portion of a framework developed to devise new strategies for assessing human systems integration. The goal of this line of inquiry is to demonstrate how theoretically sound constructs and measurement techniques from domains outside the military sciences can aid in our diagnoses of team processes when technology is designed as a performance aid. Overall, we found that the Team Decision Efficiency score was sensitive to the TADMUS Training/DSS intervention. Specifically, we see that incorporating a DSS into team processes can have a differential impact on team decision efficiency suggesting potential benefits to process and performance. This difference manifests itself to a greater extent on scores related to planning and execution decision processes.

Our rationale for this metric was that combining individual mental effort scores with overall team performance scores can be indicative of the effectiveness of training and systems interventions. By simultaneously considering individual measures of workload across multiple scenarios in conjunction with team performance we were able to illustrate how interventions reduced relative workload. The positive Team Decision Efficiency scores suggest that the Training/DSS resulted in less cognitive demand and better performance.

These analytical techniques are important because they allow us to determine the relative effectiveness of technology-enabled team processes, thereby identifying differing forms of improvement techniques for either design or training remediation. Specifically, rather than just noting performance was low, measures of efficiency allow us to determine where perceptions of workload are high versus low (see Cuevas, Fiore, & Oser, 2002). What we suggest is that this efficiency score can serve to identify human performance improvements in, <u>and problems with,</u> new training strategies and decision aiding systems. In particular, with evidence-based training and aiding systems, the efficiency score can serve to identify training remediation strategies. For example, team members reporting low workload and performing poorly may require a different form of feedback in their after action review (i.e., need to improve teamwork processes) than teams performing poorly, but reporting high workload (i.e., need to improve use of decision aiding system). As such, leveraging metrics from differing fields such as the instructional sciences allow us to produce diagnostic techniques to improve the way human-systems integration is tested in general, and how feedback is delivered and used in particular.

In conclusion, and from a broader perspective, applying such cognitive theories as CLT (see Cuevas, Fiore, Bowers, & Salas, 2004) to designing measures of human performance serves two related goals. First, from the theoretical level, it moves us closer to understanding and better diagnosing processes related to *team cognition* (Salas & Fiore, 2004). Second, from the practical level, it helps us in our efforts to transform the state of military training and decision aiding systems. In this paper we demonstrated how the Team Decision Efficiency measure can assess the combined effect of training and decision aiding. In support of analogous theorizing coming out of the instructional sciences, these techniques “can reveal important information about the cognitive consequences of instructional conditions that is not necessarily reflected by traditional performance-based measures” (p. 134, Paas & Tuovinen, 2004). Using diagnostic measurement methods can support identifying ways to reduce extrinsic cognitive load, thereby facilitating the return on investment in human systems integration design and development.

REFERENCES

Baddeley, A. (1986). Working memory. Oxford: Clarendon Press.

Baddeley, A. (1992a). Is Working Memory Working? The Fifteenth Bartlett Lecture. Quarterly Journal of Educational Psychology, 44A(1), 1-31.

Baddeley, A.D. (1992b). Working memory. Science, 255, 556-559.

Cannon-Bowers, J. A., & Salas, E. (1998). Making decisions under stress: Implications for individual and team training. Washington, DC: American Psychological Association.

Chandler, P. & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. Cognition and Instruction, 8(4), 293-332.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. Applied Cognitive Psychology, 10, 151-170.

Cohen, M.S., Freeman, J.T., & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and training strategy. In J.A. Cannon-Bowers & E. Salas (Eds.), Making decisions under stress: Implications for individual and team training (pp. 155-189). Washington, DC: American Psychological Association.

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes: Use of diagrams in computer-based training environments. Instructional Science. 30, 433 – 464.

Cuevas, H. M., Fiore, S. M., Bowers, C. A., & Salas, E. (2004). Fostering constructive cognitive and metacognitive activity in computer-based complex task training environments. Computers in Human Behavior, 20, 225-241.

Fiore, S. M., Cuevas, H. M., Scielzo, S., & Salas, E. (2002). Training individuals for distributed teams: Problem solving assessment for distributed mission research. Computers in Human Behavior, 18, 125 - 140.

Fiore, S.M., Johnston, J., Paris, C., & Smith, C. A. P. (in press). Evaluating Computerized Decision Support Systems for Teams: Using Cognitive Load and Metacognition Theory to Develop Team Cognition Measures. To be published in the *Proceedings of the 11th International Conference on Human-Computer Interaction*.

Hart, S.G., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: Elsevier.

Johnston, J.H., Poirier, J, & Smith-Jentsch, K.A. (1998). Decision making under stress: Creating a research methodology. In J.A. Cannon-Bowers & E. Salas (Eds.), Making decisions under stress: Implications for individual and team training (pp. 39-59). Washington, DC: American Psychological Association.

Johnston, J. H., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (1997). Performance measurement tools for enhancing team decision making. In M. T. Brannick, E. Salas, &

C. Prince (Eds.), Assessment and management of team performance: Theory, research, and applications (pp. 311-330). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. Applied Cognitive Psychology, 13, 351-371.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. Journal of Educational Psychology, 88, 49-63.

Mayer, R. E. (2001). Multimedia learning. Cambridge, UK: Cambridge University Press.

Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology, 90, 312-320.

McCarthy, J., Johnston, J. H., & Paris, C. (1998). Toward development of a tactical decision making under stress integrated trainer [CD-ROM] (pp. 897-904). Proceedings of the 20th Annual Interservice/Industry Training Systems and Education Conference.

Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. Journal of Educational Psychology, 91, 358-368.

Morrison, J.G., Kelly, R.T., Moore, R.A., & Hutchins, S.G. (1998). Implications for decision-making research for decision support and displays (pp. 375-406). In J.A. Cannon-Bowers & E. Salas (Eds.), Making decisions under stress: Implications for individual and team training (pp. 39-59). Washington, DC: American Psychological Association.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. Journal of Educational Psychology, 87, 319-334.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. Journal of Educational Psychology, 84, 429-434.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. Educational Psychologist, 38, 1-4.

Paas, F., & Tuovinen, J. (2004). Exploring Multidimensional Approaches to the Efficiency of Instructional Conditions. Instructional Science, 32, 133-152.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P.W.M. (2003). Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist, 38, 63-71.

Paas, F. & Van Merriënboer, J. (1993). The efficiency of instructional conditions: an approach to combine mental effort and performance measures. Human Factors, 35, 737-743.

Paas, F., Van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. Perceptual and Motor Skills, 79, 419–430.

Paris, C. R., Johnston, J. H., & Reeves, D. (2000). A schema-based approach to measuring team decision-making in a Navy combat information center. In C. McCann & R. Pigeau (Eds.), The human in command: Exploring the Modern Military Experience (pp. 263-278). New York: Kluwer Academic/Plenum Publishers.

Salas, E., & Fiore, S. M. (Editors). (2004). Team Cognition: Understanding the factors that drive process and performance. Washington, DC: American Psychological Association.

Scielzo, S, Fiore, S. M., Cuevas, H. M, & Salas, E. (2004). Diagnosticity of mental models in cognitive and metacognitive processes: Implications for synthetic task environment training. In S.G. Schiflett, L. R. Elliott, E. Salas, & M. D. Coovert (Eds.), Scaled worlds: Development, validation, and applications (pp. 181-199). Aldershot, UK: Ashgate.

Smith, C.A.P., J. Johnston, and C. Paris (2004). Decision Support for Air Warfare: Detection of Deceptive Threats, Group Decision and Negotiation, 13,129-148.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team Dimensional Training: A strategy for guided team self-correction. In J.A. Cannon-Bowers & E. Salas (Eds.), Making decisions under stress: Implications for individual and team training (pp. 271-297). Washington, DC: American Psychological Association.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257-285.

Sweller, J. (1994). Cognitive Load Theory, Learning Difficulty, and Instructional Design. Learning and Instruction, 4, 295-312.

Sweller, J. (1999). Instructional Design in Technical Areas. Melbourne: Australian Council for Educational Research.

Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. Cognition and Instruction, 12, 185-233.

Sweller, J., Chandler, P., Tierner, P., & Cooper, M. (1990). Cognitive load in the structuring of technical material. Journal of Experimental Psychology: General, 119, 176-192.

Van Merrienboer, J.J.G. & Paas, F. (1990). Automation and Schema Acquisition in Learning Elementary Computer Programming: Implications for the Design of Practice. Computers in Human Behavior, 6, 273-289.

Zsambok, C.E., & Klein, G. (1997). Naturalistic decision making. Mahwah, NJ: LEA.

Figure 1.   Mean teamwork behaviors scores across conditions.

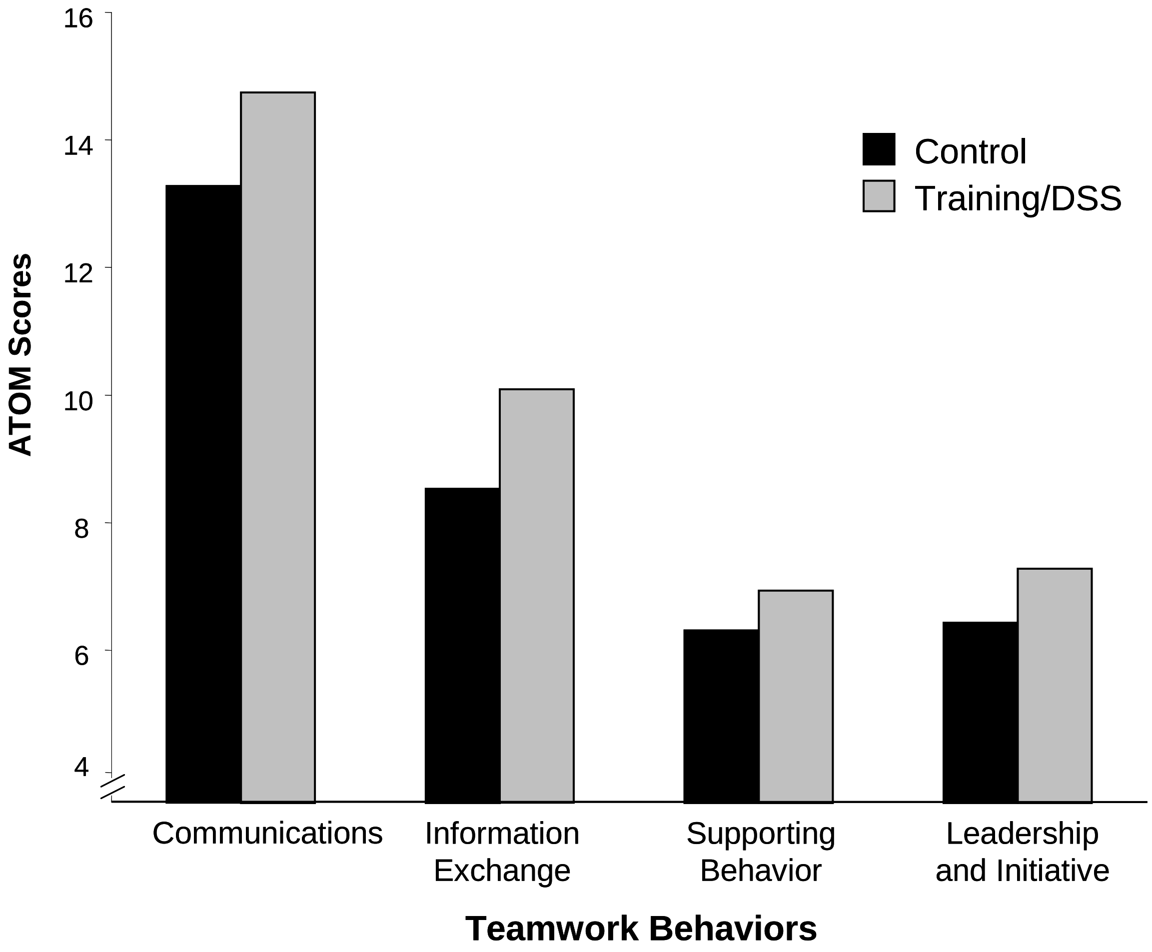
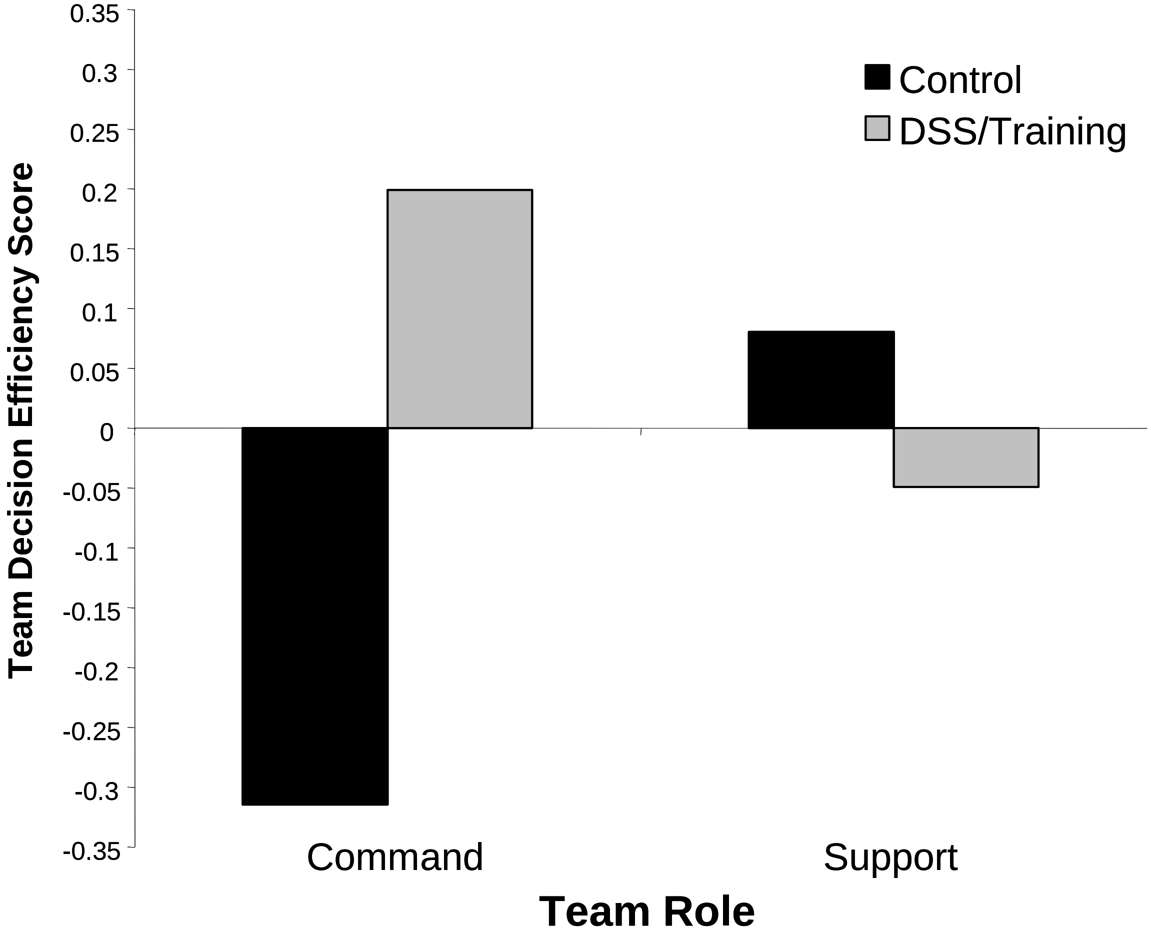


Figure 2. Standardized Team Decision Efficiency Scores for the Interaction between Condition and Role.



<u>Figure 3</u>. Standardized Decision Efficiency Scores for the interaction between Condition and Decision Type.

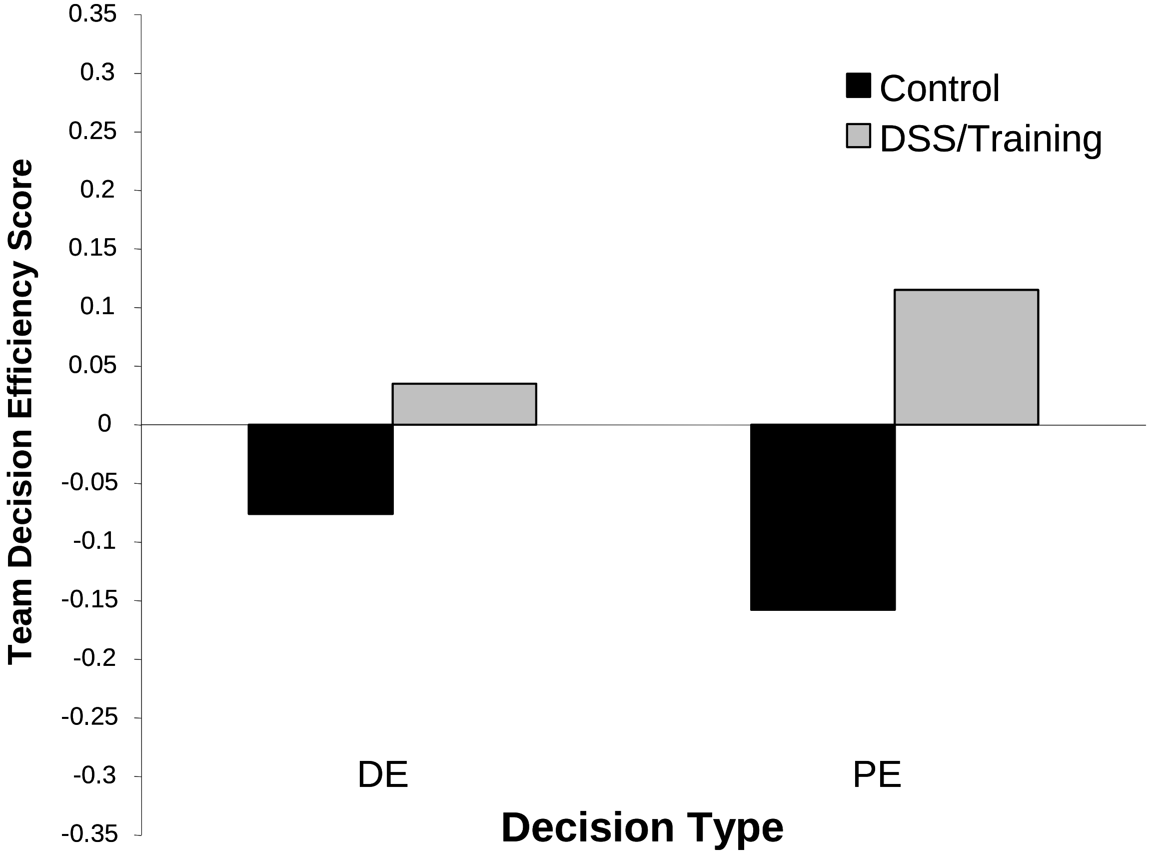
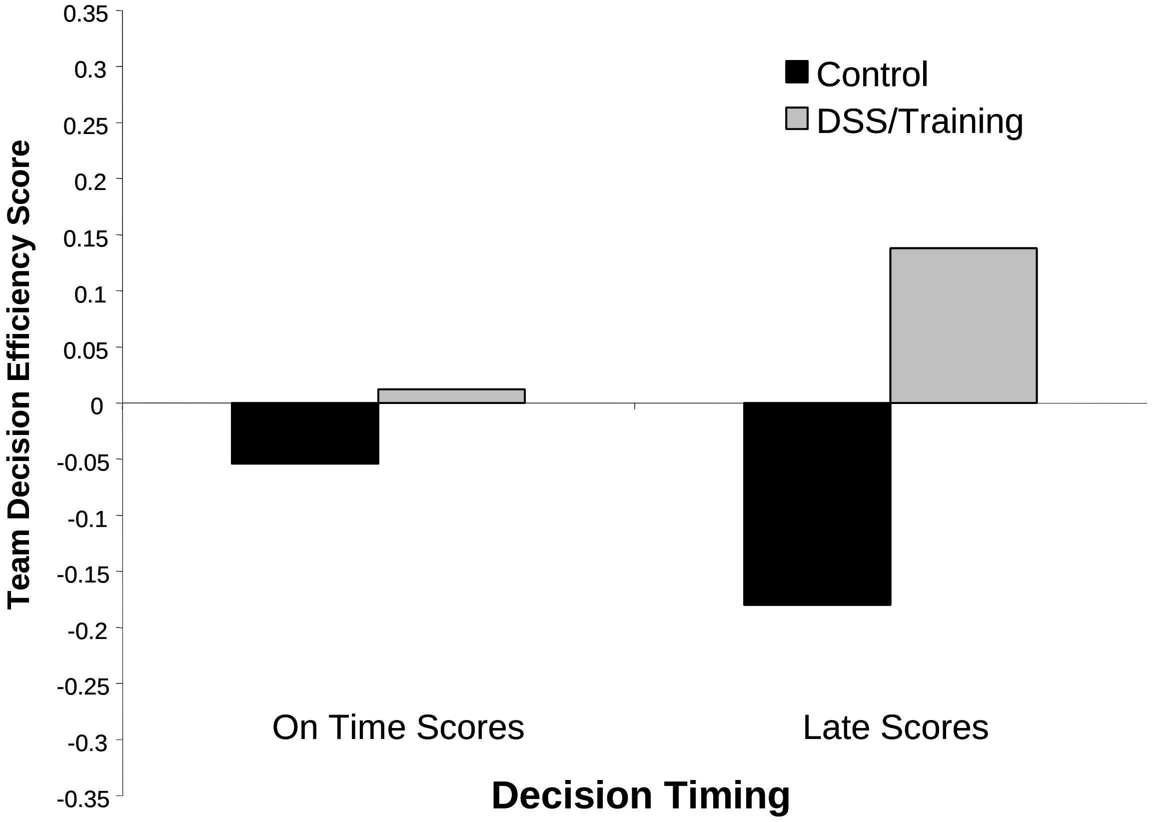


Figure 4. Standardized Decision Efficiency Scores for the Interaction between Condition and Timing of Decision.



<u>Figure 5</u>. Standardized Decision Efficiency Scores for the interaction between Condition, Timing, and Decision Task.