EVALUATION OF GEOSPATIAL DIGITAL SUPPORT PRODUCTS

Walter A. Powell* - GMU
Kathryn Blackmond Laskey - GMU
Leonard Adelman - GMU
Ryan Johnson - GMU
Shiloh Dorgan - GMU
Michael Hieb - GMU
Kenneth Braswell - AGC
Michael Powers - AGC

Center of Excellence in C4I
The Volgenau School of Information Technology and Engineering
George Mason University
4400 University Drive
Fairfax, VA 22030-4444
USA
001-703-993-1644

*Point of Contact: Walter Powell
wpowell@gmu.edu and/or (703) 993-7589/7201

Abstract

Geospatial Reasoning has been an essential aspect of military planning since the invention of cartography. Geospatial Digital Support Products (GDSPs) are ubiquitous within current military forces as well as civil and humanitarian organizations. Nevertheless, there is too little empirical evidence to quantify the military value of geospatial products to the warfighter. We conducted a hypothesis-driven experiment to evaluate the military value of the Battlefield Terrain Reasoning and Awareness – Battle Command (BTRA-BC) GDSP. Realistic scenarios and appropriate measures to assess performance were developed in collaboration with Subject Matter Experts (SMEs). The measures included time to completion, objectively assessed solution quality, subjectively assessed solution quality, and terrain understanding. BTRA-BC was integrated into the Army's Digital Topographic Support System (DTSS). A within-subjects design was employed, in which the participants completed scenarios using DTSS with and without BTRA-BC functionality. Statistical analysis of the data indicated that when the participants used BTRA-BC functionality, they created outputs faster and of higher quality without reducing their knowledge of the impact of the terrain on military decision-making. This paper discusses the scope of the current experiment, the hypotheses, the experimental design, and the results.

1. Overview

The focal point of the battlefield command post is the map. Through interactions with the map, the commander and staff collaborate to build a common operating picture (COP). This COP displays the area of operations, the militarily significant features of the terrain, the locations of adversary and friendly forces, and the evolving plan. A generation ago, planning centered on a paper map, its overlays of acetate covered with marks of grease pencils wielded by the staff members congregated around it. Today the paper map has been replaced with a digitized map projected onto a large-screen display. The grease pencil has become a mouse that officers use to draw objects and select pre-computed overlays from a pull-down menu of options. The map and overlays are stored in the computer as data structures. They are processed by algorithms that can generate, in seconds, information products it would take soldiers many hours of tedious effort to duplicate. These products can be sent instantly to relevant consumers anywhere on the Global Information Grid.

This reality of 21st Century Command and Control places a major responsibility on researchers who develop tools to support soldiers as they perform their duties. It is essential that we accurately assess the value of the tools we develop to assist in planning and situation awareness and use this assessment to shape future research and development efforts. Research indicates that sound methodologies for assessing the value of decision support tools for task performance, coupled with effective development processes that make use of the feedback thus obtained, can dramatically improve the effectiveness of decision support (Adelman, 1992; Boehm, et al., 1984; Hicks and Hartson, 1993). Intense research and development efforts are underway at many organizations which are moving the state of the art forward and pushing the latest generation of GDSPs into the field to meet the current urgent need. A rapid, effective development process is necessary if we are to provide warfighters with tools that provide force multipliers and save lives.

Geospatial Decision Support Products (GDSPs) transform commercial geographic information systems (GIS) into useful military services for Network Centric Operations. Because of their basis in commercial GIS, they have widespread applicability to fire, police, disaster relief, and other domains characterized by the need to evaluate geospatial information. GDSPs can do much more than simply speed up calculations; they are changing the way military operations are conducted. Development of these tools are shaped by military necessity, but as the new century dawns, military decision-making is itself being shaped by the automated tools that provide warfighters with a more robust situational awareness

This paper describes a project underway at the U.S. Army Engineer Research and Development Center (ERDC) to evaluate the value added to military decision making through the use of GDSPs. The specific GDSP to be evaluated is the suite of Battlespace Terrain Reasoning and Awareness – Battle Command (BTRA-BC) Tools (U.S. Army, 2003). The BTRA-BC program, which builds upon a commercial GIS tool (ARCINFO), has resulted in mature components that have been integrated into the Army's Digital Topographic Support System (DTSS), a system that provides topographic engineering support to terrain and topographic technicians as they assist military planners (Herrmann, 2002). DTSS provides geospatial data generation, collection, management, information processing, and services. The BTRA-BC GDSPs create information and knowledge products that empower soldiers with information to enhance their understanding of terrain and weather as it impacts their functional responsibilities. The BTRA-BC capabilities evaluated in this study include identification of obstacles, production of a Modified Combined Obstacles Overlay (MCOO), and generation of mobility corridors. Our experiments provided essential information to evaluate the contribution of the BTRA-BC tools in particular, and GDSPs in general to enhance the military decision making process.

This paper is organized as follows. Section 2 describes the overall scope of our research program and the scope of our initial set of experiments. Section 3 discusses the primary hypotheses which will be examined. Section 4 lays out the design of the experiment described herein and the reasoning which led to this design. Section 5 describes the measures we used to quantify the results. Sections 6 and 7 present the statistical analysis and a brief discussion of the importance of evaluation during development.

2. Scope of Experiments

Our ultimate objective is to evaluate the benefit to commanders at the brigade level and below of combining fully developed GDSPs with currently available Command and Control planning tools. The scope was limited in the first experiment, and will expand successively in later experiments. The experiment with which this paper is concerned is limited to the Intelligence Preparation of the Battlefield (IPB) process, and specifically to the terrain analysis portion of IPB.

The baseline for this series of experiments is the currently fielded DTSS suite of tools, as implemented using ARC-GIS 9.1. The DTSS tool suite consists of a package of software tools used to generate tactical decision aids for producing a number of products, including: (1) off-road and on-road speed products; (2) Combined Obstacle Overlays (COOs); (3) shaded time/distance, maneuver networks, and predictions; (4) masked/visible areas for observation; and (5) fields of fire, cover and concealment, obstacles, key terrain, and avenues of approach.

The suite of GDSPs under evaluation is the Battlefield Terrain Reasoning and Awareness – Battle Command (BTRA-BC) tool set. Our first experiment evaluated the most current version of BTRA-BC and assessed the value of BTRA-BC functionality beyond that available in DTSS. BTRA-BC consists of multiple GDSPs that generate information about the terrain and generate products which aid in analyzing the effects of terrain. Each of GDSPs utilize terrain feature data, digital elevation models, information about tactics, and techniques and system performance. BTRA-BC GDSPs produce information addressing: (1) observation, cover and concealment, obstacles and mobility, key terrain, and avenues of approach (OCOKA), (2) integrated products defining operational Positions of Advantage, (3) advanced mobility analysis, (4) digital ground and air maneuver potential, and (5) tactical structures relating information produced by the other components. Specific GDSPs support: (1) predictive multi-criteria, multi-objective maneuver, and logistical route analysis for ground platforms and forces, (2) predictive sensor performance (e.g., infrared [IR], millimeter-wave [MMW], seismic, and acoustic), (3) situation assessment, and (4) predictive threat assessment.

The research, development, products, and architectural approach of BTRA-BC are designed to enhance future force's networked Battle Command and Intelligence, Surveillance, and Reconnaissance (ISR) processes through the incorporation of actionable terrain and weather information and decision support tools. The BTRA-BC approach is wholly consistent with the Army's Future Combat System's (FCS's) System of Systems and the Defense Information Systems Agency's (DISA’s) Network Centric Enterprise Services concepts. If successful, BTRA-BC will be capable of benefiting the FCS C4ISR appliqué and the Joint Distributed Common Ground Station family of ISR systems.

3. Hypotheses

In order to evaluate the military value of BTRA-BC, we needed to establish an operational definition of the value of a geospatial product with respect to military decision-making. Discussions with both military operational planners and members of the BTRA-BC development team clarified the areas where GDSPs in general and BTRA-BC GDSPs in particular would be valuable to the military decision-maker. The first and most obvious perceived value of a GDSP is in its ability to reduce the time spent generating a given tactical decision product. Since the timeframe available to military decision makers is limited, the time saved using GDSPs to produce the desired output can free up time for a more thorough analysis of the large amount of data available. A more complete analysis is expected to result in a higher quality output which will be of more value to the military decision-maker.

Another valuable contribution of a GDSP is automation. Many of the initial tasks traditionally done by terrain analysts with paper maps are sufficiently rote in nature that a GDSP, given digital data and the appropriate parameters, can perform these functions more quickly and with less error than a human. With automation, however, comes the concern that automating these tasks may reduce the analyst’s familiarity with the terrain and understanding of its impact on the military planning process. The experts we consulted believe that the automated tasks are not analytical but procedural, and that using the output of the GDSP will not compromise the analysis of the data or the level of understanding of the analyst. The experiment tests this prediction.

From the discussion above, we constructed our three primary hypotheses. In comparison with analysts using currently available tools (DTSS without BTRA-BC), we hypothesized that trained and experienced terrain technicians who use DTSS with BTRA-BC would:

1. *Produce certain terrain-dependent Intelligence Preparation of the Battlefield (IPB) outputs* ***more quickly***. Rationale: The automation in BTRA-BC should allow the participants to complete the repetitive, tedious, and rote tasks more quickly.
2. *Produce a* ***higher quality*** *output*. Rationale: The automaton in BTRA-BC should minimize errors of omission in calculation and standardize the graphical representation of important terrain features.
3. *Display as* ***good an understanding*** *of the impact of the given terrain on military decision-making*. Rationale: The judgment required to complete the required tasks will still be required when using BTRA-BC.

The secondary hypotheses investigated became apparent as the operational definition of military value and the design of the experiment evolved. The automation of previously manual tasks, which adds value to using a GDSP, would likely reduce the variation in the output. Because this reduction in variation does not necessarily add value, this was not considered a primary hypothesis. The structure of the experiments requires the repetition of various tasks and there was concern that a learning effect might affect results. The consensus of experts in terrain analysis indicated that this would be a very minor effect; therefore, learning effects are considered secondary hypotheses. The secondary hypotheses investigated included:

4. *The output generated with BTRA-BC tools would be* ***more uniform*** *i.e. have less variance in the first two of the three categories above (speed and quality), than that generated without the use of BTRA-BC.* Rationale: Less variation in the output when using BTRA-BC is expected due to the level of automation incorporated into BTRA-BC.
5. *There would not be a* ***learning effect due to experimental design***. Rationale: The participants have previous training and extensive experience using the C2 planning environment used in the experiment. The tasks the participants are asked to perform are those that they normally perform with C2 planning tools and the participants will be trained to proficiency on BTRA-BC. A single additional usage should not lead to a learning effect. However, repetition of tasks in a within-subjects experiment creates an opportunity for a System Order effect where the temporal order in which the systems are encountered can affect the performance of the participant. Consequently, the order that the participants used the systems was randomized and counterbalanced. In addition, we tested statistically for a learning effect due to the experimental design.
6. *The participants will consider DTSS with BTRA-BC superior to DTSS without BTRA-BC for planning with respect to the time to completion, quality of the plan, their terrain understanding, the usability of the system, and overall.*

4. Study Design

The study design employs a factorial design with three independent variables: System (with and without BTRA-BC functionality), System Order (whether the first scenario is worked with or without BTRA-BC functionality), and Scenario Order (whether scenario 1 or 2 is worked first). System was a within-subject variable because all participants worked a scenario with and without BTRA functionality. A within-subjects design is particularly valuable when the number of available participants is limited, as in the current case. Results from the sets of tasks can be compared for each participant, thus eliminating participant-specific differences that might add variability to the results. System Order and Scenario Order were between-subjects variables because any given participant can only be in one ordered sequence for these variables. This factorial design allowed us to obtain potentially statistically significant results in the System variable, our variable of primary interest, while minimizing the total number of subjects required.

The participants performed the same tasks on two similar military planning scenarios, where one of the tasks was performed with BTRA-BC functions in addition to DTSS functions and the other task with DTSS functions only. The two conditions were essentially identical except for the use of BTRA-BC in addition to currently deployed geospatial tools. The order of the conditions were randomly selected and counterbalanced. Randomizing the order of the conditions enabled the analysis to control for learning effects.

The instructions, tasks, requested outputs, and evaluation of these outputs were the same in both conditions, with the exception of geographic references necessitated by the requirement to have different geographic areas for each scenario. Different geographic areas are required to prevent participants from just repeating their responses from the first condition when they form responses for the second condition. The two areas were carefully selected for their geographic similarity such that the complexity of the tasks performed by the participants and the expected results would be as nearly identical as possible.

The participants were Army and Marine Corps enlisted personnel who had all been trained as Terrain Analysts. All participants had completed the Basic Terrain Analysis Course (BTAC) at the National Geospatial Intelligence Agency University (NGAU), and were currently enrolled in the Advanced Terrain Analyst Course (ATAC). The 18 participants consisted of 1 Army Chief Warrant Officer, 10 Army Staff Sergeants, 3 Army Sergeants, 3 Marine Corps Sergeants, and a Marine Corps Corporal. Their recent operational experience with terrain analysis varied from several years of continuous experience to only formal training. No analysis was performed to investigate dependence of results on the participants’ characteristics. Even though the analysis treated the group as homogeneous in its experience, the design was counterbalanced, as described below, to control for individual differential experience.

The participants were split into two groups that were evenly balanced as to the ability and knowledge of the participants as determined by the instructors of the ATAC class. The first group performed the set of tasks first without BTRA-BC and then with BTRA-BC. The second group reversed the order of tasks. The groups were further divided into two subgroups while maintaining the balance of ability and knowledge. The first of these subgroups in each group performed the tasks on terrain area one first then terrain area two while the second subgroup reversed the order of the terrain areas. This procedure allowed us to control for variability due to the experience of the participants, the order that the systems were used, and the order that the terrain areas were used.

The tasks consisted of that portion of the Intelligence Preparation of the Battlefield (IPB) beginning with analyzing the specific terrain given a Consolidated Obstacle Overlay (COO) up to the point of generating potential Avenues of Approach (AAs). Specific tasks included (1) identifying Mobility Corridors (MC), (2) categorizing MCs by size, (3) grouping MCs to form potential Avenues of Approach (AA), (4) planning routes for three vehicle types, (5) identifying choke points on potential AAs, (6) calculating transit times, and (7) recommending subordinate Areas of Responsibility, in this case recommending battalion boundaries.

The participants produced a graphic overlay depicting the results of the above tasks. In order to gather data for the measures in section 5, the participants also completed a questionnaire which assessed their understanding of the effects of terrain on the military planning process. A separate questionnaire assessed their subjective experience with both systems.

Prior to beginning the tasks, both groups of participants received standardized training on the use of BTRA-BC. The training was sufficient to allow the participants to perform the required tasks given the participants' level of experience with automated systems and included training on the modes and features unique to BTRA-BC. The last phase of the training required the participants to perform tasks based on the training that were similar to those that the participants encountered during the experiment, but of lesser complexity.

5. Measures

The criteria for assessing the effectiveness of BTRA-BC were (1) the rapidity with which the requested outputs can be produced, (2) the quality of those outputs, (3) the level of understanding of the participants of the impact of terrain on military decision making, (4) the uniformity of the participants' scores on the first 3 measures, and (5) the perception of the participants themselves of the merits of the additional BTRA-BC functionality for the first 3 measures.

Time to Completion. The evaluation of how quickly the desired outputs can be produced was measured objectively and independently of the experimental condition by logging the amount of time it took participants to complete the tasks. As the maximum duration of each trial was 4.5 hours, the actual time each participant spent performing a scenario using DTSS with or without BTRA-BC functionality was calculated by taking the difference between the start and stop times and subtracting any break time.

Quality. We considered there to be two measures which, when combined, constitute the outputs' quality: (1) the information presented and (2) how the information was presented. We used two measures of the quality of output. The first measure, which we called objective scoring, considered only the information presented, It was objective in that it considered only criteria that could be counted or revealed by answering yes/no questions. The objective criteria included: (1) sizing of MCs, (2) actual number of choke points on each AA, (3) number of choke points identified on each AA, (4) sizing of identified choke points, (5) smallest choke point on each AA, (6) AA transit of no go areas, (7) number of AAs in each BN AO, (8) Buffering of each AA, and (9) AAs contained within operational boundaries.

The second measure, which we called subjective scoring, made use of subject-matter experts (SMEs) to score participants' solutions. The SMEs judged the quality of the output with respect to the usefulness to the commander. The evaluations were based on mutually agreed-upon criteria. Each SME independently evaluated the quality of the output, and then the SMEs discussed their evaluations to arrive at consensus scores for the subjective scoring. In both the objective and subjective evaluation, scorers rated each question on a 5-point Likert scale. The focus of the subjective evaluation was relatively narrow, focusing on the TSO being evaluated.

The factors considered in the subjective criteria included: (1) directness of the AAs, (2) clarity of presentation of AAs, (3) sizing and clarity of presentation of MCs, (4) sizing and clarity of presentation of choke points, (5) clarity of recommended AAs, (6) completeness of AAs, (7) the appropriateness of the recommended battalion boundary, and (8) overall clarity and presentation.

Due to the differences in the graphical representation produced by DTSS with and without BTRA-BC, we were unable to obtain blind scoring for the subjective evaluation. BTRA-BC's graphics are not easy to replicate manually in DTSS. Although the evaluators were not told which of the products were produced using BTRA-BC, given the number of products they graded it is likely that they would be able to determine which were produced with BTRA-BC. Although the outputs are distinguishable as to their source, the evaluators were independent SMEs and we treated their evaluation as such.

Terrain Understanding. To evaluate the participants' understanding of the terrain and their understanding of the impact of the specific terrain on military decision-making, we administered a questionnaire. The questions could not be answered directly from the outputs of DTSS or BTRA-BC. The answers required judgment and reasoning about the terrain and its effect on the military decision making, and not just regurgitating data presented by DTSS or BTRA-BC. Like the subjective evaluation, the SMEs evaluated the participants' answers on a 5-point Likert scale. The questions addressed reasoning about the recommended AA, the recommended BN boundary, a recommended location for a brigade forward logistics base, analysis of the MCs, and analysis of the chokepoints.

Participant Evaluation. A questionnaire was used to elicit the participants' reaction to DTSS with BTRA-BC when compared to DTSS without BTRA-BC, for speed, quality, and terrain understanding. Specifically, participants were asked to indicate how quickly they could complete the following five tasks using just DTSS versus using DTSS with BTRA: (1) identifying potential AAs; (2) accomplishing route planning; (3) identifying choke points on AAs; (4) calculating travel times; and (5) identifying Battalion Areas of Responsibility. Participants' mean response for these five tasks represented their opinion about how fast they could perform tasks using DTSS without versus with BTRA. Participants also compared the quality of their plans using the two systems for the same five tasks, with their mean score indicating how well they thought they performed tasks without versus with BTRA. Lastly, they answered three questions about their terrain understanding without BTRA versus with BTRA. The participants answered each of the 13 questions using the 5-point Likert scale shown in Figure 1.

DTSS assists much more	DTSS assists more	DTSS and BTRA are equal	BTRA assists more	BTRA assists much more
1	2	3	4	5

Figure 1: Comparison Scale

Two versions of the questionnaire were created: one with DTSS occupying the left side of the scale and the other with BTRA occupying the left side of the scale. The scale was reversed for the second version to control for a possible effect of the scale ordering on participants' responses. The participants were randomly assigned one of the two versions of the questionnaire. After collecting the data from the questionnaire, we used a single scale by converting the scale used on one version to the other for quantitative analysis.

6. Analyses of Results

Time to Completion. A repeated-measures analysis of variance (ANOVA) indicated that participants' average time to completion when they used DTSS with BTRA-BC ($\bar{x} = 1.140$, $s = 0.231$) was significantly faster ($p < 0.001$) than when they used DTSS without BTRA-BC ($\bar{x} = 3.120$, $s = 0.890$). On average, participants completed the tasks using DTSS with BTRA-BC 64% faster than without BTRA-BC.

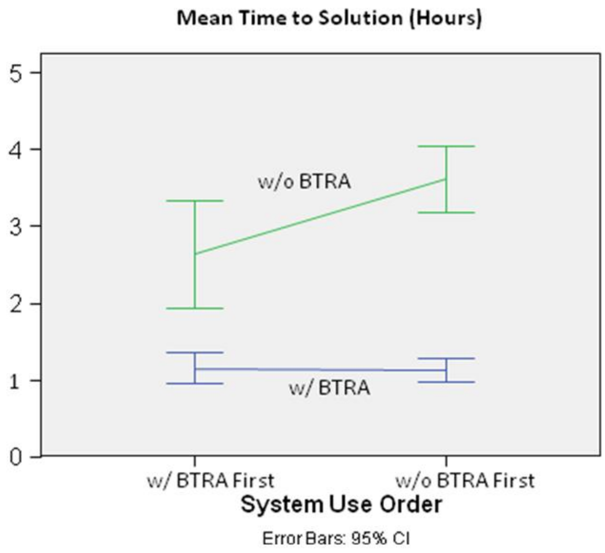

Figure 2: Time to Completion (in hours)

The repeated-measures ANOVA also provided strong statistical evidence that a learning effect ($p < 0.01$) was present due to the order that the systems were used, as displayed in Figure 2. The average time to completion for DTSS without BTRA-BC was faster for the participants who used DTSS with BTRA-BC first ($\bar{x} = 2.635$, $s = 0.909$), than the participants who used DTSS without BTRA-BC first ($\bar{x} = 3.613$, $s = 0.567$). Whether the participants used DTSS with BTRA-BC first or used DTSS without BTRA-BC first did not have a significant effect on their average time to completion for DTSS with BTRA-BC.

An F-test ($p < 0.0001$) for unequal variances indicated a significant difference in variance between the time to completion for DTSS with BTRA-BC and DTSS without BTRA-BC. The variance in time to completion when the participants used DTSS with BTRA-BC ($s^2 = 0.053$) was significantly lower than when they used DTSS without BTRA-BC ($s^2 = 0.793$).

Objective Quality. A repeated-measures ANOVA indicated strong statistical evidence ($p < 0.001$) that participants' average objective quality score when they used DTSS with BTRA-BC ($\bar{x} = 3.850$, $s = 0.626$) was significantly higher than when they used DTSS without BTRA-BC ($\bar{x} = 2.920$, $s = 0.609$). Because the data were somewhat non-normal, a Wilcoxon Signed Ranks test also was performed. The test confirmed a System Order effect. The mean rank and sum of ranks when participants used DTSS with BTRA-BC (5.10, 25.50) were significantly less when they used DTSS without BTRA-BC (10.63, 127.50), thus reinforcing the ANOVA results.

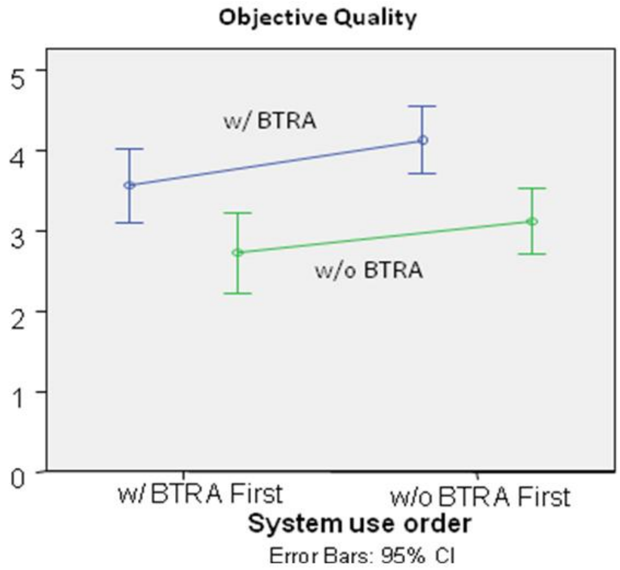

Figure 3: Objective Quality

The repeated-measures ANOVA on the objective quality data also showed significant evidence ($p = 0.038$) of a between-subjects System Order effect. The participants who used DTSS without BTRA-BC first scored higher, on average, on the objective quality measure for both DTSS with BTRA-BC ($\bar{x}$ = 4.133, s = 0.534) and DTSS without BTRA-BC ($\bar{x}$ = 3.114, s = 0.527) than the participants who used DTSS with BTRA-BC first, as shown in Figure 3. The participants who used DTSS with BTRA first, on average, scored lower on both DTSS with BTRA-BC ($\bar{x}$ = 3.565, s = 0.605) and DTSS without BTRA-BC ($\bar{x}$ = 2.725, s = 0.651).

A repeated-measures ANOVA was performed on each of the nine objective quality measures. In four of the five objective quality measures where the average score for DTSS with BTRA-BC was higher than that of DTSS without BTRA-BC, there was strong evidence that DTSS with BTRA-BC was superior ($p < 0.0001$). Although this result was not adjusted for multiple comparisons, it is strong enough to inspire confidence. In contrast, for the four measures in which the average score for DTSS without BTRA-BC was higher than that for DTSS with BTRA-BC, only one measure was statistically significant ($p = 0.049$).

An equal variance F-test ($p = 0.440$) did not indicate a significant difference in variance between the objective quality scores for DTSS with BTRA-BC and DTSS without BTRA-BC. Participants' scores for objective quality did not exhibit significantly lower variance when they used DTSS with BTRA-BC ($s^2 = 0.392$) than when they used DTSS without BTRA-BC ($s^2 = 0.371$).

Subjective Quality. The statistical analysis of the subjective quality data also yielded important insights. A repeated-measures ANOVA provided strong statistical evidence ($p = 0.003$) that participants' average subjective quality score when they used DTSS with BTRA-BC ($\bar{x}$ = 3.400, s = 0.425) was significantly higher than when they used DTSS without BTRA-BC ($\bar{x}$ = 2.720, s = 0.749), as shown in Figure 4. There was no indication of a significant learning effect or interaction.

An equal variance F-test ($p = 0.012$) indicated a significant difference in variance between the subjective quality scores for DTSS with BTRA-BC and DTSS without BTRA-BC. The variance in subjective quality when participants used DTSS with BTRA-BC ($s^2 = 0.180$) was significantly lower than when they used DTSS without BTRA-BC ($s^2 = 0.561$).

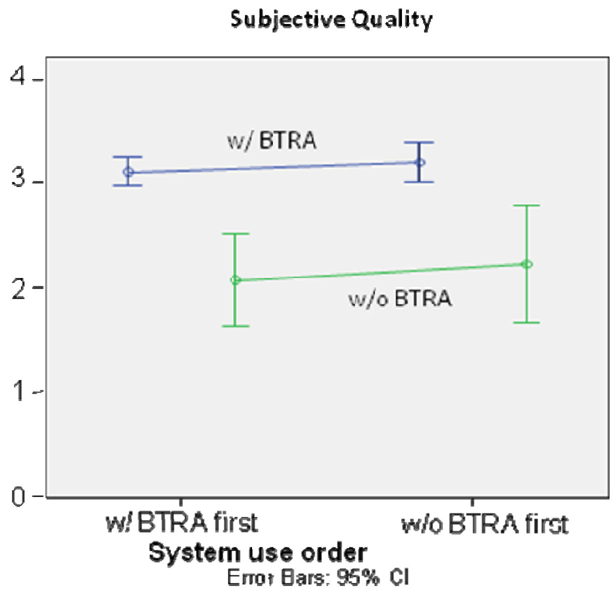

Figure 4: Subjective Quality

Terrain Understanding. A repeated-measures ANOVA provided some support that ($p = 0.059$) that knowledge and understanding of terrain was greater when participants used DTSS with than without BTRA-BC ($\bar{x}$ = 3.185, s = 0.861) ($\bar{x}$ = 2.565, s = 0.950). There was no indication of any other significant effects or interactions. In addition, an equal variance F-test ($p = 0.345$) did not indicate a significant difference in variance between the terrain understanding scores for DTSS with BTRA-BC ($s^2 = 0.741$) than without BTRA-BC ($s^2 = 0.902$).

Participant Evaluation. We calculated the participants' mean answers to questions measuring their opinions about the relative effectiveness of their planning with and without BTRA-BC on three hypotheses: the (1) speed and (2) quality with which they thought they could perform five tasks, and their (3) terrain understanding.

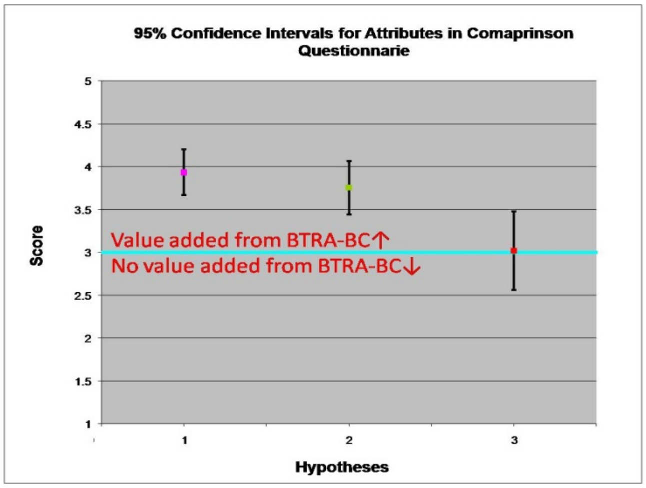

Figure 5: Participant Evaluation data

Student's t-tests were conducted for each hypothesis to assess whether the mean scores were significantly different than 3.0, the neutral point on our five-point scale. Analysis from questions relating to Hypotheses 1 [Speed] ($\bar{x}$ = 3.933, s = 0.566) and Hypothesis 2 [Quality] ($\bar{x}$ = 3.755, s = 0.661) yielded strong evidence ($p < 0.0001$) that participants' mean evaluation scores for DTSS with BTRA-BC were higher than those for DTSS without BTRA-BC. Analysis for data relating to Hypothesis 3 [Terrain Understanding] ($\bar{x}$ = 3.019, s = 0.973) did not yield statistically significant evidence (p = 0.937) of a difference, although the mean was slightly higher for the with BTRA condition. Figure 5 displays the means and 95% confidence intervals for each of the three groups of questions related to the three primary hypotheses.

7. Discussion

The analyses of all four measures strongly support the primary hypotheses: participants using DTSS with BTRA-BC produced outputs (1) faster, with (2) higher quality, and (3) had as good, if not better, an understanding of the impact of the terrain. Table 1 below summarizes the data for the averages, variances, and significance for each measure. Statistical significance at or below the p = .05 significance level is indicated by boldface type.

		Average		Variance	
		w/ BTRA	w/o BTRA	w/ BTRA	w/ BTRA
Time to Completion		**1.136**	3.124	**0.053**	0.793
Quality	Objective	**3.849**	2.920	0.392	0.371
	Subjective	**3.399**	2.719	**0.180**	0.561
Terrain Understanding		3.185	2.565	0.741	0.902

Table 1: Summary of metric average (bold indicates statistically better)

Hypothesis 1: Time to Completion – The strong statistical evidence supporting the first hypothesis suggests that it is likely that terrain analysts currently using DTSS would complete tasks similar to these much more quickly if the BTRA-BC functionality were added. This time saving would likely occur at all levels which have access to the BTRA-BC tools and could thus reduce the IPB portion of the Military Planning task.

The learning effect for DTSS without BTRA-BC, indicated in the statistical analysis of the time to completion data, was unexpected and we hypothesize that this effect was due to the participants being less familiar with solving the terrain analysis problem than we had expected. Experience with DTSS with BTRA-BC prior to using DTSS without BTRA-BC appears to have improved their understanding of how to approach the problem as evidenced by the reduction in time to completion for the DTSS without BTRA-BC condition.

Hypothesis 2: Quality – Both objective and subjective measures of quality strongly support the hypothesis that participants using DTSS with BTRA-BC produced higher quality outputs. The subjective measures of quality were not as concrete as the objective measures and the time to completion measure, because the subjective quality scores were determined by SME evaluation. Significant pains were taken to ensure that the SMEs were qualified to render informed and independent judgments. For objective quality, participants' scores showed, on average, a 31.8% improvement when DTSS with BTRA-BC was used. For subjective quality, participants scored 25% higher, on average, when DTSS with BTRA-BC was used.

Of the nine objective measures, five had a significant difference between the scores for the two systems. Four were in favor of DTSS with BTRA-BC being superior and one was in favor of DTSS without BTRA-BC being superior. The single objective measure for which the score for DTSS without BTRA-BC was superior was a yes/no question addressing whether or not the recommended route crossed an operational boundary. None of the manually generated routes (DTSS without BTRA-BC) crossed an operational boundary whereas four of the eighteen system generated (DTSS with BTRA) routes crossed an operational boundary. This measure of objective quality provides an interesting confirmation of an effect of decision support systems noted in the literature, that automated tools tend to increase errors of commission. Errors of commission occur when users of automated tools "inappropriately follow automated information or directives [even] …when other information in the environment contradicts or is inconsistent" (Skitka et al., 1999). We hypothesize that this was an instance of automation resulting in errors of commission in that at least some of the participants accepted the system-generated routes which contained errors as correct and did not evaluate the graphical presentation of the routes.

In response to this finding, the BTRA-BC routing algorithm has since been amended to prevent generated routes from crossing operational boundaries. This programming solution will prevent this specific error of commission from occurring in the future, but our results reconfirm that errors of commission can occur.

Hypothesis 3: Terrain Understanding – There is strong statistical evidence to support the hypothesis that, on average, the participants displayed *as good an understanding* of the impact of the given terrain on military decision making with BTRA-BC functionality. This hypothesis was phrased in this way because we were concerned that the use of DTSS with BTRA-BC might limit the scope of participants' terrain analysis, and thereby decrease their understanding of the terrain's impact on military operations. Even though the participants' overall understanding of the impact of terrain improved, incorrect instances of accepting the system's output still occurred, as shown in the previous discussion about errors of commission in our quality measure.

Hypothesis 4: Uniformity – The statistical evidence to support this hypothesis was mixed. Participants produced more uniform output when using DTSS with BTRA-BC for two of the four measures: the subjective quality measure ($p = 0.01$) and time to completion measure ($p <$

0.0001). The objective quality (p = 0.44) and terrain understanding measures (p = 0.34) did not support the hypothesis.

We believe that the uniformity observed for two of the four measures was partially due to the automation of certain functions by DTSS with BTRA-BC. The automation offloads certain rote tasks from the user to the system. The time to completion measure demonstrates this well. When using DTSS with BTRA-BC, much of the output is generated by the BTRA-BC algorithms; the time to generate this output is relatively constant. In contrast, the subjective quality data most likely exhibited a higher degree of uniformity when participants used DTSS with BTRA-BC because the generated output graphics were well designed, easily understood, and used the same symbology across all of the outputs.

Hypothesis 5: Learning Effect – Learning effects of varying magnitude and significance were observed in the experimental data. The real question is whether the learning effects observed were due to the experimental design. If the learning effect was due to experimental design, then there should have been improvement in time to completion (learning effect) for the second scenario for both conditions. However, the task was performed faster for the second scenario only when participants worked the first scenario with BTRA-BC functionality.

Participants who used DTSS with BTRA first completed their scenario without BTRA-BC 28% faster than the participants who used DTSS without BTRA-BC first. This would seem to indicate that the participants learned from using DTSS with BTRA-BC first, but not when using DTSS without BTRA-BC first. This analysis confirms observations made during the experiment that the participants were not as well versed in solving the terrain analysis problem as we had expected. Although all the participants were all enrolled in the Advanced Terrain Analysis Course (ATAC), not all the participants had the same level of recent operational experience with terrain analysis. Participants with less recent experience seemed to be initially less familiar with how to approach the tactical problem using DTSS than those with more recent experience. Those less experienced participants may have experienced steep learning curves concerning how to approach the terrain analysis problem when using DTSS with BTRA first. Overall, from the analysis of time to completion data, not only did DTSS with BTRA-BC significantly reduce the time required to complete the problem, but exposure to DTSS with BTRA-BC also seems to have taught the participants about the nature of the problem.

There was a significant learning effect for the objective quality measures. The participants who used DTSS without BTRA first scored higher, on average, for both DTSS with BTRA-BC and DTSS without BTRA-BC. The explanation for this interaction is unclear and requires further research, but we submit the following as one possible explanation. First, participants who were using DTSS without BTRA as their first trial may have been more careful in their analyses than those participants who used DTSS without BTRA after using DTSS with BTRA. The participants who used DTSS without BTRA second may have been less patient when having to do manually analyses that had been done previously with BTRA and thus taken less care in their analyses. Second, the participants who used DTSS without BTRA first, based on their analyses without BTRA, may have been more careful in their analyses when evaluating DTSS with BTRA output.

Hypothesis 6: Participant Evaluation – The participants' subjective evaluations of the relative merits of DTSS with BTRA-BC were consistent with the results for speed and quality presented previously. Participants believed DTSS with BTRA-BC allowed them to produce terrain analysis products more quickly (hypothesis 1) and with higher quality (hypothesis 2). They did not think it improved their terrain understanding (hypothesis 3).

We believe that most of the participants were impressed with the performance and functionality of DTSS with BTRA-BC. Although, a few of the participants were skeptical of the BTRA-BC functionality because they believed that they relied too much on the automation of certain processes, most of the participants indicated that they would like to have the BTRA-BC functionality in the field in the near future.

References

Adelman, L. *Evaluating Decision Support and Expert Systems*, Wiley, 1992.

Boehm, B.W., Gray, T.E. and Seewaldt, T. Prototyping vs. Specification: A Multi-Project Experiment. Proceedings of Seventh International Conference on Software Engineering, New York: ACM and IEEE, 473-484, 1984.

Herrmann, C.G. (Maj). DTSS Puts the "Visual" in Battlespace Visualization, *Military Intelligence Professional Bulletin*, Oct-Dec, 2002.

Hicks, D. and Hartson, H.R. Developing User Interfaces: Ensuring Usability Through Product and Process. Wiley, 1993.

Skitka, L.J., K.L. Mosier, and M. Burdick, *Does automation bias decision-making.* International Journal of Human-Computer Studies, 1999. **51**: p. 991-1006.

U.S. Army Corps of Engineers, *Battlefield Terrain Reasoning and Awareness (BTRA) Fact Sheet*, U.S. Army Engineer Research and Development Center, Topographic Engineering Center, 2003, http://www.tec.army.mil/fact_sheet/BTRA.pdf.