

# An Experimental Approach to Measuring the Effects of a Controller Conflict Probe in a Free Routing Environment

Dr. Karol Kerns

**Abstract**--The unstructured traffic environment that is gradually evolving in the National Airspace System may necessitate decision aiding tools to help controllers manage aircraft traffic efficiently and within acceptable safety and workload levels. An experiment was designed to assess the effect of conflict detection and trial planning resolution aids and traffic conditions on the performance of controllers and the efficiency of flight. Twelve controllers participated in a high fidelity simulation study in the Indianapolis Center dynamic simulation facility using the User Request Evaluation Tool (URET). Study results indicated that URET clearly improves the acceptability of operations and may enhance safety, particularly in the unstructured environment. In addition, the study findings strongly support a shift toward more strategic air traffic control with conflict probe as well as a significant increase in the contribution of the D controller to the primary sector tasks. Because of study limitations, it was not possible to associate any quantitative estimates of airspace user benefits with the potentially more efficient resolution strategies. The results and experimental approach are discussed in terms of their contribution to measurement methods and issues in the evolutionary progression toward a free flight-based air traffic system.

**Index Terms**--Conflict probe, controller decision support, benefits assessment, human factors

## I. INTRODUCTION

Current plans to modernize the National Airspace System (NAS) recognize that both procedural changes and new technologies are necessary to realize significant improvements in the efficiency and capacity of the NAS. Under the free flight concept [13], the use of procedural restrictions by air traffic control (ATC) to predefine routes will be gradually reduced, allowing airspace users to fly more direct and fuel efficient routes. Often, these preferred routes are fuel-efficient paths that do not conform to charted airways or other predefined routes. Today, air traffic controllers rely on structured routes and traffic organization for managing separation problems. The random conflict locations that occur with unstructured routes are not easily detected from the flight information currently available on the flight progress strip. To offset the additional demands placed on controllers operating in the

unstructured routing environment, the free flight concept also proposes a concomitant deployment of a conflict probe decision aid [14].

In the Free Flight Phase 1 (FFP1) Program, the core capabilities of several decision support systems will be evaluated at a number of operational sites. The FFP1 User Request Evaluation Tool (URET) capability is a set of tools designed to help controllers detect and resolve conflicts. The URET uses data from the en route Host computer system (HCS) to model 4-D trajectories for all aircraft with flight plans. Trajectories are updated from radar position reports and flight plan amendments. A conflict probe function carries out continuous, automatic conflict detection and alerts the appropriate controller when a conflict is found. The URET also provides a trial planning capability which allows the controller to specify a trial flight plan modification and determine whether the trial plan creates another conflict.

A URET prototype was deployed to the Indianapolis Air Route Traffic Control Center for field trials in January 1996. Since then, several incremental versions of the URET capabilities have been delivered to the Indianapolis and Memphis centers and evaluated in live operations at selected sectors [2] [1]. The FFP1 version of URET is expected to be in place and provide measurable benefits by the end of 2002.

Although research has begun to address assessment methods for controller decision support tools, this work is immature and the reliability and applicability of the findings are limited. Technical guidance on methods and measures is needed to support the evolutionary system development process envisioned in [14]. This experiment was conducted in the dynamic simulation (DYSIM) facility at Indianapolis Center (ZID) to identify and quantify benefits associated with use of the URET in the current and emerging unstructured traffic environments. The primary objectives of this study were to (1) determine whether the conflict probe capability can yield more efficient flight paths within acceptable safety and workload limits and (2) determine whether these effects generalize across current traffic and emerging free routing conditions. Since the URET capability also reflects important concepts about how to manage traffic complexity and controller workload,

a third objective of the study was to analyze controller performance impacts of the conflict probe capability.

The objectives and results of the study reported here were limited by the evolutionary state of the URET capabilities in place at the time, as well as the procedures, practices, and controller proficiency levels that pertained at the time. Consequently, while the experiment was designed to provide a “snapshot” measurement of benefit potential in that state, the preliminary assessment reported here was more focused on the development and tuning of the measures, methods, and tools that would be needed to routinely assess progress in achieving benefits through the URET life cycle. In the evolutionary development environment, the experimental approach has practical use in validating a thread of core methods and measurements that can be applied in laboratory and operational settings [15]. This paper documents the design and results of the initial study and discusses the validity of the approach and measurement methods for future evaluations.

### *A. Background*

Since about 1980 efforts to develop controller decision aids for conflict prediction and resolution have been underway [5]. Several conflict probe decision aids are now under development. The URET is a derivative of an earlier research and development program conducted by the FAA and the MITRE Corporation Center for Advanced Aviation System Development for several years prior to the start of the FFP1 program. And both the National Aeronautics and Space Administration (NASA) and the European Organization for the Safety of Air Navigation (EUROCONTROL) have been researching conflict probe tools. In addition, some research has been performed on free routing, since the free flight concept was formulated in 1995. Within this body of work, there are relatively few empirical or controlled studies and at this point the results of these studies are equivocal. For this experiment, the prior research mainly gives methodological background and identifies lessons learned in measurement approaches.

The research literature generally shows that conflict probe decision aids reduce or have a neutral impact on controller workload [9][8][7][2][1]. However, the observed workload effects depend on (1) design and performance characteristics of the decision aids [17][12] and (2) whether single- or two-controller teams were assessed [8][7][9][12]. The research further shows that objective and subjective workload measures are needed to assess the effects of traffic and automation influences. Objective measures, such as physiological measures or secondary task performance, have been shown to be reliable indicators of increased traffic demands. In addition to measuring the influence of traffic, the subjective measures are also sensitive to automation influences that may limit acceptance and use of new technologies [8][7].

Empirical evidence of increased flight efficiency with advanced decision aids is limited and somewhat ambiguous

[11][9][17]. Despite inconsistencies in results, this research indicates limitations of some measures and a lack of clarity in defining metrics and their relationship to user benefits. For example, [17] points out that in the en route environment a measure of time in sector has significant limitations as an indicator of user benefits. Losses or gains in flight time through a single sector may be counterbalanced by changes elsewhere. Lessons drawn from this research indicate that measures of total distance and total flight time from origin to destination are the most promising criteria for making comparisons and measuring changes in flight efficiency. Distance flown captures the effects of horizontal maneuvers while time flown captures the effects of all types of maneuvers. Calculating these measures over the total (origin/destination) trip ensures that the net effect of the maneuvers on flight time and distance is accurately measured. Other flight efficiency measures identified in literature include the number of maneuvers issued to an aircraft and the time spent at or near the flight’s requested altitude.

Although research has not addressed the effects of decision aids in the free routing environment, studies have been conducted to investigate the effect of unstructured traffic on controller tasks and performance [3][4]. This research indicates that unstructured routings increase controller workload. This research further suggests that free routing reduces lead times for controller decision making and the efficiency of the controller’s conflict resolutions.

### *B. Overview of the Experimental Approach*

The human-in-the-loop simulation study described in this report used an experimental design to control for the effects of as many extraneous variable as was possible and isolate the effects of two variables, the conflict probe capability and traffic conditions. Test scenarios were built from samples of recorded traffic in the Indianapolis airspace with matched traffic samples created by assigning the same flights to direct routes. Higher demand scenarios were created by adding flights to the direct route samples. Data were collected under high fidelity simulation conditions with URET-trained, airspace-qualified controllers, comparing performance with and without the conflict probe. Outcome measures representing user and operator impacts were defined. User outcomes included measures of flight efficiency. Operator outcomes included measures of safety, acceptability, and workload. Based on previous research, multiple indicators of the hypothesized outcomes were included to assess their sensitivity to the experimental treatments and to examine relations among measures.

## II. EXPERIMENT METHOD

### *A. Description of Conflict Probe Capability*

The basic roles and responsibilities of the Radar (R) and Data (D) controller team are not expected to change with the implementation of the conflict probe, but it is expected to improve the situation awareness and the ability of the D

controller to support sector operations. Instead of using the paper flight strips to detect potential conflicts between aircraft that may not yet be in the sector, the D controller will use notifications displayed by the conflict probe problem detection function. Conflict notifications are presented to the controller when an event driven conflict probe checks active aircraft flight plans and detects a problem in or near the boundary of the controller's sector. As shown in Figure 1, the URET presents conflict information in the context of a textual aircraft list that contains flight identifiers and flight plan information for all active and inbound sector flights. Controller notifications are color-coded into two priorities (red and yellow) indicating predicted minimum separation distance.

*Figure 1. Example of URET aircraft list display*

Interactive conflict resolution is facilitated by visualization and trial planning aids that synthesize trajectory data and allow the controller to see the effect of proposed resolutions in the context of the current and future traffic situation. The controller may select a conflict from the aircraft list for closer analysis and display current and trial plan route trajectories on a graphic plan display. As shown in Figure 2, the graphic display depicts the aircraft's current trajectory and the trajectory of the conflicting aircraft with the area of violation highlighted. On the actual display, the area of violation is shown in red or yellow. The graphic display may also be manipulated by the controller to dynamically change the display time, depicting aircraft movement from current to future trajectory positions.

*Figure 2. Example of URET graphic display*

The trial planning function allows the controller to specify altitude, speed, route, or a combination of amendments as trial plans. URET uses simple operations to specify trial plans, such as selecting amendment parameters from menus and drawing reroutes on the graphic display by adding or deleting waypoints. When the controller requests a trial plan, feedback is provided, indicating whether the trial plan creates another conflict. An automated coordination function enables the D controller to transmit planned trial amendments to other sectors for approval and implementation.

Two important functions of the FFP1 conflict probe capability were not available in URET at the time of this study. The first was a two-way interface that will allow the controller to transfer a trial plan to the HCS when it is issued as a clearance to the pilot. At the time of this study, the controller was required to reenter a trial plan as a flight plan amendment to the HCS after a clearance was issued. The second capability that was not available in URET for this study was the automated coordination capability. Without this capability, the controller was required to use the voice communication system to relay trial plans to surrounding sectors for approval and implementation.

## *B. Sector and Traffic Scenario Descriptions*

The experiment scenarios were based on recorded flights from the Falmouth high altitude sector at (ZID). Three types of scenarios were developed for the experiment: (1) baseline structured traffic, (2) unstructured traffic, and (3) unstructured traffic with increased volume. Two alternate forms of each type of scenario were created. The baseline scenarios consisted of two equivalent 30-minute time samples of recorded air traffic. The two unstructured scenarios were composed of the same sets of flights but the aircraft were placed on direct routes between the origin and destination. Prior to testing, the baseline and unstructured scenarios were analyzed and matched in terms of peak aircraft and conflict counts. These scenarios had peak aircraft counts ranging between 10 and 12 and conflict counts ranging between 7 and 9. Finally, a third pair of higher volume unstructured traffic scenarios was constructed by adding flights to the two unstructured traffic samples to increase the traffic by 25%. The high volume scenarios had peak aircraft counts of 18 and conflict counts of 14.

## *C. Test Facilities*

The ZID DYSIM laboratory was used for the experiment. The DYSIM laboratory contains multiple suites of PVD controller workstations, a full voice communications system, and flight strip printers. The R- and D-side stations are identical to the operational en route control area. The URET workstation comprises a 17-inch color display, trackball, and keyboard that was placed at the D-side station of the experimental sector. When the URET workstation was used, it took the place of most of the strip bays at the D position. When URET was not in use, it was slid back into the console and the strip bays were replaced. Figure 3 shows the configuration of equipment that was used in this experiment.

*Figure 3. ZID DYSIM laboratory configuration*

The HCS operated in the DYSIM mode which is typically used for controller training. In this mode, the HCS generates simulated radar targets from the scenario and allows a simulated operator position to fly the aircraft using keyboard entries to input instructions received via voice communications. The HCS collected data on aircraft position, time, and event variables. Throughout the experiment, the URET system was running for data collection purposes even when it was not being used by the controller. It collected data on predicted aircraft conflicts, distance and time flown, and controller interactions with the conflict probe capabilities.

## *D. Participants*

Twelve full performance level controllers participated in six teams. Participation was limited to controllers who were qualified to work the simulated airspace and trained in the

use of the URET capabilities. The participants' ATC experience ranged between 8 and 21 years with an average of 11 years.

Participants' experience with URET ranged between 1 hour and 24 months with an average of 4 months. During the experiment, the most experienced controller was assigned to the D side; the average D-controller URET experience was 6.8 months. Participants also rated their URET experience on a seven-point scale from (1) not very experienced to (7) extremely experienced (mean experience rating = 3).

#### *E. Design*

To address the effects of conflict probe and traffic conditions, the experiment used a within subjects design. Both independent variables—automated conflict probe and traffic condition—were manipulated within subjects. The conflict probe variable was defined by two levels—on or off. Traffic condition was defined by three levels—structured, unstructured, and high volume unstructured. Combining these two independent variables resulted in six test conditions.

The twelve participants were divided into six controller teams. Each of the teams completed six test sessions, one for each of the test conditions. The order of presentation of test conditions was randomized. Alternate forms of the scenario types were presented and counterbalanced to control for practice and learning effects.

#### *F. Training*

During the experiment, the participants were asked to control sector traffic as they normally would and to apply these operational procedures to the traffic situations encountered in the test scenarios. Under all traffic conditions, the controller was required to issue standard altitude clearances for arrival traffic. Under the structured traffic condition, the controller was also required to issue standard routing restrictions for all traffic.

Participants were told that their primary task during experiment was to maintain traffic separation. During the test sessions, participants were asked to think aloud while working the traffic scenarios and report any conflict situations that would require control actions, as soon as they recognized them. They were instructed that while reporting conflicts is important, they should give first priority to maintaining separation. Both R and D controllers were asked to report conflicts and describe what actions, if any, they were taking to resolve them. Hands-on practice in the DYSIM laboratory used a separate training scenario and allowed participants and observers to practice reporting and recording conflict information.

#### *G. Procedure*

After training, six test sessions were conducted. These sessions yielded the experimental data for the study. Participants maintained their assigned position throughout the tests. During testing, each team was accompanied by an operational expert from the facility that served as supervisor, evaluating the team's performance during the test runs. In addition, two observers sat next to each controller and recorded verbal reports on conflicts.

A test session consisted of a 30-minute traffic simulation run followed by participant and supervisor ratings of the simulation. Each participating controller rated the perceived workload and operational acceptability of the run. Supervisors rated the controller team's performance and assessed the overall safety of the run. For each conflict reported during the simulation, observers recorded the time it was detected, the aircraft involved, the type of resolution maneuver issued, and the time that the resolution was issued. After all six test sessions, participants completed a questionnaire and a debriefing discussion was conducted.

#### *H. Dependent Measures*

**Aircraft Separation Violations:** Aircraft pairs with violations of the separation minima were identified and tabulated for each run along with the time of the violation and the closest approach distance

**Supervisor Safety Assessments:** Operational assessment of any potentially unsafe conditions during a test run were made on a 4-point scale from (1) operations were unsafe, and unacceptable to (4) the margin of safety was higher than normal for this sector.

**Controller Operational Acceptability Rating:** The Controller Acceptability Rating Scale (CARS) was used to measure how effectively the controller and the automated system performed together in controlling traffic [10]. The CARS describes the joint controller-automation system performance in terms of four rating levels: controllability, tolerability, satisfaction and acceptability. The R and D controllers located the performance within one of these levels and then assigned a numeric rating indicating the degree of problems or deficiencies experienced in the simulation.

**Controller Workload:** Subjective workload was assessed using the multidimensional NASA Task Load Index (TLX) [6]. An overall workload score and six subscale scores were computed for the R and D positions.

Handoff acceptance latency was used as an objective workload measure with longer latencies indicating higher workload[16]. In this experiment, an automatic handoff offer was initiated when an aircraft was a constant time from the boundary of the experimental sector. Handoff accept latency was measured as the elapsed time from the handoff offer to the controller's entry to accept the handoff.

**Delta Time and Distance:** Calculations of flying time and distance used the URET trajectory estimate of origin to destination flying time and distance traveled. Based on a consistent problem start time, delta time<sup>1</sup> and distance was defined as the difference between the actual trajectory time/distance flown in the test scenario and the initial URET trajectory time/distance calculated for the aircraft at the start of the scenario. A delta distance and an absolute delta distance were computed for each aircraft

**Number of maneuvers per aircraft:** Observers logs documented maneuvers issued by the R controller and the HCS recorded the number of aircraft entering the sector on each run. The number of maneuvers issued per aircraft equaled the number of maneuvers issued divided by the number of aircraft entering the sector.

**Conflict Detection Lead Time:** Conflict detection lead-time was measured as the time in minutes and seconds from the conflict detection time recorded in the observer log to the predicted violation time recorded in the URET conflict report.

**Conflict Resolution Lead Time:** Conflict resolution lead-time was measured as the time in minutes and seconds from the conflict resolution time recorded in the observer log to the predicted violation time recorded in the URET conflict report

**Type of Resolution Maneuvers:** Based on the log data, maneuvers were tabulated by the following categories (1) altitude, (2) vector, (3) route amendment, (4) speed, and (5) multiple maneuvers.

**D Controller Activity:** D controller conflict resolution activity recorded in the observer log was tabulated in terms of the number of actions taken to coordinate conflict resolutions with (1) the R controller and (2) other sector controllers.

### III. RESULTS AND DISCUSSION

The statistical analyses discussed throughout this section included two-way repeated measures analyses of variance (ANOVAs) for the interval level measures and paired sample Students T tests for follow-on comparisons. Nonparametric Friedman tests were conducted for the ordinal level measures and follow-on comparisons used Wilcoxon matched pairs signed ranks tests.

---

<sup>1</sup>A measurement problem was encountered in the analysis of time flown and reliable estimates of delta time flown could not be calculated. A random parameter used in the DYSIM to model flight profiles caused flight times to vary randomly with each test run and precluded repeatable measurements of total (origin/destination) time flown.

#### A. Safety and Acceptability

It can be argued that any improvements in flight efficiency as result of new procedures or decision aids have the possibility of being safety compromising. This study found no evidence of safety problems under any of the test conditions. No separation violations occurred under any of the test conditions. In all 36 simulation runs, supervisors' safety assessments indicated that ATC performance was typical and acceptable or higher than normal for the sector. Nor were there any significant differences in supervisor safety ratings as a function of traffic or conflict probe condition. Although a Friedman test did not show any significant differences in supervisor safety ratings between test conditions, conflict probe tended to be associated with higher safety ratings in the unstructured and high volume unstructured traffic conditions.

The CARS scores were analyzed separately for R and D controllers using a Friedman test. The results indicate that conflict probe significantly improved the operational acceptability of the unstructured traffic conditions. However, this effect was observed only for the D controller. Figures 4 and 5 show the CARS scores for the R and D controllers, respectively.

*Figure 4. Mean R controller CARS scores as a function of conflict probe and traffic condition*

*Figure 5. Mean D controller CARS scores as a function of conflict probe and traffic condition*

As shown in Figure 4, the R controller ratings reflect an interaction of traffic condition and automation. The Friedman test indicated a significant difference between the test conditions,  $\chi^2(5)=11.6$ ,  $p \leq .04$ . Comparing the trendlines across traffic conditions indicates that the markedly reduced acceptability with conflict probe that occurred in the structured condition was reduced and then disappeared in the unstructured and high volume unstructured conditions, respectively. Observations and controller feedback during the simulation suggest that this pattern of results may reflect a problem of reduced R controller access to flight data when the conflict probe was in use. During the conflict probe test conditions, the D controller frequently used the graphic display to analyze conflicts. When the graphic display was in use, it covered the aircraft list, concealing flight data. This caused the R controller to invoke the flight plan readout function in order to obtain the flight data. R controllers needed to refer to flight plan data more in the structured traffic conditions because the aircraft's route was more complex and the controllers were expected to issue standard procedural routings. This may account the reduced level of acceptability.

As shown in Figure 5, D controller ratings of operational acceptability reflected the opposite pattern of results. Again, the Friedman test indicated a significant difference

in the test conditions,  $\chi^2(5)=10.69, p \leq .05$ . Comparing the trendlines across traffic conditions indicates that, while acceptability was essentially equivalent in structured conditions, there was a continuing drop in acceptability under the structured and high volume unstructured conditions without conflict probe. In fact, without the conflict probe, system performance approached an unacceptable rating in the high volume unstructured condition.

Taken overall, the evidence on operational acceptability shows that the conflict probe produced clearly improved acceptability under the unstructured and high volume unstructured conditions. The R controller acceptability ratings indicate a human computer interface (HCI) problem that may affect tools use. The safety data indicate that while conflict probe is not required for safe operations in the unstructured traffic environment, the conflict probe capability may improve the safety margin in an unstructured traffic environment.

### B. Controller Workload

An overall TLX score and six subscale scores were analyzed separately for R and D controllers. In general, TLX scores fell below 50 on the 100 point scale, indicating that light to moderate workloads were experienced under all test conditions (see Figure 6). The statistical analyses showed that traffic condition had a significant effect on R and D controller workload,  $F(2,4)=19.02$  and  $15.18, p \leq .01$  for R and D controllers, respectively. The main effect for conflict probe was not significant. As shown in the figure, R and D controllers differed in terms of how they rated the workload experienced under different traffic conditions.

*Figure 6. Mean TLX subjective workload for R and D controllers as a function of conflict probe and traffic conditions*

Although not significant, conflict probe by traffic condition interaction trends suggest that conflict probe reduced controller workload under the high volume unstructured traffic condition, particularly for the D controller, whereas conflict probe increased workload under structured traffic conditions, particularly for the R controller. The marked increase in R controller workload ratings under the structured traffic condition with conflict probe may be another indicator of the flight data access problem that was noted in the previous section.

Five of the six TLX subscales revealed significant effects of traffic condition on components of controller workload. For both R and D controllers, scores on three subscales (physical demand, temporal demand, and frustration) were not affected by changes in traffic structure but increased with traffic load. Consistent with the overall TLX scores, the results on mental demand and effort indicate that the R controller ratings were influenced by traffic structure while

the D controllers ratings were influenced exclusively by traffic load.

The ANOVA for handoff acceptance latency revealed a marginally significant effect for traffic condition,  $F(2,4)=5.45, p \leq .07$ . Mean handoff acceptance latencies for the six test conditions are shown in Figure 7. Handoff latency increased from the structured condition to the unstructured condition and increased again from the unstructured to the high volume unstructured condition. These data also suggest a conflict probe by traffic condition interaction. Although not significant, the direction of the trend was consistent with the subjective workload results. Conflict probe increased R controller workload in the structured condition whereas conflict probe decreased R controller workload in the unstructured and high volume unstructured conditions. Again, this increase may be indicative of a flight data access problem.

*Figure 7. Mean handoff acceptance latency as a function of conflict probe and traffic conditions*

While the overall pattern of workload impacts is complex, the results appear to support these findings.

- Changes in traffic conditions representing the introduction of unstructured routings and higher traffic loads significantly affected controller workload. The pattern of workload impacts differed for R and D controllers indicating that changes in traffic structure are more likely to affect the R controller's tactical role while changes in traffic load are more likely to affect the D controller's strategic role.
- There appears to be an interaction of traffic condition with conflict probe. In contrast to previous research [8] [7], both the subjective and objective workload results of this study indicate that conflict probe automation tended to reduce workload under the high volume unstructured traffic condition.
- As in previous research [8] [7], the subjective and objective measures tended to disassociate. The objective workload measure indicated increased controller workload under unstructured traffic conditions while the subjective measure indicated reduced workload under unstructured traffic conditions.
- The results also indicate that the URET HCI tested in the present study impeded the R controller's access to flight data and increased R controller workload.

### C. Flight Efficiency

By far, the most important benefits associated with unstructured routings are realized when the user's preferred route is filed and accepted by the ATC system. However, there is also a collateral in-flight benefit, especially for air carrier aircraft, when ATC interventions are minimized and the predictability of the flight schedule is maintained. Therefore, for this study, efficiency improvements resulting from conflict probe were hypothesized to depend on (1) the ability of the controllers to manage the unstructured traffic, and (2) the efficiency of the resolutions generated by the controllers.

Data relative to the first hypothesis were discussed in the previous section and show that controllers were able to manage the unstructured traffic conditions presented in the simulation, even without conflict probe. In terms of the second hypothesis, two measures indicating the efficiency of conflict resolutions were analyzed. In the structured traffic condition, it was possible for the controller to resolve conflicts by giving the aircraft a shortcut and clearing it direct to a downstream fix. Consequently, it was arguable whether conformity with the aircraft's schedule and planned arrival time was generally more beneficial than distance savings. Therefore both the increase or decrease in distance flown and the magnitude of the deviation from the planned route were analyzed. For both delta and absolute delta distance, there was a significant main effect of traffic condition,  $F(2,4)=28.70$ ,  $p\leq.004$ . This results confirms that a distance measure of controller interventions was valid only in the structured traffic condition. There was no significant main effect for conflict probe and no interaction of conflict and traffic condition.

Conflict resolution in the structured traffic condition changed the aircraft's horizontal path by about 1 nautical mile. The data on delta distance show that conflict resolutions shortened the aircraft's paths. Although not significant, the average path savings were slightly greater without conflict probe. Conversely, the absolute delta distances indicate that the magnitude of the deviation from the flight plan due to conflict resolution was slightly smaller with the conflict probe. Again, this difference was not significant.

The study results on the number of maneuvers issued per aircraft, shown in Figure 8, suggest two trends: a reduced rate of maneuvers in the unstructured and high volume unstructured conditions and a reduced rate of maneuvers with conflict probe. The ANOVA indicated a marginally significant main effect for traffic condition,  $F(2,4)=5.06$ ,  $p\leq.08$ . The main effect for conflict probe was not significant.

*Figure 8. Mean number of maneuvers per aircraft as a function of conflict probe and traffic condition*

In sum, the statistical analysis of flight efficiency produced inconclusive results. Although not significant, the study results on the number of maneuvers issued are consistent

with previous research [9] and suggest increased flight efficiency with decision aids. Results on distance flown did not reproduce path savings of the magnitude reported in previous field research [11]; however, they are consistent with results of simulation research [17]. There were many differences between this study and [11] in terms of the context and methodology. Since [11] was a field test, it was not possible to achieve the degree of experimental control present in this study. A major difference between the studies was that in [11], the controllers who were actively controlling the traffic did not use the conflict probe tools. In the field test results, inferences about flight path and cost savings are based on selected conflict situations and on solutions developed by a dedicated test controller who interacted with the conflict probe tools.

#### *D. Controller Task Performance*

Controller teams operated more strategically with conflict probe, detecting and resolving conflicts earlier. Data on conflict detection and resolution lead times are shown in Figure 9. In general, the data show that controllers detected and reported conflicts about 6 minutes in advance of the predicted conflict start time and issued a resolution maneuver about 5 minutes in advance of the predicted conflict start time. These results are consistent with the results of a previous analysis of recorded data from ZID, which calculated URET warning time and conflict resolution time [2]. This analysis further revealed that controllers detected and resolved conflicts about 1 minute earlier with conflict probe than they did without the capability. The main effects of conflict probe on conflict detection and resolution lead-time were marginally significant,  $F(1,5)=5.07$ ,  $p\leq.07$  and  $F(1,5)=2.45$ ,  $p\leq.10$ , for conflict detection and conflict resolution, respectively. Although not significant, a conflict probe by traffic conditions interaction trend suggests that, compared to the structured condition, controllers resolved conflicts earlier under the unstructured and high volume unstructured conditions with conflict probe.

*Figure 9. Mean conflict detection lead time and conflict resolution lead time as a function of conflict probe and traffic condition*

The statistical analysis of resolution strategies examined the distributions of vertical and lateral maneuvers between test conditions (see Figure 10). The significant main effect of traffic condition is evident in the increased level of altitude resolution maneuvers in the unstructured and high volume unstructured conditions as compared to the structured conditions,  $F(2,4)=6.30$ ,  $p\leq.05$ . The marginally significant conflict probe by traffic condition interaction is depicted primarily by the marked increase in the use of altitude resolutions that occurred with conflict probe in the structured traffic condition,  $F(2,4)=5.20$ ,  $p\leq.07$ . Controller feedback during the simulation suggested that lateral maneuver options were constrained in the unstructured conditions by the random locations of surrounding traffic

and by the obvious negative impact of the lateral maneuvers on the horizontal path of the flight.

*Figure 10. Types of maneuvers issued in six test conditions*

D controllers were more proactive with the conflict probe, coordinating resolutions with their R controller and the surrounding sectors (see Figure. 11). Results of Friedman tests showed significant difference between the test conditions on both coordination measures,  $\chi^2(5) = 15.62$  and  $10.97$ ,  $p \leq .05$  for other sectors and R controller, respectively. As expected, follow-on analyses revealed that D controllers coordinated significantly more often when they operated with conflict probe, with their R controller ( $z = -2.20$  and  $-1.91$ ,  $p \leq .05$  for other sectors and R controller, respectively).

*Figure. 11. Number of D controller coordinations with other sectors and R controller as a function of conflict probe*

The findings on controller task performance and strategies strongly support a shift toward more strategic ATC with conflict probe as well as a significant increase in the contribution of the D controller to the primary sector tasks. This study did not find any evidence of the shift toward more tactical conflict detection and resolution under unstructured traffic conditions that was suggested by previous research [3]. However, it is interesting to note that the shift toward more strategic resolution of conflicts with conflict probe was most pronounced under the unstructured and high volume unstructured conditions. Further research, with the FFP1 URET capability, including automated coordination, and using a multi-sector simulation environment, is needed to assess the impact of strategic conflict resolution on inter-sector coordination and sector team responsibilities.

#### *E. Assessment of the Methodology*

The results obtained in this study support some remarks about the validity, efficiency and utility of the approach. Two ways to assess the methodology are to examine (1) how well the measures fit the operational concept and hypotheses regarding the effects of the experimental variables, and (2) how well the measures of the same outcome “go together”.

Data from the experiment indicate that the measures of operator outcomes--conflict detection lead time, conflict resolution lead time, acceptability and workload--exhibited the expected relationship to the experimental variables. Since previous research [8] [7] showed that subjective workload measures may be sensitive indicators of controller acceptability, a preliminary analysis was conducted to assess whether the CARS correlates with controller workload and the extent of the overlap between the two measures. Although the scale was based on an existing flight deck measurement instrument, the CARS was

developed recently and its validity and reliability is not well established. A regression analysis was run using the six TLX subscales as predictors and the CARS as the criterion. Overall, there was a significant multiple correlation between the CARS and the TLX subscales ( $R = .45$ ,  $p \leq .01$ ). Three of the subscales accounted for a significant proportion of the variance in the CARS: frustration, mental demand, and temporal demand. A significant negative relationship between frustration and CARS indicated that high acceptability was associated with low frustration. High acceptability was also associated with low mental workload and with high temporal demand. The effort, physical demand and (controller) performance scales did not make significant contributions to predicting the CARS.

These preliminary results tend to indicate that the CARS is accurately measuring selected facets of workload that influence the controller's use of and satisfaction with the ATC decision aids. They also suggest that the CARS does not completely overlap with the TLX. The CARS is designed to provide a global measure indicating the controller's assessment of whether the experienced levels of workload, safety, and performance are appropriate compatible with operational use. This kind of data is economical to collect and should have general application in a field evaluation. Further validation research should be conducted on the CARS to (1) include a wider range of operational contexts and traffic levels, representing under- and overload, (2) examine the agreement among controller ratings, and (3) simultaneously validate the relationship between CARS and factors over and above workload, such as decision aiding.

Data from the experiment on user outcomes measures--delta time and distance, number of maneuvers per aircraft--were inconclusive and indicate that further work is needed to establish valid measures of user benefits. The inability to estimate the deltas in time flown as a global metric was a serious obstacle to quantifying user benefits and to assessing the validity of the other flight efficiency measures. Further validation of flight efficiency measures using an experimental approach would be an economical strategy to include in FFP1 benefits assessments before undertaking continuous data collection over long periods in an uncontrolled field evaluation. A controlled study of the experimental variables, including a comparison with existing procedures as a baseline, is needed to support a quantitative assessment of user benefits. The opportunity for comparisons with baseline data is much more limited in the field. The field evaluation should focus more on verification of previously established experimental results [15].

#### IV. CONCLUSIONS

This study is the first conducted to look at the benefits of conflict detection and resolution aids in the context of a free routing environment. Empirical evidence from this study leads to the general conclusion that the conflict probe has favorable impacts in terms of safety, acceptability, and

workload in the free routing environment. Study constraints prevented a firm conclusion regarding the impact of the conflict probe on flight efficiency and user benefits.

Despite a number of study limitations, the experimental approach provided a suitable means to rigorously investigate the quality and characteristics of proposed outcome measures of the FFP1 URET. Although further analysis and empirical data are needed to validate the CARS, preliminary findings showed that the CARS appears to be a promising measure of operational acceptability. A similar approach should be undertaken to bring empirical evidence to bear on the validation of a flight efficiency measure or index of measures.

## REFERENCES

- [1] D. J. Brudnicki, K. C. Bowen, S. E. Gordon, and A. L. McFarland, "User Request Evaluation Tool (URET) Evaluation Report," MTR-98W0000045, The MITRE Corporation, McLean, VA, 1998.
- [2] D. J. Brudnicki and A. L. McFarland, "User Request Evaluation Tool (URET) Conflict Probe Performance and Benefits Assessment," MP97W0000112, The MITRE Corporation, McLean, VA, 1997.
- [3] L. S. Carlson, L. R. Rhodes, and M. Cullen, "Effects of Unstructured Routes on En Route Controllers' Work Activities and Operational Environment," MTR-96W0000019, The MITRE Corporation, McLean, VA, 1996.
- [4] M. R. Endsley, R. Mogford, K. R. Allendoerfer, M. D. Snyder, and E. S. Stein, "Effects of Free Flight Conditions on Controller Performance, Workload, and Situation Awareness," DOT/FAA/CT-TN97/12, The National Technical Information Service, Springfield, VA, 1997.
- [5] L. Goldmuntz, J. T. Kefaliotis, D. Weathers, L. A. Kleiman, R. A. Rucker, and L. Schuchman, "Automated En Route Air Traffic Control (AERA) Concept," MU61, The MITRE Corporation, McLean, VA, 1981.
- [6] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," In P. A. Hancock and N. Meshati (Eds.) *Human Mental Workload*, Elsevier Science Publishers B.V.(North-Holland), New York, NY, pp. 139-183, 1988.
- [7] B. G. Hilburn, P. G. Jorna, E. A. Byrne, and R. Parasuraman, "The Effect of Adaptive Air Traffic Control Decision Aiding on Controller Mental Workload," In M. Mouloua and J. M. Koonce (Eds.) *Human-Automation Interaction: Research and Practice*, Lawrence Erlbaum Associates, Mahwah, N J, pp. 84-91, 1997a.
- [8] B. G. Hilburn, P. G. Jorna, and R. Parasuraman, "The Effect of Advanced ATC Strategic Decision Aiding Automation on Mental Workload and Monitoring Performance: An Empirical Investigation in Simulated Dutch Airspace. In R. S. Jensen and L. A. Rakovan" (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, April 1995.
- [9] P. Krois and A. Marsden, "FAA ODID IV: En Route Baseline Comparison Simulation Final Report," Crown Doc. No. C012-001-008, Crown Communications, Inc., Washington, DC, 1997.
- [10] K. K. Lee and T. J. Davis, "The Development of the Final Approach Spacing Tool (FAST): A Cooperative Controller-Engineer Design Approach." In D. Brandt and T. Martin (Eds.) *Automated Systems Based on Human Skill, A Proceeding Volume from the 5th IFAC Symposium* (pp. 219-226), Berlin, Germany, Pergamon, Oxford, UK, 1995.
- [11] B. D. McNally, R. E. Bach, and W. Chan, "Field Test Evaluation of the CTAS Conflict Prediction and Trial Planning Capability," Paper presented at the American Institute of Aeronautics and Astronautic Guidance, Navigation, and Control Conference, Boston, MA, August 1998.
- [12] F. Rainback, D. Hudson, and A. Lucas, "PD/1+trial report" (DOC-97-70-10), EUROCONTROL Agency, Brussels, Belgium, 1997.
- [13] RTCA, "Final Report of RTCA Task Force 3: Free Flight Implementation," RTCA, Inc., Washington, DC, 1995.
- [14] RTCA, "Government/Industry Operational Concept for the Evolution of Free Flight: Addendum 1, Free Flight Phase 1 Limited Deployment of Select Capabilities," RTCA, Inc., Washington, DC, 1998.
- [15] F. Schick, "Methods and Measurements for the Evaluation of ATM Tools in Real-Time Simulations and Field Tests," *Proceedings of the 2<sup>nd</sup> USA Europe Air Traffic Management R&D Seminar*, The Federal Aviation Administration Office of Aviation Research, Washington, DC, 1998.
- [16] N. Tofukuji, "An En Route ATC Simulation Experiment for Sector Capacity Estimation," *IEEE Transactions on Control System Technology*, 1(3), 138-143, 1993.
- [17] R. Whitaker, D. Marsh, and H. Schroter, "PD/1 Final Report," DOC-96-70-24, EUROCONTROL Agency, Brussels, Belgium, 1997.