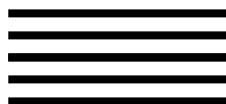


PROGRAMME FOR  
HARMONISED AIR TRAFFIC  
MANAGEMENT RESEARCH  
IN EUROCONTROL



**DOC 96-70-24**



EUROPEAN ORGANISATION FOR THE SAFETY OF AIR NAVIGATION, EUROCONTROL



**PD/1 FINAL REPORT**  
**Annex C**  
**Results**

PHARE/NATS/PD1-10.2/SSR;1.1



**EUROCONTROL**

EUROCONTROL  
96 rue de la Fusée  
B-1130 BRUXELLES

Prepared by: D Hudson  
S Kay  
D Marsh

Date: January 1997

## **REVISION HISTORY**

June 1996	Draft version 0
July 1996	Draft version 1, incorporating DRA and NATS review comments
September 1996 PD3CG	Draft version 2, with minor changes following review by
October 1996	Version 1.0, with minor changes, approved by the PCC
January 1997	Version 1.1, issued as a final document in Volume 2, approved by the PHARE Management Board

## LIST OF CONTENTS

1. INTRODUCTION .....	8
1.1. GENERAL.....	8
1.2. NOTES ON METHOD.....	8
2. WORKLOAD.....	10
2.1. INTRODUCTION.....	10
2.2. SUBJECTIVE MEASURES .....	11
2.2.1. ISA scores.....	11
2.2.2. TLX scores.....	19
2.3. OBJECTIVE MEASURES .....	30
2.3.1. Number of ATC instructions issued .....	30
2.3.2. R/T communications .....	31
2.3.3. Telephone communications.....	35
2.4. SUMMARY AND DISCUSSION OF RESULTS.....	36
2.4.1. Introduction.....	36
2.4.2. Workload differences between ORG 0 and ORG 1.....	36
2.4.3. Workload differences between ORG 1 and ORG 2 (70%).....	37
2.4.4. Workload differences between ORG 0 and ORG 2 (70%).....	37
3. CAPACITY.....	39
3.1. INTRODUCTION.....	39
3.2. AVERAGE AIRCRAFT SEPARATIONS .....	41
3.3. MINIMUM SEPARATION INFRINGEMENTS .....	41
3.3.1. Number of minimum safe separation infringements per run.....	41
3.3.2. Duration of minimum safe separation infringements .....	44
3.3.3. ‘Serious’ separation infringements .....	45
3.4. PERCENTAGE OF TIME ‘PICTURE’ MAINTAINED BY CONTROLLERS....	46
3.5. STCAs.....	47
3.6. DISCUSSION OF RESULTS .....	48
4. QUALITY OF SERVICE.....	50
4.1. INTRODUCTION.....	50
4.2. PERCENTAGE OF TIME AIRCRAFT AT REQUESTED CRUISE FLIGHT LEVEL.....	51
4.3. AVERAGE TIME SPENT BY AIRCRAFT IN SECTOR.....	54
4.4. SUMMARY AND DISCUSSION OF RESULTS.....	55
5. CONCLUSIONS.....	57
6. ABBREVIATIONS.....	59

## LIST OF FIGURES

Figure 2-1 - Comparison between ORGs and controller role of mean ISA scores .....	16
Figure 2-2 - Comparison between ORGs and controller role of the overall TLX scores .....	21
Figure 2-3 - Comparison between ORGs of the average scores for the TLX factors recorded by the Tactical Controllers .....	24
Figure 2-4 - Comparison between ORGs of the average scores for the TLX factors recorded by Planning Controllers .....	27
Figure 2-5 - Comparison between ORGs of the average total number of ATC instructions issued per run per sector.....	31
Figure 2-6 - Comparison between ORGs and sectors of corrected number of R/T calls .....	33
Figure 2-7 - Comparison between ORGs and sectors of the percentage of time spent using R/T per Tactical Controller per run .....	34
Figure 2-8 - Duration of phone calls .....	35
Figure 3-1 - Hypothetical example of the relationship between ATC system, controller workload and traffic level.....	40
Figure 3-2 - Distribution of the number of safe separation infringements (sector 10) .....	42
Figure 3-3 - Distribution of the number of safe separation infringements (sector 11) .....	43
Figure 3-4 - Difference in number of infringements by each controller in each sample, comparing ORG 0 to ORG 2 (70%).....	44
Figure 3-5 - Distribution of the duration of longest infringement per run (sector 10).....	45
Figure 3-6 - Distribution of the duration of longest infringement per run (sector 11).....	45
Figure 3-7 - Percentage of time the 'picture' is maintained by Tactical Controllers .....	47
Figure 3-8 - Percentage of time the 'picture' is maintained by Planner Controllers .....	47
Figure 3-9 - Comparison between ORGs and sectors of the number of STCAs per run.....	48
Figure 4-1 - Comparison between ORGs of the average percentage of time aircraft were at their requested cruising flight level.....	53
Figure 4-2 - Comparison between ORGs of the average percentage of time aircraft were within 2000' of their requested cruising flight level.....	53
Figure 4-3 - Comparison between ORGs of the average time spent in sector by aircraft (shown as a percentage change over ORG 0) .....	55

## LIST OF TABLES

Table 2-1 - Results for the comparison of ISA means between ORG 0 and ORG 1.....	12
Table 2-2 - Results for the comparison of ISA means between ORG 0 and ORG 2 (70%) .....	13
Table 2-3 - Results for the comparison of the ISA means between ORG 1 and ORG 2 (70%).....	14
Table 2-4 - Results for the comparison of the ISA means between ORG 2 (30%) and ORG 2 (70%).....	15
Table 2-5 - Summary of mean ISA results .....	16
Table 2-6 - Results for the comparison of the ISA means between controller roles .....	17
Table 2-7 - Results for the comparison of the ISA means between sectors .....	18
Table 2-8 - Results for the overall TLX scores recorded by Planning Controllers .....	19
Table 2-9 - Results for the overall TLX scores recorded by Tactical Controllers .....	20
Table 2-10 - Results for the Effort Expended scores recorded by Tactical Controllers .....	22
Table 2-11 - Results for the Mental Demand scores recorded by Tactical Controllers .....	22
Table 2-12 - Results for the Time Pressure scores recorded by Tactical Controllers .....	22
Table 2-13 - Results for the Physical Demand scores recorded by Tactical Controllers .....	23
Table 2-14 - Results for the Frustration Experienced scores recorded by Tactical Controllers .....	23
Table 2-15 - Results for the Effort Expended scores recorded by Planning Controllers .....	25
Table 2-16 - Results for the Mental Demand scores recorded by Planning Controllers .....	25
Table 2-17 - Results for the Time Pressure scores recorded by Planning Controllers .....	26
Table 2-18 - Results for the Physical Demand scores recorded by Planning Controllers.....	26
Table 2-19 - Results for the Frustration Experienced scores recorded by Planning Controllers .....	27
Table 2-20 - Unweighted individual TLX factors between ORG 0 and ORG 2 (70%) .....	28
Table 2-21 - Comparison between controller roles of overall TLX scores .....	29
Table 2-22 - Comparison between sectors of overall TLX scores .....	29
Table 2-23 - Results for the number of ATC instructions issued.....	30
Table 2-24 - Results for the total corrected number of R/T calls made.....	32
Table 2-25 - Results for the percentage of time spent using R/T per Tactical Controller per run .....	34
Table 3-1 - Number of aircraft per hour used in defining the traffic samples.....	39
Table 3-2 - Results for the average number of minimum safe separation infringements .....	42
Table 3-3 - Results for the average duration of minimum safe separation infringements .....	44
Table 3-4 - Results for the percentage of time the 'picture' is maintained .....	46
Table 4-1 - Results for the percentage of time the aircraft were at their requested cruise flight level.....	52
Table 4-2 - Results for the percentage of time aircraft were within 2000' of their requested cruise flight level.....	52
Table 4-3 - Results for the average aircraft time spent in each sector.....	54

This page is intentionally blank

## 1. INTRODUCTION

### 1.1. GENERAL

This annex describes the results from the statistical analysis of the data recorded during the first PHARE Demonstration trial (PD/1). Note that this is a technical annex, and the terminology used herein is quite specialised and detailed; for a summary of the results, refer to the relevant sections of the main report. The categories of data recorded and the analysis methodology employed are described fully in Annex A: Experimental Design and Methods.

The aims of PD/1 were:

- to determine the effect on controller workload and traffic throughput of the introduction of computer assistance tools from the PHARE Advanced Tools (PATs) programme;
- to determine the effect on controller workload and traffic throughput of the increasing proportion of 4-D FMS aircraft with full duplex datalink;
- to gain a degree of controller approval for the computer assistance tools introduced.

The quantitative analysis reported in this annex addresses the first two aims of the PD/1 trial, the third aim being addressed in Annex D. However, it must be emphasised that the results presented herein cannot be regarded in isolation; a pure statistical analysis of any set of results from a real-time ATC simulation must be considered in conjunction with the comments made by the participating controllers and the observations made by the experimenters during the trial. These comments and observations are reported in Annex F, with the main report containing the summary conclusions of the numerous analyses conducted.

While PD/1's aims concentrate on an examination of controller workload and traffic throughput, these are not the only measures by which the effectiveness of changes to an ATC operational concept can be measured. The statistical analysis, therefore, also considered the quality of the service provided to the airlines - this being an obvious, and extremely important, measure of the success of any ATC system.

The analysis has been conducted as a comparison between the three organisations, or ORGs, described in Annex A. Therefore, the effect of the introduction of the PATs (and the new operational procedures they support) was examined by comparing ORG 1 with the baseline ORG 0. The effect of the introduction of the PATs and 4D FMS aircraft with full duplex datalink (and the new operational procedures they support) was examined by comparing ORG 2 with the baseline ORG 0. In this annex, references made to ORG 2 (30%) or ORG 2 (70%) refer to the introduction of 30% and 70% 4D FMS and datalink equipped aircraft respectively.

### 1.2. NOTES ON METHOD

The analysis methodology has been described in detail in Annex A. However it is important to stress here some significant points.

Firstly, note that the statistical tests of significance were all one-tailed throughout the analysis, except where stated otherwise. The level of significance was 5%. Since so many statistical

tests were carried out during the analysis, it is recommended that the reader be aware of false positives: with a significance level of 5%, it is to be expected that on average 1 test in 20 tests where the null hypothesis was rejected is in fact a false positive. This means that the observed difference did indeed occur by chance and the null hypothesis should not have been rejected.

A different set of four controllers participated in the trials each week, controlling the same set of traffic samples. An investigation was carried out at an early stage of the analysis into whether the measures from different controllers in the same position in different weeks produced similar responses. Hence, the null hypothesis was tested that there were no significant differences between controllers in the same sector, role and on the same traffic sample in different weeks. In this case, the alternative hypothesis was simply that there did exist differences between the responses of different controllers.

The Kruskal-Wallis non-parametric one-way analysis of variance was used to test this hypothesis for both a subjective measure of workload (mean ISA scores) and also for an objective measure of one element of workload (the mean duration of R/T communications). It was found there were significant differences between measures of the controllers' workload for the different sets of controllers. The null hypothesis was rejected with a high degree of certainty (at the 0.1% level of significance for both mean ISA scores and mean duration of R/T communications).

This analysis showed that in general, the response to a given scenario was highly dependent on the controller. This drawback to the statistical analysis was overcome through the use of the Wilcoxon Signed Ranks test. The Signed Ranks Test is only used here to compare measures produced by the same controller, hence eliminating the effects of controller specific responses. In contrast, the Wilcoxon-Mann-Whitney test is more prone to the biases introduced by variations in controller responses, so its use has been limited to those occasions where insufficient data were available to use the Wilcoxon Signed Ranks test, or in cases where matched pairs of data were not available, for example where comparisons were made between controller roles.

## 2. WORKLOAD

### 2.1. INTRODUCTION

The first two objectives of the PD/1 trial both concentrate on determining the effect on controller workload of the changes to the ATC system. The limiting factor in increasing en-route capacity in the well developed parts of Europe is primarily controller workload; so, if the controllers' workload can be reduced, the capacity of the airspace should increase. The aim of the PD/1 operational concept is, therefore, to reduce the controllers' workload to allow them to cope with the higher traffic levels that are predicted for the period 2005 to 2010.

In conducting the statistical analysis, both subjective and objective measures of controller workload (see Annex A) have been analysed. The following set of null hypotheses (referred to as  $H_0$ ) were tested to examine directly the first two aims of PD/1:

- the workload of ORG 0 is the same as that of ORG 1 ( $H_{0.1}$ );
- the workload of ORG 0 is the same as that of ORG 2 with 70% of the aircraft 4-D FMS and datalink-equipped ( $H_{0.2}$ );
- the workload of ORG 1 is the same as that of ORG 2 with 70% of the aircraft 4-D FMS and datalink-equipped ( $H_{0.3}$ );
- the workload of ORG 2 with 30% of the aircraft 4-D FMS and datalink-equipped is the same as that of ORG 2 with 70% of the aircraft 4-D FMS and datalink-equipped ( $H_{0.4}$ ).

In addition to examining the effect of the PATs and the introduction of 4-D FMS and datalink-equipped aircraft over the complete airspace studied in PD/1, their effect on the individual sectors and on the specific controller roles were examined. The following null hypotheses were tested for selected measures:

- the workload of the Tactical Controller is the same as that of the Planning Controller ( $H_{0.5}$ );
- the workload in sector 10 is the same as that in sector 11 ( $H_{0.6}$ ).

The results of these tests are presented in Sections 2.2 and 2.3, and are discussed in Section 2.4. Note that the experimental design, as described in Annex A, does not allow a comparison between ORG 2 (30%) and the other cases examined over *all* the traffic sample volumes, i.e. high, medium and low. Time constraints placed on the PD/1 trial meant that ORG 2 (30%) could not be conducted using the low traffic sample. Therefore, all comparisons between ORG 2 (30%) and the other organisations use only the medium and high traffic samples.

The experimental design was developed to assess the difference between traffic samples with 30% and 70% of the aircraft equipped with 4-D FMS and datalink. However, it must be remembered that ORG 1 is equivalent to ORG 2 (0%), i.e. with no datalink-equipped 4-D FMS aircraft; therefore, when applicable, comparisons may be made between ORG 0, ORG 1, ORG 2 (30%) and ORG 2 (70%): but only using the medium and high traffic samples.

## 2.2. SUBJECTIVE MEASURES

### 2.2.1. ISA scores

ISA scores are a subjective measure of workload, and are described fully in Annex A. Four measures using the recorded ISA scores were examined:

- the mean ISA score per run;
- the peak ISA score per run;
- the percentage of ISA scores greater than or equal to 4 per run;
- the percentage of ISA scores equal to 5 per run.

The mean ISA scores per run of the different ORGs were initially compared across both sectors, both controller roles, and across all three traffic levels together. Those results that were statistically significant at the 5% level were then investigated more deeply, by repeating the comparisons taking each controller role separately, and each sector separately. Finally the ORGs were compared within each individual sector, role, and traffic level. In this manner, it is possible to see whether the results are actually general results or were caused by one particularly extreme result.

The Wilcoxon Signed Ranks (WSR) and Wilcoxon-Mann-Whitney (WMW) tests are standard statistical tests for use with non-parametric data (i.e. data for which we can make no *a priori* assumption about their distribution). Both the WSR and WMW tests rank data points and then compare the sums of ranks. The key difference between the tests is that the data points ranked in WSR are the differences between matched ISA scores (i.e. from the same controller in the same sector and role, and controlling the same volume of traffic). The data points ranked in WMW are the actual ISA scores in the two ORGs being compared.

The WSR test is a more powerful technique than the WMW test (in that, given the same ISA scores, the WSR test can give conclusions at a better degree of certainty than the WMW test). In essence, the comparison of matched pairs of results removes one source of variability between the two ORGs. However, the dependence of the WSR test on comparing ISA scores from matched simulation runs often reduces the number of data that can be used (due to runs which were excluded from the analysis). The WMW test does not need matched ISA scores. A reduction in the number of data points used in a test reduces its ability to detect differences. The impact of all this is that the ISA results presented usually show a better degree of statistical significance for the WSR test than for the WMW test. The occasions where the contrary is true are caused by there being only a limited number of *matched* ISA scores.

Results at the 10% level of significance are included in the tables presented in this and subsequent sections to display trends rather than to imply any kind of sound statistical differences. Although the null hypothesis is accepted for results at the 10% level, the general direction of the trend is shown but enclosed in square brackets to indicate it is not statistically significant.

#### 2.2.1.1. Comparison of mean ISA scores between ORGs

Table 2-1 to Table 2-4 below give the results for the comparisons between ORGs of the mean ISA scores. Table 2-6 to Table 2-7 give the results for the comparisons of mean ISA scores between controller roles and sectors. The tests used were the Wilcoxon Signed Ranks (WSR)

test, and the Wilcoxon-Mann-Whitney (WMW) test - the results for each test being identified in the tables by the font used, normal (WSR) or *italic* (WMW). Figure 2-1 on gives a graphical view of the results of the mean ISA scores.

In Table 2-1 to Table 2-4, and in Table 2-6 to Table 2-7, P signifies the probability of obtaining the observed difference, or a more extreme observation, by chance if the null hypothesis  $H_0$ , were indeed true, where  $H_0$  is the hypothesis that there is no difference between the ORGs. Note that by  $ORG\ a > ORG\ b$ , it is meant that the mean ISA scores for ORG a were **significantly** greater than those for ORG b and this difference is unlikely to have occurred by chance. The hypotheses tested are  $H_{0.1}$  to  $H_{0.4}$  for the comparison between ORGs of the mean ISA scores and  $H_{0.5}$  to  $H_{0.6}$  for the comparison of mean ISA scores between controller roles and sectors respectively. These follow the hypotheses as stated in Section 2.1.

ORGs compared	Sectors	Role	Traffic Volume	P(WSR) <i>P(WMW)</i>	Significant?	Conclusion
ORG 0 vs ORG 1	Both	Both	All	<0.001 <i>&lt;0.001</i>	At 0.1% level <i>At 0.1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Tactical		0.0012 <i>0.0015</i>	At 0.2% level <i>At 0.2% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Planner		<0.001 <i>&lt;0.001</i>	At 0.1% level <i>At 0.1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Both	Low	0.0625 <i>&gt;0.15</i>	At 10% level <i>No</i>	[ORG 1 > ORG 0] <i>Accept <math>H_{0.1}</math></i>
		Tactical		- <i>&gt;0.15</i>	Not enough data <i>No</i>	- <i>Accept <math>H_{0.1}</math></i>
		Planner		- <i>&gt;0.15</i>	Not enough data <i>No</i>	- <i>Accept <math>H_{0.1}</math></i>
		Both	Medium	0.0011 <i>&lt;0.001</i>	At 0.2% level <i>At 0.1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Tactical		0.0790 <i>0.0193</i>	At 10% level <i>At 2% level</i>	[ORG 1 > ORG 0] <i>ORG 1 &gt; ORG 0</i>
		Planner		0.0017 <i>0.0064</i>	At 0.2% level <i>At 1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Both	High	<0.001 <i>&lt;0.001</i>	At 0.1% level <i>At 0.1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Tactical		0.0062 <i>0.0052</i>	At 1% level <i>At 1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Planner		0.0019 <i>0.0021</i>	At 0.2% level <i>At 0.5% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
	10	Planner	All	<0.001 <i>0.011</i>	At 0.1% level <i>At 2% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Tactical		<0.001 <i>&lt;0.001</i>	At 0.1% level <i>At 0.1% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
	11	Planner		<0.001 <i>0.0023</i>	At 0.1% level <i>At 0.5% level</i>	ORG 1 > ORG 0 <i>ORG 1 &gt; ORG 0</i>
		Tactical		>0.15 <i>&gt;0.15</i>	No <i>No</i>	Accept $H_{0.1}$ <i>Accept <math>H_{0.1}</math></i>

Table 2-1 - Results for the comparison of ISA means between ORG 0 and ORG 1

Table 2-1 shows that comparing ORG 0 and ORG 1, there is a significant increase in the perceived workload of both the Planner and Tactical Controller for the high and medium traffic levels. For the low traffic volume samples no difference in workload could be determined with the limited data available.

ORGs compared	Sectors	Role	Traffic Volume	P(WSR) P(WMW)	Significant?	Conclusion
ORG 0 vs ORG 2 (70%)	Both	Both	All	0.011 0.055	At 2% level At 10% level	ORG 2 (70%) > ORG 0 [ORG 2 (70%) > ORG 0]
		Tactical		>0.15 >0.15	No No	Accept $H_{0-2}$ Accept $H_{0-2}$
		Planner		<0.001 0.0026	At 0.1% level At 0.5% level	ORG 2 (70%) > ORG 0 ORG 2 (70%) > ORG 0
		Both	Low	0.098 0.150	At 10% level No	[ORG 0 > ORG 2 (70%)] Accept $H_{0-2}$
		Tactical		0.0527 0.0253	At 10% level At 5% level	[ORG 0 > ORG 2 (70%)] ORG 0 > ORG 2 (70%)
		Planner		>0.15 >0.15	No No	Accept $H_{0-2}$ Accept $H_{0-2}$
		Both	Medium	0.0023 <0.001	At 0.5% level At 0.1% level	ORG 2 (70%) > ORG 0 ORG 2 (70%) > ORG 0
		Tactical		0.0874 >0.15	At 10% level No	[ORG 2 (70%) > ORG 0] Accept $H_{0-2}$
		Planner		0.0020 <0.001	At 0.5% level At 0.1% level	ORG 2 (70%) > ORG 0 ORG 2 (70%) > ORG 0
		Both	High	0.0783 >0.15	At 10% level No	[ORG 2 (70%) > ORG 0] Accept $H_{0-2}$
		Tactical		>0.15 >0.15	No No	Accept $H_{0-2}$ Accept $H_{0-2}$
		Planner		0.0017 0.049	At 0.2% level At 5% level	ORG 2 (70%) > ORG 0 ORG 2 (70%) > ORG 0

Table 2-2 - Results for the comparison of ISA means between ORG 0 and ORG 2 (70%)

The overall impact of moving from ORG 0 to ORG 2 when 70% of the aircraft are equipped with 4-D FMS and datalink (ORG 2 (70%)) is a significant increase in mean ISA. However, a more detailed picture emerges when the two controller roles are considered separately.

The Planning Controller's workload is significantly higher for ORG 2 (70%) for the high and medium traffic levels. For the low traffic volume, no significant change was recorded.

On the other hand, the Tactical Controller's workload shows no significant change between the two organisations. At a more detailed level, this is true at medium and high traffic levels. However, at low traffic levels, the workload actually decreases from ORG 0 to ORG 2 (70%). This small but significant improvement in workload is masked by the increase in the Planner's workload in the combined result for the two roles.

ORGs compared	Sectors	Role	Traffic Volume	P(WSR) <i>P(WMW)</i>	Significant?	Conclusion	
ORG 1 vs ORG 2 (70%)	Both	Both	All	<0.001 <i>0.0019</i>	At 0.1% level <i>At 0.2% level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>	
		Tactical		<0.001 <i>&lt;0.001</i>	At 0.1% level <i>At 0.1% level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>	
		Planner		>0.15 <i>&gt;0.15</i>	No <i>No</i>	Accept $H_{0.3}$ <i>Accept <math>H_{0.3}</math></i>	
		Both	Low	0.0156 <i>0.0103</i>	At 2% level <i>At 2% level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>	
		Tactical		0.0527 <i>0.115</i>	At 10% level <i>No</i>	[ORG 1 > ORG 2 (70%)] <i>Accept <math>H_{0.3}</math></i>	
		Planner		- <i>&gt;0.15</i>	Not enough data <sup>1</sup> <i>No</i>	- <i>Accept <math>H_{0.3}</math></i>	
		Both	Medium	>0.1 <i>&gt;0.15</i>	No <i>No</i>	Accept $H_{0.3}$ <i>Accept <math>H_{0.3}</math></i>	
		Tactical		0.00998 <i>&gt;0.15</i>	At 1% level <i>No</i>	ORG 1 > ORG 2 (70%) <i>Accept <math>H_{0.3}</math></i>	
		Planner		>0.15 <i>0.119</i>	No <i>No</i>	Accept $H_{0.3}$ <i>Accept <math>H_{0.3}</math></i>	
		Both	High	0.0014 <i>0.0021</i>	At 0.2% level <i>At 0.5 % level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>	
		Tactical		<0.001 <i>0.0013</i>	At 0.1% level <i>At 0.2% level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>	
		Planner		>0.15 <i>0.146</i>	No <i>No</i>	Accept $H_{0.3}$ <i>Accept <math>H_{0.3}</math></i>	
		10	Both	Medium and High	0.0055 <i>0.0260</i>	At 1% level <i>At 5% level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>
		11			0.0532 <i>&gt;0.15</i>	At 10% level <i>No</i>	[ORG 1 > ORG 2 (70%)] <i>Accept <math>H_{0.3}</math></i>
		10	Tactical	Medium and High	<0.001 <i>0.0019</i>	At 0.1% level <i>At 0.2% level</i>	ORG 1 > ORG 2 (70%) <i>ORG 1 &gt; ORG 2 (70%)</i>
		11			0.0170 <i>&gt;0.15</i>	At 2% level <i>No</i>	ORG 1 > ORG 2 (70%) <i>Accept <math>H_{0.3}</math></i>
		10	Planner	Medium and High	>0.15 <i>&gt;0.15</i>	No <i>No</i>	Accept $H_{0.3}$ <i>Accept <math>H_{0.3}</math></i>
		11			>0.15 <i>&gt;0.15</i>	No <i>No</i>	Accept $H_{0.3}$ <i>Accept <math>H_{0.3}</math></i>

Table 2-3 - Results for the comparison of the ISA means between ORG 1 and ORG 2 (70%)

The comparison of ORG 1 and ORG 2 (70%) reinforces the conclusions reached by the ORG 0/ORG 1 and ORG 0/ORG 2 (70%) comparisons; the Planner's workload in

<sup>1</sup> This is due to 'unacceptable runs' (see Annex A) and tied observations

ORG 2 (70%) is of the same order as that in ORG 1, whereas the Tactical Controller shows a significant decrease in the workload from ORG 1 to ORG 2 (70%).

ORGs compared	Sectors	Role	Traffic Volume	P(WSR) P(WMW)	Significant?	Conclusion	
ORG 2 (30%)  vs  ORG 2 (70%)	Both	Both	Medium  and  High	0.145 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$	
		Tactical		0.0104 0.129	At 2% level No	ORG 2 (30%) > ORG 2 (70%) Accept $H_{0-4}$	
		Planner		>0.15 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$	
		Both	Medium	>0.15 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$	
		Tactical		>0.15 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$	
		Planner		0.107 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$	
		Both	High	0.0742 >0.15	At 10% level No	[ORG 2 (30%) > ORG 2 (70%)] Accept $H_{0-4}$	
		Tactical		0.037 0.095	At 5% level At 10% level	ORG 2 (30%) > ORG 2 (70%) [ORG 2 (30%) > ORG 2 (70%)]	
		Planner		>0.15 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$	
		10	Both	All	0.0167 >0.15	At 2% level No	ORG 2 (30%) > ORG 2 (70%) Accept $H_{0-4}$
		11	Both	All	>0.15 >0.15	No No	Accept $H_{0-4}$ Accept $H_{0-4}$

Table 2-4 - Results for the comparison of the ISA means between ORG 2 (30%) and ORG 2 (70%)

Finally, when comparing the two levels of aircraft equipped with 4D FMS and datalink in ORG 2, it can be seen that while the Planning Controller's workload is the same in both cases, the Tactical Controller's workload is significantly higher in the 30% case than the 70% case for the high traffic volumes. This implies that the percentage of aircraft equipped with 4D FMS and datalink has a positive impact on the Tactical Controller's perceived workload.

Table 2-5 summarises the mean ISA results, showing the impact of changing organisation on the workload of the two controllers, taking each traffic volume separately. The '=' indicates that there was no statistically significant change in workload for that controller role and traffic volume. The '↑' and '↓' symbols indicate statistically significant increases and decreases in workload respectively, for that change of ORG. The '.' symbol simply indicates that no comparison was possible between those ORGs for that traffic level.

ORGs compared	Planner (low med high)	Tactical (low med high)
From ORG 0 to ORG 1	= ↑ ↑	= ↑ ↑
From ORG 0 to ORG 2 (70%)	= ↑ ↑	↓ = =
From ORG 1 to ORG 2 (70%)	= = =	= ↓ ↓
From ORG 2 (30%) to ORG 2 (70%)	. = =	. = ↓

Table 2-5 - Summary of mean ISA results

Figure 2-1 shows a graph of the mean ISA results. For each combination of organisation and controller role, the distribution of mean ISA score per run was calculated. The 25th, 50th and 75th percentiles of this distribution are shown. Only medium and high traffic volumes were included, since ORG 2 (30%) was not evaluated against low volume traffic. For example, from the graph it can be seen that the median of mean ISA for the Tactical Controller is about the same in ORG 0 and ORG 2 (70%) (actually 2.16 and 2.15 respectively).

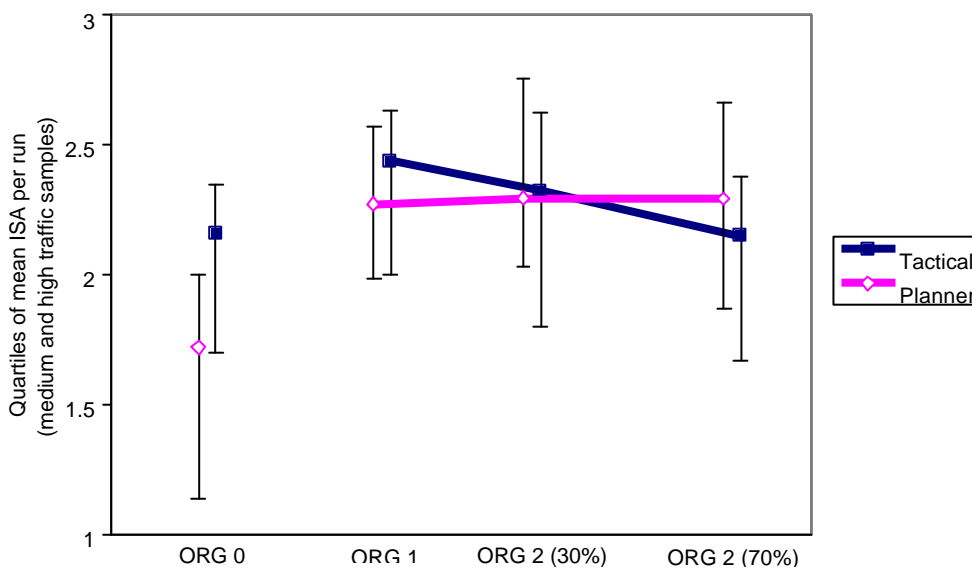


Figure 2-1 - Comparison between ORGs and controller role of mean ISA scores

In Figure 2-1, the line connecting ORG 1 to the ORG 2 values is interpreting ORG 1 as ORG 2 (0%), ie ORG 2 with no aircraft with 4-D FMS and datalink. The suggestion of the figure is that, in the interval 0% to 70%, workload is a linear function of the proportion of 4-D FMS and datalinked aircraft. This hypothesis has not been tested statistically, but might be worth further consideration (taking ISA and TLX jointly).

#### 2.2.1.2. Comparison of mean ISA scores between controller roles

The comparison of mean ISA scores between the two controller roles allows an assessment to be made of the impact of the introduction of computer assistance tools and 4-D FMS/datalink on the split of workload between the Tactical and Planning Controllers. The results are presented in Table 2-6, which assesses whether the null hypothesis that “the workload of the

Tactical Controller is the same as that of the Planning Controller ( $H_{0.5}$ )” is valid. The test used was the Wilcoxon-Mann-Whitney (WMW) test. The Wilcoxon Signed Ranks test did not apply because the comparison was between controllers rather than the response of each controller to differing scenarios.

The results in Table 2-6 show that:

- the mean ISAs of the Tactical Controllers were significantly greater than those of the Planning Controllers in ORG 0 for high traffic levels
- the mean ISAs were the same for both controllers in ORG 1
- there is at least an indication of a trend for the Planning Controllers to record higher mean ISAs than the Tactical Controllers in ORG 2 (70%).

Thus Table 2-6 shows that PD/1 has gone some way to achieving one of its objectives, viz transferring workload from the Tactical to the Planning Controller.

Roles Compared	ORG	Sectors	Traffic Volume	$P(WMW)$	Significant?	Conclusion
Tactical vs Planner	ORG 0	Both		0.0157	At 2% level	Tactical > Planner
		10	All	>0.15	No	Accept $H_{0.5}$
		11		0.049	At 5% level	Tactical > Planner
		Both	Low	>0.15	No	Accept $H_{0.5}$
			Medium	0.0605	At 10% level	[Tactical > Planner]
			High	0.0311	At 5% level	Tactical > Planner
	ORG 1	Both	All	>0.15	No	Accept $H_{0.5}$
			Low	>0.15	No	Accept $H_{0.5}$
			Medium	>0.15	No	Accept $H_{0.5}$
			High	>0.15	No	Accept $H_{0.5}$
	ORG 2 (30%)	Both	All	>0.15	No	Accept $H_{0.5}$
	ORG 2 (70%)	Both	All	0.0645	At 10% level	[Planner > Tactical]
		10		>0.15	No	Accept $H_{0.5}$
		11		0.0783	At 10% level	[Planner > Tactical]
		Both	Medium	0.0836	At 10% level	[Planner > Tactical]
			High	0.133	No	Accept $H_{0.5}$

Table 2-6 - Results for the comparison of the ISA means between controller roles

### 2.2.1.3. Comparison of mean ISA scores between sector 10 and 11

Table 2-7 shows the comparison of the mean ISAs recorded for sectors 10 and 11 and assesses whether the null hypothesis “the workload in sector 10 is the same as that of sector 11 ( $H_{0.6}$ )” is valid. The test used was the Wilcoxon-Mann-Whitney (WMW) test. The

Wilcoxon Signed Ranks (WSR) test did not apply since the comparison was between different controllers.

The mean ISAs of the controllers in sector 10 were found to be significantly greater than the mean ISAs of the sector 11 controllers over nearly all the organisations. It is first important to note that the workload scores in sector 10 are significantly higher than those of sector 11 in the baseline organisation. This means that sector 10 is inherently more difficult to control than sector 11. Second, the introduction of the computer assistance tools and 4-D FMS/datalink did not change the relative difficulty of controlling the two sectors although, for ORG 2, the difference between the sectors was only a trend, not a significant result.

ORG	Controller role	Traffic Volume	$P(WMW)$	Significant?	Conclusion
All	Planner	All	0.0072	At 1% level	10 > 11
	Tactical		<0.001	At 0.1% level	10 > 11
ORG 0	Both		0.0195	At 2% level	10 > 11
ORG 1			<0.001	At 0.1% level	10 > 11
ORG 2 (30%)			0.0677	At 10% level	[10 > 11]
ORG 2 (70%)			0.0917	At 10% level	[10 > 11]

Table 2-7 - Results for the comparison of the ISA means between sectors

#### 2.2.1.4. Other ISA scores

As stated in Section 2.2.1, ISA statistics other than the mean ISA per run were also investigated; namely, the peak ISA per run, the percentage of ISA scores = 5 per run, and the percentage of ISA scores  $\geq 4$  per run. These measures were found to correlate very highly with the mean ISA scores, at the 0.1% level and, accordingly, they yielded very similar results when making comparisons between the ORGs.

For the peak ISA scores, as for the mean ISA scores, it was found that ORG 1 had significantly higher peak workload than ORG 0, and that ORG 2 with 30% 4D FMS and datalink proved to have significantly higher workload than ORG 2 with 70% 4D FMS and datalink. However, no significant differences were found between ORG 0 and ORG 2 with 70% 4D FMS and datalink.

For the percentage of ISA scores  $\geq 4$ , it was observed that ORG 1 had a significantly greater percentage than ORG 0, for both Tactical and Planner roles. The Planner role in ORG 2 with 70% 4D FMS and datalink also had a significantly greater percentage than ORG 0 and ORG 1. The Tactical role in ORG 2 with 70% 4D FMS and datalink, on the other hand, showed no significant difference from ORG 0, and was found to have a significantly smaller percentage than both ORG 1 and ORG 2 with 30% 4-D FMS and datalink. No evidence was found of a difference of percentages between ORG 2 with 30% 4D FMS and datalink and ORG 2 with 70% 4-D FMS and datalink for the Planning Controller role.

For the percentage of ISA scores = 5, no significant conclusions could be drawn since too few of the simulation runs (28 from 329 runs, or 89 ISA responses from 10,903 in total) recorded ISA scores = 5. This must be considered as an encouraging result. If a significant number of

ISA scores of 5 had been recorded, then the acceptability of the PD/1 operational concept would need to be questioned.

The total number of ISA scores = 4 and 5 was only 2.0% of the total in ORG 0, 11.9% in ORG 1 and 5.8% in ORG 2 (70%). Again this is encouraging, since NATS experience in evaluating other simulations suggests that when ISA scores of 4 and 5 are fewer than 6% of all the ISA scores, the operational concept under consideration is acceptable for implementation.

### 2.2.2. TLX scores

The TLX scores were analysed: first, to test the consistency of TLX and the ISA scores; and second, to establish which particular factors of workload were most important. Table 2-8 and Table 2-9 below give the results using the Wilcoxon Signed Ranks test for the overall TLX scores for the Planning and Tactical Controllers respectively. Figure 2-2 on page 21 illustrates the tables. The null hypotheses tested were identical to those for the ISA tests, i.e.  $H_{0-1}$  to  $H_{0-4}$  (see Section 2.1).

In Table 2-8 to Table 2-19 and Table 2-21 to Table 2-22 below, P signifies the probability of obtaining the observed difference, or a more extreme observation, by chance if the null hypothesis  $H_0$  were indeed true, where  $H_0$  is the hypothesis that there is no difference between the ORGs. Note that by ORG a > ORG b, it is meant that the mean TLX scores for ORG a were **significantly** greater than those for ORG b and this difference is unlikely to have occurred by chance.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	<0.001	At 0.1% level	ORG 1 > ORG 0
	10		0.0345	At 5% level	ORG 1 > ORG 0
	11		0.0081	At 1% level	ORG 1 > ORG 0
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
	10		0.0086	At 1% level	ORG 2 (70%) > ORG 0
	11		0.0163	At 5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0184	At 5% level	ORG 2 (70%) > ORG 1
	10		0.0174	At 5% level	ORG 2 (70%) > ORG 1
	11		0.1571	No	Accept $H_{0-3}$
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	0.4617	No	Accept $H_{0-4}$

Table 2-8 - Results for the overall TLX scores recorded by Planning Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0317	At 5% level	ORG 1 > ORG 0
	10		0.0337	At 5% level	ORG 1 > ORG 0
	11		0.1918	No	Accept H <sub>0-1</sub>
ORG 0 vs ORG 2 (70%)	Both	All	0.1614	No	Accept H <sub>0-2</sub>
ORG 1 vs ORG 2 (70%)	Both	All	0.0026	At 0.5% level	ORG 1 > ORG 2 (70%)
	10		0.0252	At 5% level	ORG 1 > ORG 2 (70%)
	11		0.0200	At 5% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	0.0015	At 0.5% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	0.0009	At 0.5% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.1026	No	Accept H <sub>0-4</sub>

*Table 2-9 - Results for the overall TLX scores recorded by Tactical Controllers*

As can be seen from Table 2-8 and Table 2-9 and Figure 2-2, the overall TLX scores reflect the same workload trends as the results from the ISA scores.

The overall TLX of the Planning Controllers displays a strong tendency to increase from ORG 0 to ORG 1 (significant at the 0.1% level). However, for the Tactical Controllers, the overall TLX increases significantly from ORG 0 to ORG 1 in sector 10 but not in sector 11. This sector specific increase in workload for the Tactical Controllers was also observed in mean ISA (Table 2-1, bottom). (It was also found in Section 2.2.1.3 that sector 10 was inherently more difficult to control than sector 11.)

Figure 2-2 shows a graph of the overall TLX results. For each combination of organisation and controller role, the distribution of overall TLX in each run was calculated. The 25th, 50th and 75th percentiles of this distribution are shown. Only medium and high traffic volumes were included, since ORG 2 (30%) was not evaluated against low volume traffic. Note that the overall TLX scores range from 15 to 300.

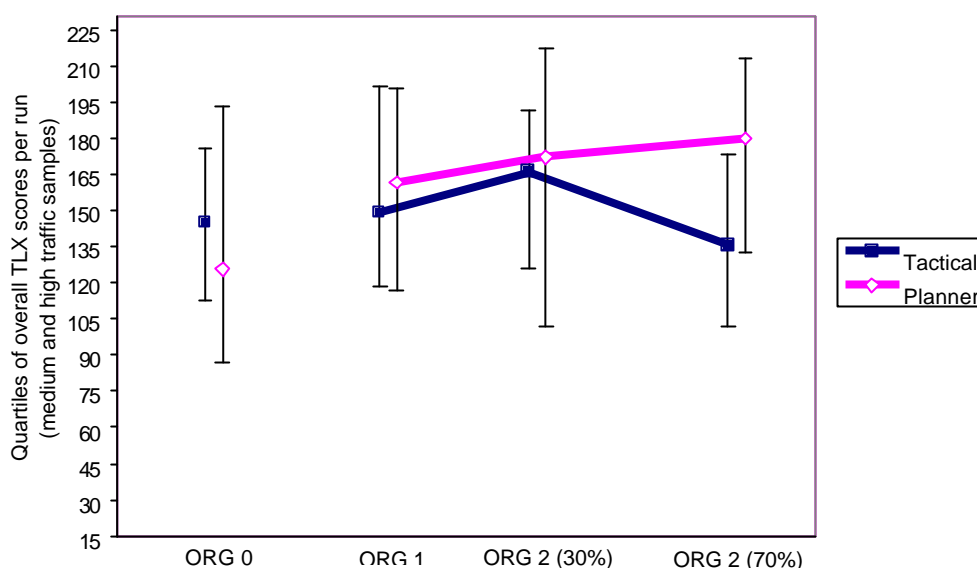


Figure 2-2 - Comparison between ORGs and controller role of the overall TLX scores

The Tactical Controllers did not significantly distinguish between the workload of ORG 0 and ORG 2 with 70% 4-D FMS and datalink, whereas the Planning Controllers had significantly greater workload in ORG 2 (70%). Comparing ORG 1 and ORG 2 with 70% 4-D FMS and datalink, the Planning Controllers had greater overall TLX scores in ORG 2 (70%), whereas the Tactical Controllers had greater overall TLX scores in ORG 1. Again this effect was also found in the comparison between mean ISA scores shown in Table 2-2 and Table 2-3. This similarity between the ISA results and the TLX results means that examination (given below) of the individual TLX workload factors might help to explain the workload trends identified by both ISA and TLX.

#### 2.2.2.1. Unweighted TLX scores - Tactical Controller

The results for the individual unweighted TLX workload factors recorded by the Tactical Controllers (namely effort expended, mental demand, time pressure, physical demand and frustration experienced) are presented in Table 2-10 to Table 2-14 below and summarised in Table 2-20 and Figure 2-3. The null hypotheses tested were identical to those for the ISA tests, i.e.  $H_{0-1}$  to  $H_{0-4}$  (see Section 2.1). Again, the Wilcoxon Signed Ranks test was used.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.3210	No	Accept H <sub>0-1</sub>
ORG 0 vs ORG 2 (70%)	Both	All	0.1973	No	Accept H <sub>0-2</sub>
ORG 1 vs ORG 2 (70%)	Both	All	0.0838	At 10% level	[ORG 1 > ORG 2 (70%)]
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	0.0340	At 5% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	0.0967	At 10% level	[ORG 2 (30%) > ORG 2 (70%)]
	11	High	0.1020	No	Accept H <sub>0-4</sub>

Table 2-10 - Results for the Effort Expended scores recorded by Tactical Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0825	At 10% level	[ORG 1 > ORG 0]
ORG 0 vs ORG 2 (70%)	Both	All	0.0773	At 10% level	[ORG 0 > ORG 2 (70%)]
ORG 1 vs ORG 2 (70%)	Both	All	0.0178	At 5% level	ORG 1 > ORG 2 (70%)
	10		0.2085	No	Accept H <sub>0-3</sub>
	11		0.0156	At 5% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	0.0103	At 5% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	0.0105	At 5% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.1826	No	Accept H <sub>0-4</sub>

Table 2-11 - Results for the Mental Demand scores recorded by Tactical Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0021	At 0.5% level	ORG 1 > ORG 0
	10		0.0018	At 0.5% level	ORG 1 > ORG 0
	11		0.1445	No	Accept H <sub>0-1</sub>
ORG 0 vs ORG 2 (70%)	Both	All	0.3077	No	Accept H <sub>0-2</sub>
ORG 1 vs vs ORG 2 (70%)	Both	All	0.0080	At 1% level	ORG 1 > ORG 2 (70%)
	10		0.0167	At 5% level	ORG 1 > ORG 2 (70%)
	11		0.0795	At 10% level	[ORG 1 > ORG 2 (70%)]
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	0.0186	At 5% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	0.0112	At 5% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.2651	No	Accept H <sub>0-4</sub>

Table 2-12 - Results for the Time Pressure scores recorded by Tactical Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.3928	No	Accept $H_{0,1}$
ORG 0 vs ORG 2 (70%)	Both	All	0.0450	At 5% level	ORG 0 > ORG 2 (70%)
	10		0.2470	No	Accept $H_{0,2}$
	11		0.0430	At 5% level	ORG 0 > ORG 2 (70%)
ORG 1 vs ORG 2 (70%)	Both	All	0.0386	At 5% level	ORG 1 > ORG 2 (70%)
	10		0.3613	No	Accept $H_{0,3}$
	11		0.0164	At 5% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	0.0013	At 0.5% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	0.0040	At 0.5% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.0474	At 5% level	ORG 2 (30%) > ORG 2 (70%)

Table 2-13 - Results for the Physical Demand scores recorded by Tactical Controllers

Examination of the individual, unweighted TLX workload factors for the Tactical Controller shows a similar pattern to the overall TLX workload factor. The Tactical Controller on sector 10 showed an increase from ORG 0 to ORG 1 in the frustration experienced and time pressure scores, and a non-significant increase in mental demand scores. As with the overall TLX workload factor and ISA scores, the Tactical Controller on sector 11 showed no increase from ORG 0 to ORG 1 with the exception of frustration experienced.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0066	At 1% level	ORG 1 > ORG 0
	10		0.0298	At 5% level	ORG 1 > ORG 0
	11		0.0447	At 5% level	ORG 1 > ORG 0
ORG 0 vs ORG 2 (70%)	Both	All	0.3790	No	Accept $H_{0,2}$
ORG 1 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
	10		0.0020	At 0.5% level	ORG 1 > ORG 2 (70%)
	11		0.0310	At 5% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	0.0172	At 5% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	0.0031	At 0.5% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.3445	No	Accept $H_{0,4}$

Table 2-14 - Results for the Frustration Experienced scores recorded by Tactical Controllers

The only significant difference noted between ORG 2 (70%) and ORG 0 for Tactical Controllers was for physical demand averaged over both sectors. This showed that ORG 0 placed a greater physical demand on the Tactical Controllers than in ORG 2 (70%), especially in sector 11. There was also a non-significant trend for the mental demand scores to be less

in ORG 2 (70%). As with the overall TLX and ISA scores, there was a statistically significant decrease in most of the individual unweighted TLX workload factors for both sectors combined when moving from ORG 1 to ORG 2 (70%) and from ORG 2 (30%) to ORG 2 (70%) - especially in sector 10. The only exception was effort expended, where the ORG 1/ORG 2 (70%) result was not statistically significant.

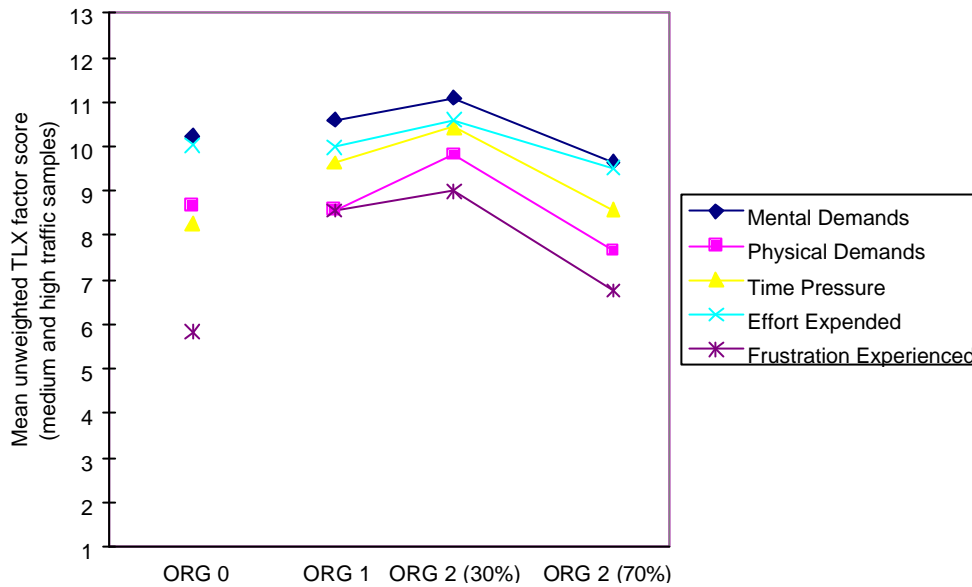


Figure 2-3 - Comparison between ORGs of the average scores for the TLX factors recorded by the Tactical Controllers

For Figure 2-3, note that TLX factor scores scale covers the range from 1 (minimum contribution of factor to workload) to 20 (maximum).

#### 2.2.2.2. Unweighted TLX scores - Planning Controller

The results for the individual unweighted TLX workload factors recorded by the Planning Controllers (namely effort expended, mental demand, time pressure, physical demand and frustration experienced) are presented in Table 2-15 to Table 2-19 below and summarised in Table 2-20 and Figure 2-4. The null hypotheses tested were identical to those for the ISA tests, i.e.  $H_{0,1}$  to  $H_{0,4}$  (see Section 2.1). Again, the Wilcoxon Signed Ranks test was used.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0144	At 5% level	ORG 1 > ORG 0
	10		0.0721	At 10% level	[ORG 1 > ORG 0]
	11		0.0677	At 10% level	[ORG 1 > ORG 0]
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
	10		0.0026	At 0.5% level	ORG 2 (70%) > ORG 0
	11		0.0181	At 5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0264	At 5% level	ORG 2 (70%) > ORG 1
	10		0.0297	At 5% level	ORG 2 (70%) > ORG 1
	11		0.1652	No	Accept H <sub>0-3</sub>
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	0.1452	No	Accept H <sub>0-4</sub>

Table 2-15 - Results for the Effort Expended scores recorded by Planning Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.2824	No	Accept H <sub>0-1</sub>
ORG 0 vs ORG 2 (70%)	Both	All	0.0256	At 5% level	ORG 2 (70%) > ORG 0
	10		0.0476	At 5% level	ORG 2 (70%) > ORG 0
	11		0.1368	No	Accept H <sub>0-2</sub>
ORG 1 vs ORG 2 (70%)	Both	All	0.0374	At 5% level	ORG 2 (70%) > ORG 1
	10		0.0381	At 5% level	ORG 2 (70%) > ORG 1
	11		0.2715	No	Accept H <sub>0-3</sub>
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	0.1501	No	Accept H <sub>0-4</sub>

Table 2-16 - Results for the Mental Demand scores recorded by Planning Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0168	At 5% level	ORG 1 > ORG 0
	10		0.0780	At 10% level	[ORG 1 > ORG 0]
	11		0.0698	At 10% level	[ORG 1 > ORG 0]
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
	10		0.0137	At 5% level	ORG 2 (70%) > ORG 0
	11		0.0013	At 0.5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0282	At 5% level	ORG 2 (70%) > ORG 1
	10		0.0412	At 5% level	ORG 2 (70%) > ORG 1
	11		0.2070	No	Accept H <sub>0-3</sub>
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	0.0786	At 10% level	[ORG 2 (30%) > ORG 2 (70%)]

Table 2-17 - Results for the Time Pressure scores recorded by Planning Controllers

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	0.0506	At 10% level	[ORG 1 > ORG 0]
ORG 0 vs ORG 2 (70%)	Both	All	0.0027	At 0.5% level	ORG 2 (70%) > ORG 0
	10		0.0247	At 5% level	ORG 2 (70%) > ORG 0
	11		0.0170	At 5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0061	At 1% level	ORG 2 (70%) > ORG 1
	10		0.0040	At 0.5% level	ORG 2 (70%) > ORG 1
	11		0.1532	No	Accept H <sub>0-3</sub>
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	0.4039	No	Accept H <sub>0-4</sub>

Table 2-18 - Results for the Physical Demand scores recorded by Planning Controllers

The individual unweighted TLX workload scores do not highlight any particular reason for the workload increase from ORG 0 to ORG 1 and ORG 2. All factors except mental demand and physical demand showed a significant increase from ORG 0 to ORG 1. From ORG 0 to ORG 2 (70%) and from ORG 1 to ORG 2 (70%) all factors showed a significant increase - particularly in sector 10. There was no evidence of a change in the scores of any factor from ORG 2 (30%) to ORG 2 (70%). These results again are similar to those obtained from the overall TLX scores.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	<0.001	At 0.1% level	ORG 1 > ORG 0
	10		0.0015	At 0.5% level	ORG 1 > ORG 0
	11		0.0035	At 0.5% level	ORG 1 > ORG 0
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
	10		0.0014	At 0.5% level	ORG 2 (70%) > ORG 0
	11		0.0315	At 5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0312	At 5% level	ORG 2 (70%) > ORG 1
	10		0.0512	At 10% level	[ORG 2 (70%) > ORG 1]
	11		0.1583	No	Accept H <sub>0,3</sub>
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	0.3069	No	Accept H <sub>0,4</sub>

Table 2-19 - Results for the Frustration Experienced scores recorded by Planning Controllers

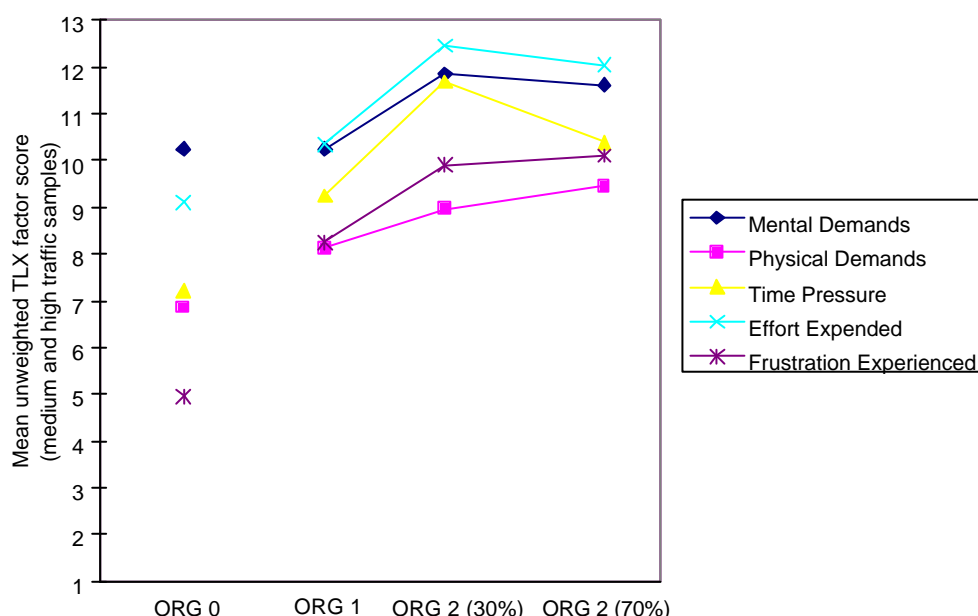


Figure 2-4 - Comparison between ORGs of the average scores for the TLX factors recorded by Planning Controllers

### 2.2.2.3. Unweighted TLX scores - comparison between roles

The individual, unweighted TLX workload scores vary with organisation differently for the Tactical and Planning Controllers. The difference between the mental demand scores for the Tactical and Planning Controllers is an interesting illustration of this: the Planning Controllers, who are assisted by most of the new computer tools, identified a significant increase in mental demand from ORGs 0 and 1 to ORG 2 (70%); the Tactical Controllers, on the other hand, showed a non-significant increase from ORG 0 to ORG 1, a significant decrease from

ORG 2 (30%) to ORG 2 (70%) and a trend (at the 10% significance level) for a *decrease* from ORG 0 to ORG 2 (70%).

Going from the baseline organisation, ORG 0, to the most advanced organisation, ORG 2 (70%), resulted in a significant increase in all factors for the Planning Controller. However, for the Tactical Controller, the only trend (mental demand) and significant result (physical demand) were *reductions* in workload.

Both roles reported that ORG 1 was significantly more frustrating than ORG 0. For the Tactical Controllers, ORG 1 was also more frustrating than ORG 2 (70%), the net result being that there was, for them, no difference in frustration levels between ORG 0 and ORG 2 (70%).

These results are summarised in Table 2-20 below for both sectors combined. The '=' symbol indicates that there was no statistically significant change in the factor score for that controller role, when going from ORG 0 to ORG 2 (70%). The '↑' and '↓' symbols indicate that the TLX factor significantly increases and decreases respectively, in that controller role, when going from ORG 0 to ORG 2 (70%). The square brackets indicate the general direction of the trend, where the result was significant at the 10% level only.

TLX Factor	Planner Controller	Tactical Controller
Effort Expended	↑	=
Mental Demand	↑	[↓]
Time Pressure	↑	=
Physical Demand	↑	↓
Frustration Experienced	↑	=

*Table 2-20 - Unweighted individual TLX factors between ORG 0 and ORG 2 (70%)*

#### 2.2.2.4. Comparison between controller roles of overall TLX scores

As for the mean ISA scores, the null hypothesis  $H_{0.5}$  was tested using overall TLX scores that there was no difference between the workload of the two controller roles for each of the ORGs. Once again, the test used was the Wilcoxon-Mann-Whitney (WMW) test. The Wilcoxon Signed Ranks test did not apply because the comparison was between controllers rather than the response of each controller to differing scenarios.

Roles Compared	ORG	Sectors	Traffic Volume	$P(WMW)$	Significant?	Conclusion
Tactical vs Planner	ORG 0	Both	All	0.1609	No	Accept $H_{0.5}$
	ORG 1	Both	All	0.4269	No	Accept $H_{0.5}$
	ORG 2 (30%)	Both	All	0.4579	No	Accept $H_{0.5}$
	ORG 2 (70%)	Both	All	0.0077	At 1% level	Planner > Tactical
		10		0.0422	At 5% level	Planner > Tactical
		11		0.0285	At 5% level	Planner > Tactical

Table 2-21 - Comparison between controller roles of overall TLX scores

The results in Table 2-21 show that for all ORGs except ORG 2 (70%) there was no significant difference in the overall TLX scores between the Planning and Tactical Controllers. For ORG 2 (70%) it was found that the overall TLX scores of the Planning Controllers were significantly greater than those of the Tactical Controllers, in both sectors. Figure 2-2 on page 21 illustrates this.

#### 2.2.2.5. Comparison between sectors of overall TLX scores

As for the mean ISA results, the null hypothesis  $H_{0.6}$  was tested that no difference existed in the overall TLX scores between the two sectors. As before, the test used was the Wilcoxon-Mann-Whitney (WMW) test. The Wilcoxon Signed Ranks test did not apply because the comparison was between controllers rather than the response of each controller to differing scenarios. The results of each test are presented in Table 2-22 below.

Sectors Compared	ORG	Roles	Traffic Volume	$P(WMW)$	Significant?	Conclusion
10 vs 11	ORG 0	Both	All	0.3233	No	Accept $H_{0.6}$
	ORG 1	Both	All	0.2740	No	Accept $H_{0.6}$
	ORG 2 (30%)	Both	All	0.0920	At 10% level	[10 > 11]
	ORG 2 (70%)	Both	All	0.0023	At 0.5% level	10 > 11

Table 2-22 - Comparison between sectors of overall TLX scores

No significant difference was found in overall TLX scores between the two sectors in any ORG except ORG 2 (70%), in which the scores for sector 10 were found to be significantly greater than those of sector 11. There was also a non-significant trend in ORG 2 (30%) for the scores of sector 10 to be greater. These results do not correspond exactly with those from the mean ISA scores (see Section 2.2.1.3).

### 2.3. OBJECTIVE MEASURES

A number of objective measures of workload were examined to compare with and complement the results obtained from the analysis of the perceived workload measures - ISA and TLX. These measures were:

- number of ATC instructions issued per aircraft (*Tactical Controller*)
- number of R/T communications, and percentage of time spent using R/T, per Tactical Controller per run (*Tactical Controller*)
- number and duration of telephone communications per Planning Controller per run (*Planning Controller*)

The tables presented in this section display the results of Wilcoxon Signed Ranks test using the same null hypotheses used for the ISA analysis, i.e.  $H_{0.1}$  to  $H_{0.4}$  (see Section 2.1).

In each table, P signifies the probability of obtaining the observed difference, or a more extreme observation, by chance if the null hypothesis  $H_0$  were indeed true, where  $H_0$  is the hypothesis that there is no difference between the ORGs. Note that by  $ORG\ a > ORG\ b$ , it is meant that the measures under ORG a were **significantly** greater than those under ORG b and this difference is unlikely to have occurred by chance.

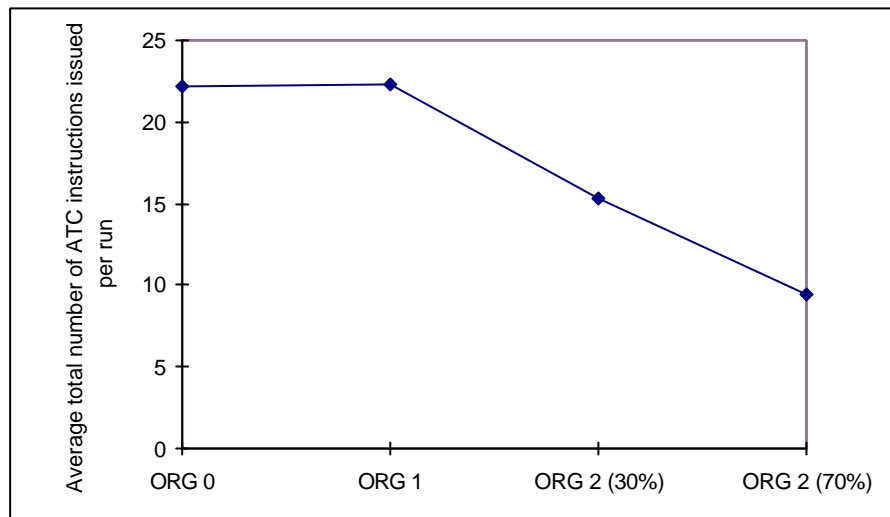
#### 2.3.1. Number of ATC instructions issued

Table 2-23 presents the results for the number of ATC instructions issued via the R/T. An ATC instruction is defined as being one of either a flight level, speed or heading change, or a go-direct instruction. The results have again been tested against the null hypotheses used for the ISA analysis, i.e.  $H_{0.1}$  to  $H_{0.4}$  (see Section 2.1), and evaluated using the Wilcoxon Signed Ranks test. As can be seen, the number of ATC instructions issued by the Tactical Controller in ORG 2 (70%) is significantly fewer than the baseline ORG 0 and ORG 1. In addition, ORG 2 (70%) has significantly fewer than any of the other ORGs, ORG 0, ORG 1 and ORG 2 (30%). No significant difference was found between ORG 0 and ORG 1.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	>0.15	No	Accept $H_{0.1}$
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 0 > ORG 2 (70%)
ORG 1 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	<0.001	At 0.1% level	ORG 2 (30%) > ORG 2 (70%)

*Table 2-23 - Results for the number of ATC instructions issued*

Figure 2-5 below summarises the above table. For each ORG, the number of ATC instructions has been averaged across medium and high traffic levels only. The number of instructions has also been averaged across both sectors.



*Figure 2-5 - Comparison between ORGs of the average total number of ATC instructions issued per run per sector*

As expected, given datalinked aircraft, there is a clear reduction in the number of ATC instructions issued. When interpreting these results, it should be noted that the task of issuing an instruction differs between the ORGs and therefore the workload associated with each instruction may be different:

- In ORG 0, the instructions are the result of the plans of the tactical controller, as in current day systems.
- In ORGs 1 and 2, the instructions are either: (a) issued (to non-datalinked aircraft) by the tactical controller, when prompted by the system, in order to follow the plan of the planning controller; (b) instructions to implement the tactical controller's response to unplanned events, such as deviation from plans.

Moreover, in ORG 2 for a proportion of the aircraft (30% or 70%), if there are no deviations then there should be no need to issue R/T instructions at all.

The average total number of ATC instructions issued per run showed a variation between sectors. There were a greater number of instructions issued in sector 10 than in sector 11: 30.1, 26.9, 20.2 and 11.3 compared to 14.3, 17.6, 10.4 and 7.6 in ORGs 0, 1, 2 (30%) and 2 (70%) respectively (over medium and high traffic volumes only). This is consistent with differences between controller workload in sectors 10 and 11 found in Sections 2.2.1.3 and 2.2.2.5. The greater number of ATC instructions issued in sector 10 may be a result of a greater number of conflicts in sector 10 due to the fact that sector 10 had more route crossing points than sector 11. However, the conduct of this analysis could not attribute this cause to the observed difference.

### 2.3.2. R/T communications

The time spent by the Tactical Controller on the R/T is a well-established measure of part of the controller workload for comparisons of similar systems: the more time that controllers spend on the R/T, the busier they are. However, care needs to be taken when comparing dissimilar systems, especially if R/T usage has been replaced by a different communication system. The PD/1 data were, therefore, analysed to examine the usage of the R/T by the Tactical Controller. R/T usage was measured by the number of R/T calls made by the

Tactical Controller and also by the percentage of time using R/T. To allow a fair comparison of runs of less than 75 minutes duration in the analysis, the number of R/T calls for those runs were 'corrected' by being linearly scaled up to the same duration as complete runs.

2.3.2.1. Number of R/T communications per Tactical Controller per run

Table 2-24 and Figure 2-6 below show that, as is to be expected, the number of R/T communications by the Tactical Controller closely replicates the results obtained from the number of ATC instructions issued (Section 2.3.1).

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	>0.15	No	Accept $H_{0,1}$
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 0 > ORG 2 (70%)
	10		<0.001	At 0.1% level	ORG 0 > ORG 2 (70%)
	11		<0.001	At 0.1% level	ORG 0 > ORG 2 (70%)
ORG 1 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
	10		<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
	11		<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	<0.001	At 0.1% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	<0.001	At 0.1% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.052	At 10% level	[ORG 2 (30%) > ORG 2 (70%)]

Table 2-24 - Results for the total corrected number of R/T calls made

Figure 2-6 shows a graph of the corrected number of R/T calls made per run. For each combination of organisation and sector, the distribution of the corrected number of R/T calls made was calculated. The 25th, 50th and 75th percentiles of this distribution are shown. Only medium and high traffic volumes were included, since ORG 2 (30%) was not evaluated against low volume traffic. For example, from the graph it can be seen that the median of the number of calls in sector 10 is much more in ORG 0 than in ORG 2 (70%) (actually 176.5 and 143 respectively).

As can be seen from Figure 2-6, a typical value for the number of R/T communications in ORG 0 and ORG 1 is about 160. However, a typical figure for the number of ATC instructions issued is about 20; the extra 140 R/T communications by the Tactical Controller were not to issue ATC instructions but: to respond to aircraft calling in; to release aircraft to the next sector; to confirm communications, or to acknowledge pilots reporting present position, etc.

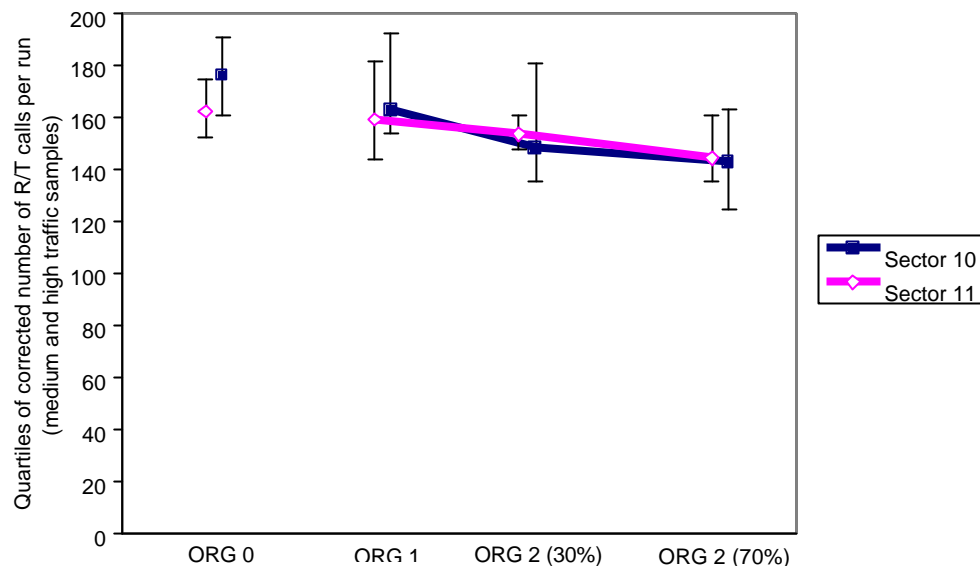


Figure 2-6 - Comparison between ORGs and sectors of corrected number of R/T calls

### 2.3.2.2. Percentage of time spent using R/T per Tactical Controller per run

In addition to the number of R/T communications issued by the Tactical Controller, it is also necessary to consider the percentage of time spent using R/T. It would, for instance, be possible for the number of communications to decrease significantly, but for their total duration to increase - which would indicate that the controllers had an increased R/T workload while dealing directly with fewer aircraft.

Table 2-25 presents the results for the comparison of the percentage of time per run spent by Tactical Controllers using R/T. The results are similar to those for the number of ATC instructions issued (Section 2.3.1). ORG 2 (70%) has a significantly shorter percentage of time spent using R/T than all other ORGs. Note that in the comparison between ORGs 0 and 1, there is significantly less time spent using R/T in ORG 0 for sector 11, whilst in sector 10 there is a trend for ORG 1 to have less R/T usage.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	>0.15	No	Accept $H_{0,1}$
	10		0.0668	At 10% level	[ORG 0 > ORG 1]
	11		0.0198	At 5% level	ORG 1 > ORG 0
ORG 0 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 0 > ORG 2 (70%)
	10		<0.001	At 0.1% level	ORG 0 > ORG 2 (70%)
	11		0.0045	At 0.5% level	ORG 0 > ORG 2 (70%)
ORG 1 vs ORG 2 (70%)	Both	All	<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
	10		<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
	11		<0.001	At 0.1% level	ORG 1 > ORG 2 (70%)
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium	<0.001	At 0.1% level	ORG 2 (30%) > ORG 2 (70%)
	10	and	<0.001	At 0.1% level	ORG 2 (30%) > ORG 2 (70%)
	11	High	0.0338	At 5% level	ORG 2 (30%) > ORG 2 (70%)

Table 2-25 - Results for the percentage of time spent using R/T per Tactical Controller per run

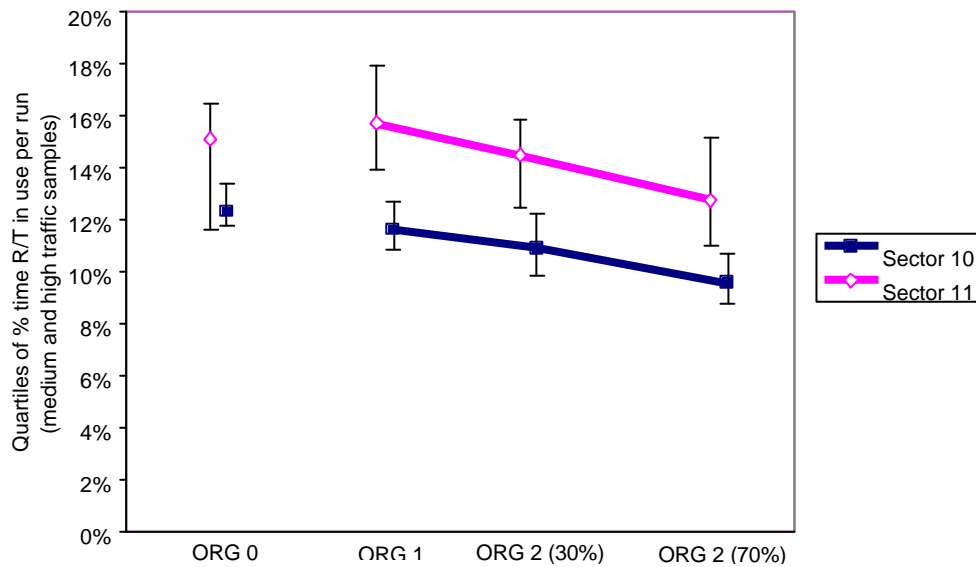


Figure 2-7 - Comparison between ORGs and sectors of the percentage of time spent using R/T per Tactical Controller per run

Figure 2-7 shows a graph of the percentage of time spent using R/T per run. For each combination of organisation and sector, the distribution of the percentage of time spent using R/T was calculated. The 25th, 50th and 75th percentiles of this distribution are shown. Only medium and high traffic volumes were included, since ORG 2 (30%) was not evaluated against low volume traffic. For example, from the graph it can be seen that the median of the percentage of time spent using R/T in sector 10 is greater in ORG 0 than in ORG 2 (70%) (actually 12.3% and 9.6% respectively).

These results indicate that the usage of R/T was greater in sector 11 than in sector 10. However, Sections 2.2.1.3 and 2.2.2.5 showed that the controllers perceived their workload to be greater in sector 10 than in sector 11. This illustrates that the usage of R/T is only one of many contributors to perceived workload. The difference in R/T usage may be attributable to the throughput figures shown in Annex A Table 3-1. The table shows that sector 11 has of the order of 20% higher traffic throughput than sector 10.

### 2.3.3. Telephone communications

As with the use of the R/T by the Tactical Controller, the use of the telephone by the Planning Controller is a well-established measure of part of controller workload; the more time that the controllers spend on the telephone to other sectors, the busier they are. The PD/1 data were, therefore, analysed to examine telephone usage by the Planning Controller.

#### 2.3.3.1. Duration of telephone communications per Planner Controller per run

For each combination of sector and organisation, the distribution was calculated of the proportion of time spent on phone calls (by the Planning Controller). Figure 2-8 shows the 25th, 50th and 75th percentiles of these distributions.

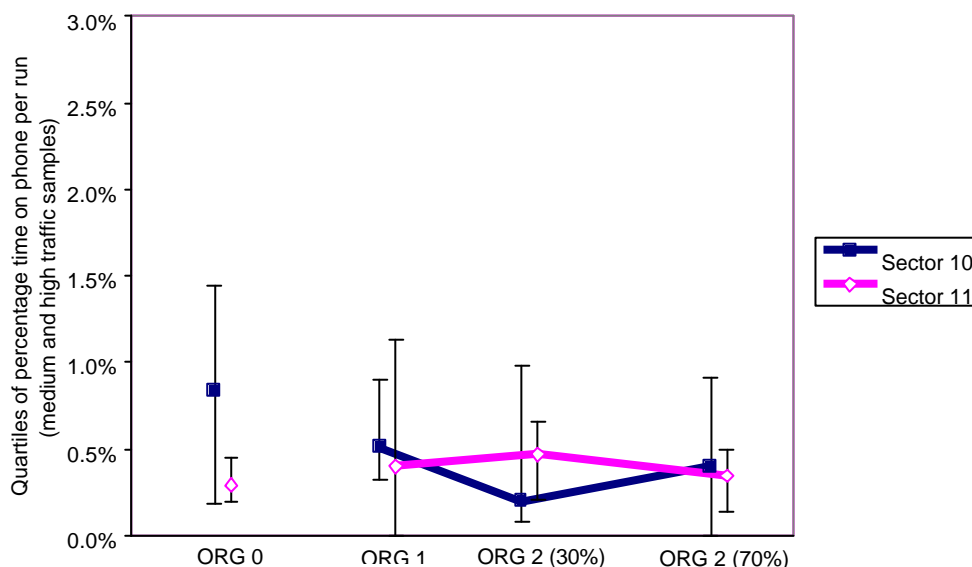


Figure 2-8 - Duration of phone calls

It is evident that Planning Controllers spend very little (around 0.5%) of their time on the phone. Given the use in PD/1 of electronic coordination, this is not at all surprising. Therefore, although statistical trends might be observed in the data, it is unlikely that they will provide any clues as to how to make reductions of any useful magnitude in overall workload. No further results will be presented.

## 2.4. SUMMARY AND DISCUSSION OF RESULTS

### 2.4.1. Introduction

The objective measures of components of controller workload which have been analysed are by no means exhaustive. Other unmeasured elements of workload may contribute significantly to the total workload of the controllers and the balance of workload between the different elements could change markedly in different scenarios; for instance less R/T could be counter-balanced by an increase in a different element. However, the subjective measures of workload can be envisaged as comprehensive measures of (perceived) workload. Hence a significant change in workload as determined by the objective measures analysed need not necessarily correspond to a significant change in the subjectively measured workload. However, if the same trends were to be seen in the objective and the subjective measures of workload, then there would be more confidence in the validity of the results.

The following subsections draw together the results for workload by comparing 3 pairs of organisations: ORG 0 v ORG 1; ORG 1 v ORG 2(70%); and ORG 0 v ORG 2 (70%).

### 2.4.2. Workload differences between ORG 0 and ORG 1

A move from ORG 0 to ORG 1 corresponds to the introduction of computer assistance tools and the associated ATC procedure changes. This move was associated with a significant increase in subjective workload (as measured by mean ISA and most of the TLX measures). However, it had a less obvious impact on the objective measures of workload: sector 11 had a significant increase in the percentage of time spent using R/T, whereas in sector 10 there was a non-significant trend in the opposite direction (the reason for this may be explained by the differing geometry of the two sectors). There was no significant increase in the number of air traffic control instructions issued by the Tactical Controller from ORG 0 to ORG 1.

Thus some, but not all, objective measures supported the subjective result; so explanations for the subjective result might need to be sought elsewhere. The individual TLX factors may explain the reasons for an increase in mean ISA with the introduction of computer assistance tools. For both Tactical and Planning Controllers, the frustration experienced showed the greatest proportionate increase from ORG 0 to ORG 1. Time pressure also showed a significant increase from ORG 0 to ORG 1 for both Tactical and Planning Controllers, especially in sector 10. It may be the case that the increase in perceived workload (as measured by mean ISA) is attributable to external factors, such as relative unfamiliarity with the computer assistance tools after a relatively short training period, rather than the workload that a fully trained controller would experience. Factors such as this could emerge in terms of increased frustration during this simulation trial.

It should also be emphasised that the controllers were using the computer assistance tools in ORG 1 to perform tasks not performed in the baseline ORG 0. Many of these extra tasks were performed by the Planning Controller, whose responsibilities in ORG 1 (and ORG 2) include using the tools to ensure that trajectories are conflict free for up to 20 minutes ahead. In this way the Planner aimed to establish a flow of traffic through the sector which was more easily managed by the Tactical Controller. Hence it would perhaps be expected that the Tactical Controllers' workload should decrease as a smoother flow of traffic was established by the Planner. It would also be expected that the Planning Controllers' subjective workload would increase, because they were required to do more.

This re-distribution of workload between the controller roles with the introduction of computer assistance tools was confirmed by the fact that in ORG 0 the Tactical Controllers have significantly higher mean ISA scores than the Planning Controllers. However, in ORG 1 there were no significant differences between the mean ISA scores.

Further analysis might show whether the Tactical and Planning Controller teams were effectively collaborating in the way expected in the design of the split of workload between the roles. For example, the question whether the Tactical Controller still subconsciously performed planning tasks remains unanswered.

#### 2.4.3. Workload differences between ORG 1 and ORG 2 (70%)

A move from ORG 1 to ORG 2 (70%) corresponds to the introduction of 70% 4-D FMS and datalink equipped aircraft, with computer assistance tools and associated procedures already in place. This move was associated with marked differences in both objective and subjective measures of workload. However, the differences were dependent on the controller role.

For Tactical Controllers there was a significant decrease in both mean ISAs and overall TLX with the introduction of 70% 4-D FMS and datalink equipped aircraft. Objective measures of workload also all decreased significantly, supporting the subjective result.

However, for the Planning Controller the overall TLX measure of workload significantly increased. There were increases in all of the components of TLX, especially in sector 10, which might explain this difference.

The introduction of 70% 4-D FMS and datalink equipped aircraft changed the tasks which the controller was expected to do, especially for the Tactical Controller. For the aircraft with datalink, clearances were no longer passed over the R/T; instead trajectories were negotiated by datalink and implemented by the 4-D FMS. The relay of many R/T clearances to aircraft in ORG 1 could be one reason for the Tactical Controllers' physical demands, time pressure and frustration experienced. All of these TLX factors decreased significantly when 70% of the aircraft were 4-D FMS and datalink-equipped.

One way to think of ORG 1 is as 'ORG 2 (0%)'. The comparison of ORG 2 (30%) and ORG 2 (70%) produced similar conclusions to those just described for ORG 1 versus ORG 2 (70%), although distinctions which could be drawn between ORG 1 and ORG 2 (70%) were not always observed between ORG 2 (30%) and ORG 2 (70%). This supports the picture of ORG 2 (30%) as an intermediate stage between ORG 1 and ORG 2 (70%). If the impact of the gradual introduction of 4-D-FMS and datalink is of interest, then further trial runs using, say, 100% 4-D-FMS and datalinked aircraft, would be useful. However, the variability of ISA and TLX suggest that it might be difficult to draw any conclusions with confidence at such a level of detail.

#### 2.4.4. Workload differences between ORG 0 and ORG 2 (70%)

A move from ORG 0 to ORG 2 (70%) corresponds to the introduction of computer assistance tools, the associated ATC procedure changes and 4-D FMS and datalinked aircraft. This move, just as in the previous case, had an impact which was dependent on the controller role.

For the Planning Controller there was a significant increase in all subjective measures of workload. The objective measures accounted for too small a portion of workload to be taken

to support or contradict this. Possible explanations are difficulties in the use of the tools or, more simply, the additional responsibilities of planning using the tools.

The mean ISA of the Tactical Controller actually decreased at low traffic volume and was not significantly changed at medium and high traffic volume. Although overall TLX was unchanged, the physical demand component of TLX decreased significantly and there was a trend for mental demand also to decrease. These may explain the ISA result. Unlike the Planning Controller case, the objective measures for the Tactical Controller were consistent with the subjective measures: they all decreased significantly. The traffic volume dependence might be interpreted as an indication that the controllers used the tools more effectively and with full confidence at low traffic volume. At higher levels of traffic, relative inexperience in using the tools may have made the controllers more likely to revert to the more familiar baseline operational procedures, hence not gaining full benefit in tool utilisation.

### 3. CAPACITY

#### 3.1. INTRODUCTION

The capacity of the sectors was recorded directly as an objective measure by the NRF using two parameters:

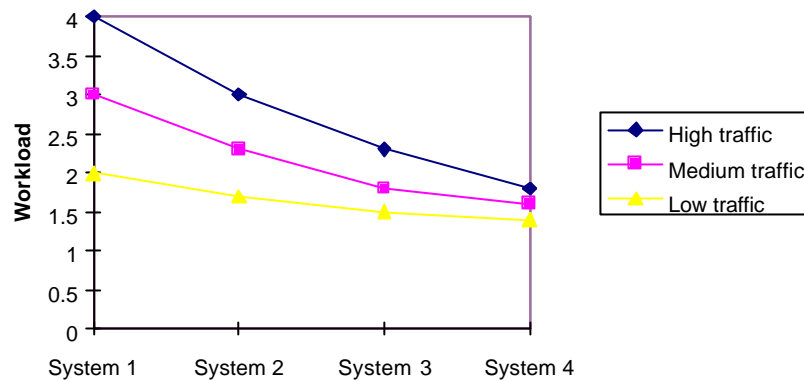
- the number of aircraft passing through the sector per hour;
- the number of aircraft assumed by the controllers at any one time.

These parameters are initially defined by the traffic samples employed during the trial (see Annex A), and are shown in Table 3-1 below. The number of aircraft ‘assumed’ is a time dependent parameter and cannot be presented in the simplified format of Table 3-1. (‘Assumed’ is a technical term; it indicates that the controller has accepted responsibility for the aircraft from the previous sector, and not yet passed it to the next sector.)

Number of aircraft per hour		Traffic sample
Sector 10	Sector 11	
30-31	33-34	Low
45-46	54-59	Medium
50-56	66-68	High

*Table 3-1 - Number of aircraft per hour used in defining the traffic samples*

For any given ATC system, the number of aircraft handled by the controllers is expected to correlate strongly with the perceived workload scores, and this was confirmed for the PD/1 trial. Thus, if the number of aircraft controlled per hour at a given workload level can be shown to increase between organisations, then it may be assumed that the capacity of the sector has increased. The hypothetical example shown in Figure 3-1 indicates that under System 4 the high traffic sample may be controlled at a workload level lower than the low traffic sample under System 1. Therefore it may be stated that an increase in capacity was achieved in moving from System 1 to System 4.



*Figure 3-1 - Hypothetical example of the relationship between ATC system, controller workload and traffic level*

For the PD/1 results, such a clear-cut statement cannot be made. As was extensively discussed in Section 2, the perceived workload for ORG 0 was not consistently significantly greater than the perceived workload for ORG 2 (70%). Therefore, it cannot be argued that an increase in capacity was shown to be achieved in moving from ORG 0 to ORG 2 (70%).

However, it is possible to consider other objective measures as an indication of whether a capacity increase may be possible by implementation of the advanced organisations. For the PD/1 analysis, the following objective measures were adopted as possible indicators of the airspace capacity effects of the ORGs:

- average lateral aircraft separation per run;
- average vertical aircraft separation per run;
- number of minimum safe separation infringements per run;
- duration of those minimum safe separation infringements;
- the percentage of time the ‘picture’ is maintained by controllers;
- number of Short Term Conflict Alerts (STCAs) activated per run;
- proportion of the time that the STCA was activated per run.

While some of these measures are obviously related to capacity, e.g. the average lateral separation, the ‘safety’ related criteria may not initially be such an obvious measure of capacity. However, it must be realised that if in one organisation the controllers cannot handle the number of aircraft ‘safely’ - as may be indicated by an increase in the number of STCAs - then the traffic sample under consideration may be too high for that sector and should be reduced. If the same traffic could be handled in a different organisation then there may be a change in capacity.

These measures were therefore used to test the following null hypotheses:

- the capacity of ORG 0 is the same as that of ORG 1 ( $H_{0,1}$ );
- the capacity of ORG 0 is the same as that of ORG 2 with 70% 4D FMS and datalink-equipped aircraft ( $H_{0,2}$ );
- the capacity of ORG 1 is the same as that of ORG 2 with 70% 4D FMS and datalink-equipped aircraft ( $H_{0,3}$ );

- the capacity of ORG 2 with 30% 4-D FMS and datalink-equipped aircraft is the same as that of ORG 2 with 70% 4-D FMS and datalink-equipped aircraft ( $H_{0.4}$ ).

Results of these tests, in terms of the various measures listed above, are given in the following subsections.

Throughout the following analysis in this section the Wilcoxon Signed Ranks test has been used. Results at the 10% level of significance have been included only to display trends and do not indicate that the null hypothesis is rejected.

As in Section 2,  $P$  signifies the probability of obtaining the observed difference, or a more extreme observation, by chance if the null hypothesis,  $H_0$ , were indeed true, where  $H_0$  is the hypothesis that there is no difference between the ORGs. Note that by ORG a > ORG b, it is meant that the measures for ORG a were **significantly** greater than those for ORG b and this difference is unlikely to have occurred by chance.

### 3.2. AVERAGE AIRCRAFT SEPARATIONS

It was proposed to consider average aircraft lateral and vertical separations as a measure of capacity. However, the measures that could be extracted are extremely difficult to interpret in the en-route context: for example an increase in lateral separation in could be argued to imply a reduction or an increase in capacity. Therefore no results are reported.

### 3.3. MINIMUM SEPARATION INFRINGEMENTS

The number and duration of minimum separation infringements may be considered as indirect measures of sector capacity (if they are constantly infringed then the traffic level is too high and the capacity is, therefore, lower than the traffic flow in the sample); however, they may also be a reflection of controller workload. For convenience, this measure is considered here rather than in Section 2.

Theoretically, if the tools are used 'perfectly' and function 'perfectly' and the controllers understand and operate the system correctly, then there should be no minimum separation infringements. In fact, there were infringements. While this section will not assess the reasons for the infringements occurring, four possibilities are mentioned:

- the controllers were overworked and therefore the traffic level was too high;
- the controllers were not well enough trained in the tools' use;
- the controllers did not obey the PD/1 Operational Concept;
- the prototype tools failed to predict some separation losses.

Item one on the above list justifies the inclusion of this measure as part of the capacity assessment.

#### 3.3.1. Number of minimum safe separation infringements per run

Table 3-2 below gives the results for the analysis of the number of minimum safe separation infringements per run.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	>0.15	No	Accept $H_{0,1}$
ORG 0 vs ORG 2 (70%)	Both	All	0.0088	At 1% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0089	At 1% level	ORG 2 (70%) > ORG 1
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	>0.15	No	Accept $H_{0,4}$

Table 3-2 - Results for the average number of minimum safe separation infringements

The results show that ORG 2 with 70% 4-D FMS equipped aircraft had a statistically significant greater number of infringements than ORG 0 and ORG 1 when all controllers, both sectors and all traffic volumes were considered. No significant difference was found between ORG 0 and ORG 1 nor between ORG 2 with 30% and 70% 4-D FMS equipped aircraft.

To illustrate the size of the difference, Figure 3-2 below displays, for each organisation, the percentage of runs in which there were 0, 1, 2 etc minimum separation infringements in sector 10. The numbers shown are for medium and high traffic volumes only, so that a valid comparison can be made with ORG 2 (30%) (for which no low traffic volumes were used). Percentages, rather than absolute numbers of infringements, are used to eliminate biases which would be introduced due to the variable number of total runs included in the analysis under each of the different organisations.

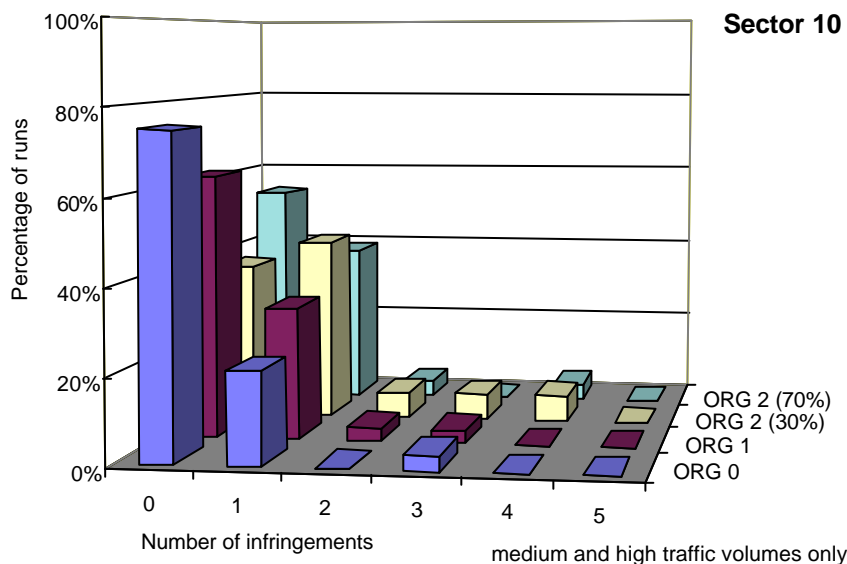


Figure 3-2 - Distribution of the number of safe separation infringements (sector 10)

Figure 3-2 shows that ORG 2 had a lower percentage than ORG 0 and ORG 1 of runs with no infringements and a higher percentage of runs with one infringement. ORG 2 also had higher proportions of runs with 2 or more infringements, although the number of runs here is small: the column for two infringements in ORG 2 (30%) represents only 1 run.

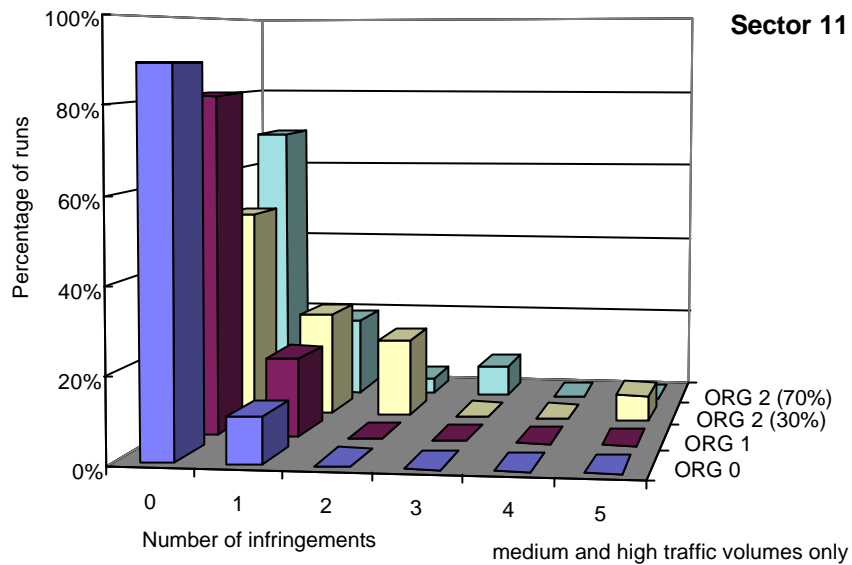
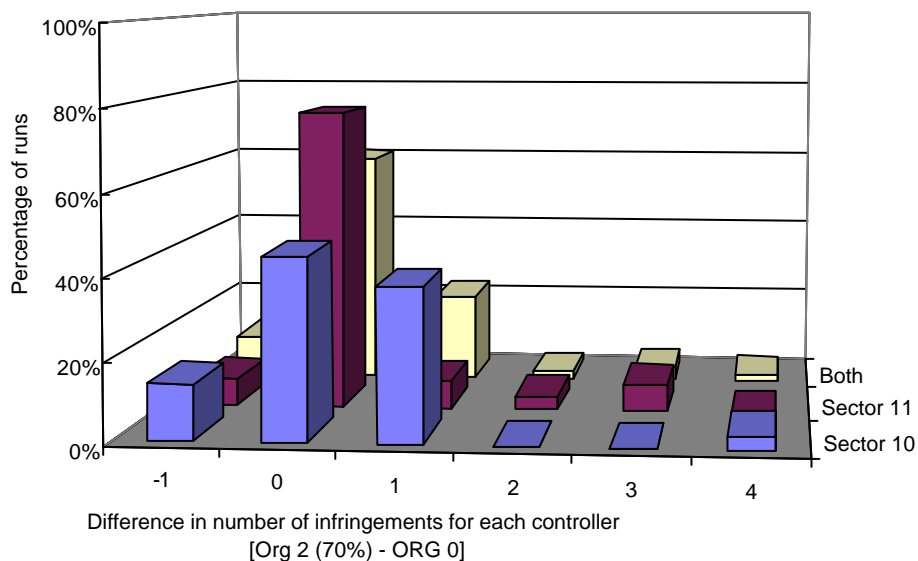


Figure 3-3 - Distribution of the number of safe separation infringements (sector 11)

Figure 3-3 shows the same results for sector 11. A similar pattern emerges, although the percentage of runs with no infringements was higher in sector 11, reinforcing the conclusions of Section 2.2.1.3 which suggested that sector 10 was more difficult to control than sector 11.

By aggregating weeks, these figures hide the responses of individual controllers. In fact, 60% of all controllers had the *same* number of infringements in each ORG 2 (70%) traffic sample as in the matched traffic sample in ORG 0. Indeed, more than 90% of controllers had fewer than 2 more infringements in ORG 2 (70%) than in ORG 0. These observations are illustrated by Figure 3-4. So, although the difference in numbers of infringements is statistically significant, it is small.



*Figure 3-4 - Difference in number of infringements by each controller in each sample, comparing ORG 0 to ORG 2 (70%)*

Further examination of the data shows large variation in the number of infringements both between weeks, controllers and sectors. For example, in two weeks, the sector 11 controller had no infringements in either ORG 0 or ORG 2 (70%). At the other extreme, the sector 10 controller in another week had 3 infringements in ORG 0 and 4 in ORG 2 (70%).

### 3.3.2. Duration of minimum safe separation infringements

Table 3-3 below gives the results for the duration of minimum safe separation infringements.

ORGs compared	Sectors	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	Both	All	>0.15	No	Accept $H_{0-1}$
ORG 0 vs ORG 2 (70%)	Both	All	0.0074	At 1% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	Both	All	0.0026	At 0.5% level	ORG 2 (70%) > ORG 1
ORG 2 (30%) vs ORG 2 (70%)	Both	Medium and High	>0.15	No	Accept $H_{0-4}$

*Table 3-3 - Results for the average duration of minimum safe separation infringements*

The results for the average duration of minimum safe separation infringements were similar to those for the number of infringements. No significant difference was found between the duration of infringements for ORG 0 and ORG 1. Similarly, there was no evidence of a significant difference between ORG 2 with 30% 4-D FMS and datalink and ORG 2 with 70% 4-D FMS and datalink. However, ORG 2 (70%) was found to have a significantly greater proportion of time with aircraft infringing safe separation standards than both ORG 0 and ORG 1.

Figure 3-5 below shows a comparison between ORGs of the duration of the longest infringement per run. As in the last section, the values are taken over medium and high traffic volumes only and the same justification for the use of percentages is applicable. For example, 75% of medium and high runs in ORG 0 had a longest infringement of 0 seconds - ie no infringements. (Not surprisingly, the same value is found in Figure 3-2.) The 'Longer than 100 seconds' columns correspond to 2 runs for ORG 0 and 1 run each for the other organisations.

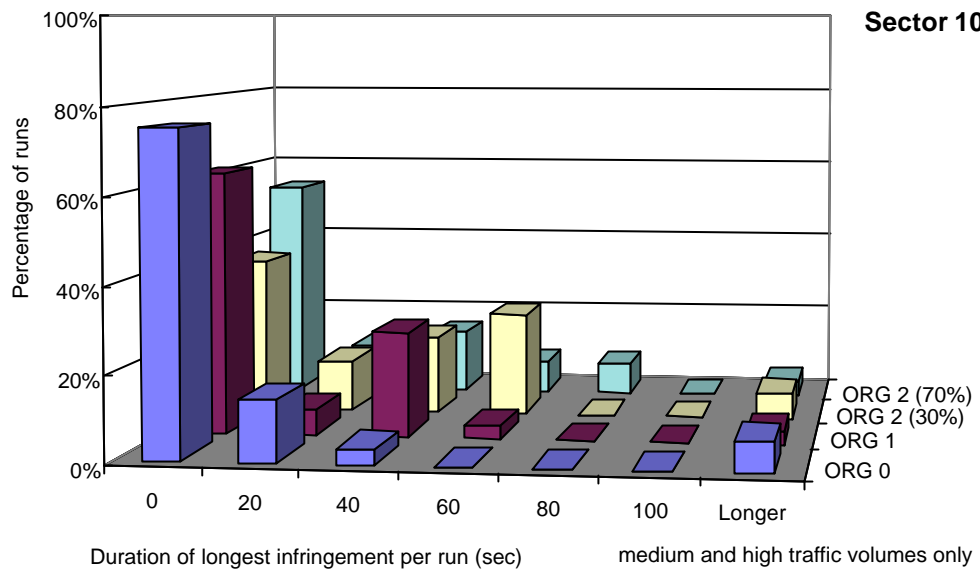


Figure 3-5 - Distribution of the duration of longest infringement per run (sector 10)

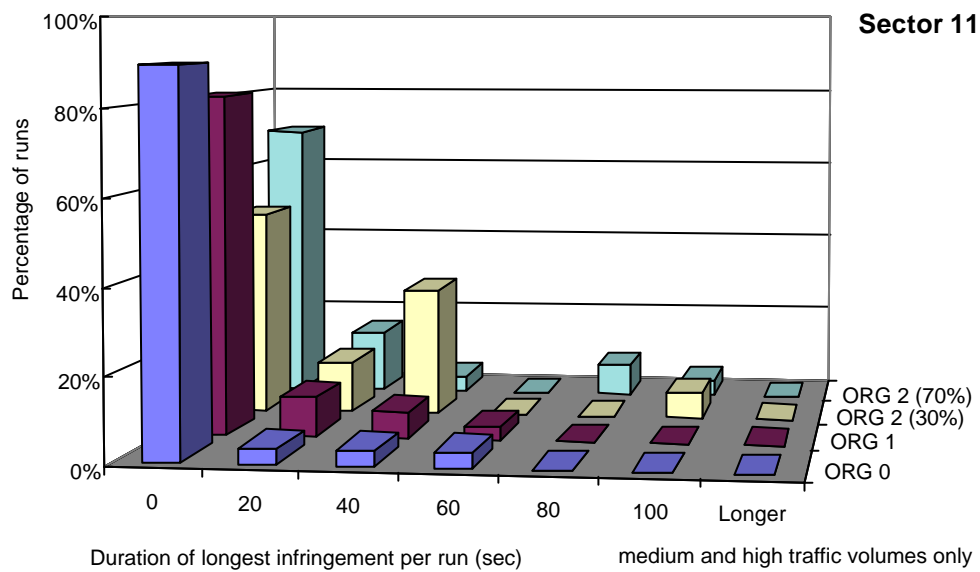


Figure 3-6 - Distribution of the duration of longest infringement per run (sector 11)

Figure 3-6 shows the same results, but for sector 11. Note again that there are fewer infringements in total for sector 11 than for sector 10 (e.g. just 3 in ORG 0). With such limited data, comparisons of the duration of infringements should be treated with caution.

### 3.3.3. 'Serious' separation infringements

The duration of infringements where both the horizontal and vertical separations were less than 50% of the minimum safe standards simultaneously were also investigated. In this way less serious minimum safe separation infringements could be filtered out from the analysis.

It was found that ORG 2 with 70% 4D FMS and datalink had significantly more of these ‘serious’ separation infringements than ORGs 0 and 1. No significant difference was found between ORG 0 and ORG 1 and between ORG 2 with 30% 4D FMS and datalink and ORG 2 (70%). This is the same as for the other infringement cases (Sections 3.3.1 and 3.3.2).

### 3.4. PERCENTAGE OF TIME ‘PICTURE’ MAINTAINED BY CONTROLLERS

As part of the post-run questionnaire, the controllers were asked the percentage of the time that they had maintained the ATC picture. This measure is, therefore, a subjective measure of both capacity and the controller’s workload. An increase in the percentage of time the picture was maintained indicates a perceived increase in ability to manage the sector airspace.

Table 3-4 presents the results for the percentage of time the controllers maintained the picture. The null hypotheses  $H_{0,1}$  to  $H_{0,4}$  (Section 3.1) were examined using the Wilcoxon-Mann-Whitney test rather than Wilcoxon Signed Ranks, because there was a large number of pairs with the same value.

As can be seen from Table 3-4, there were few differences between the organisations: the only result that was statistically significant was that in ORG 2 (70%) the picture was maintained for longer than in ORG 2 (30%). There was a trend (at the 10% significance level and shown in square brackets) for the Tactical Controllers to maintain the picture more of the time in ORG 0 than in ORG 1<sup>2</sup>. The Planning Controllers did not show any difference between ORG 0 and ORG 1. There was also a trend for the Planning Controllers to maintain the picture more often in ORG 0 than in ORG 2 (70%).

ORGs compared	Traffic Volume	Controller Role	P	Significant?	Conclusion
ORG 0 vs ORG 1	All	Both	0.0646	At 10% level	[ORG 0 > ORG 1]
		Tactical	0.0728	At 10% level	[ORG 0 > ORG 1]
		Planner	>0.15	No	Accept $H_{0,1}$
ORG 0 vs ORG 2 (70%)	All	Both	0.0601	At 10% level	[ORG 0 > ORG 2 (70%)]
		Tactical	>0.15	No	Accept $H_{0,2}$
		Planner	0.0892	At 10% level	[ORG 0 > ORG 2 (70%)]
ORG 1 vs ORG 2 (70%)	All	Both	>0.15	No	Accept $H_{0,3}$
ORG 2 (30%) vs ORG 2 (70%)	Medium and High	Both	0.0378	At 5% level	ORG 2 (70%) > ORG 2 (30%)
		Tactical	0.0557	At 10% level	[ORG 2 (70%) > ORG 2 (30%)]
		Planner	>0.15	No	Accept $H_{0,4}$

Table 3-4 - Results for the percentage of time the ‘picture’ is maintained

Figure 3-7 - and Figure 3-8 below illustrate the results given in the table. The numbers of occasions on which the picture is reported as being maintained for less than 100% of the time is small (only 25 cases out of 237 responses). The figures support the difference between the

<sup>2</sup> This is not apparent in Figure 3-7 - because the figure omits the low volume traffic samples.

Tactical and Planning Controllers identified as (not statistically significant) trends in Table 3-4: comparing ORG 0 and ORG 2 (70%), the Planning Controller tends to lose the picture less in ORG 0 and there is no difference for the Tactical Controller.

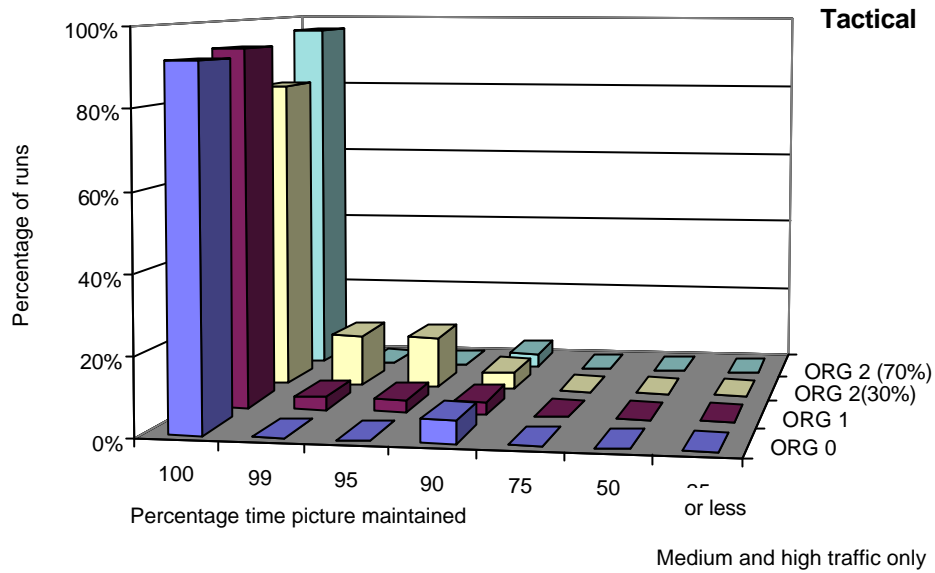


Figure 3-7 - Percentage of time the 'picture' is maintained by Tactical Controllers

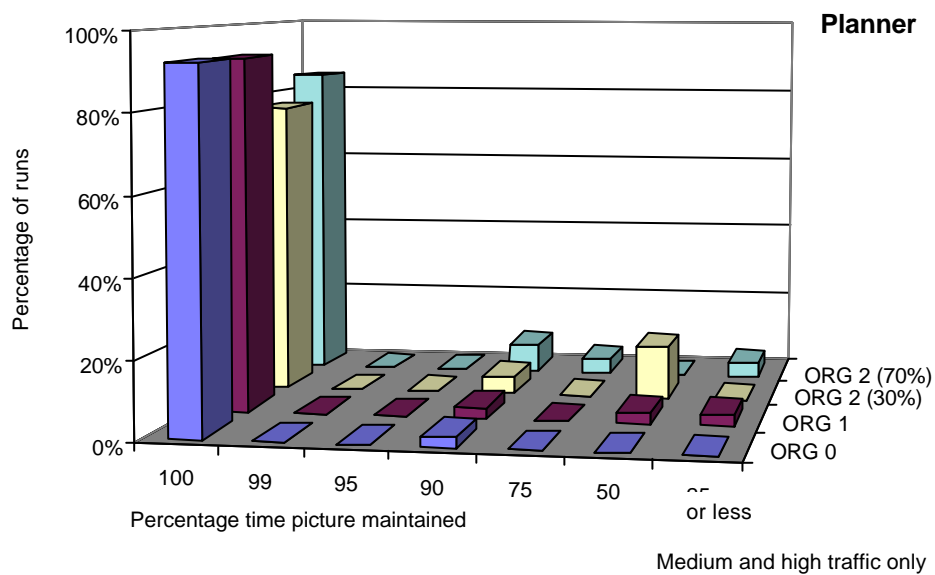


Figure 3-8 - Percentage of time the 'picture' is maintained by Planner Controllers

### 3.5. STCAs

No evidence was found to suggest a significant difference in the number of STCAs per run or in the percentage of time the STCA was activated per run, between any of the ORGs. This is emphasised by the values shown in Figure 3-9 below for the variability of the number of

STCAs per run: no visible trend exists. Note that the values in the figure are taken over medium and high traffic volumes only.

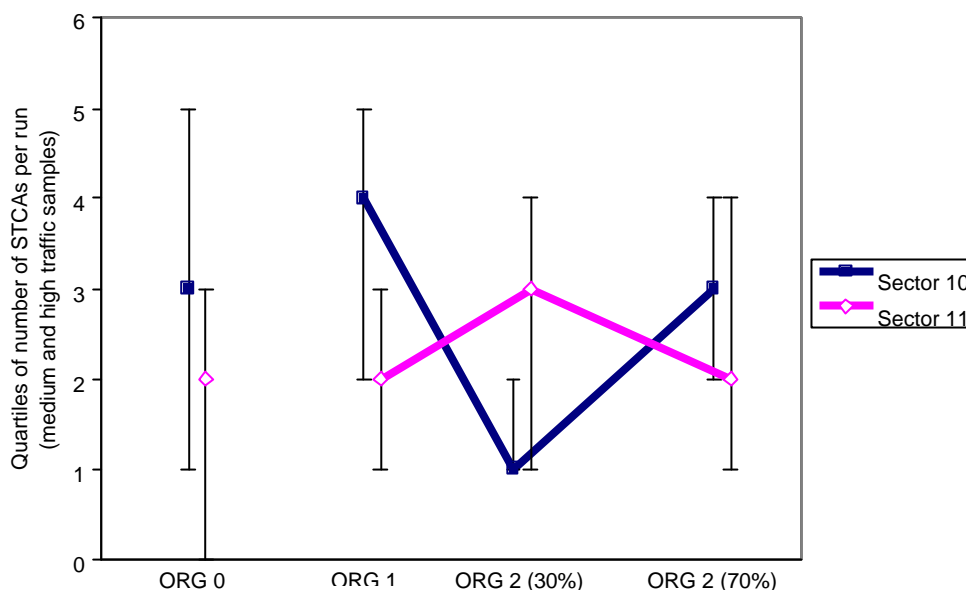


Figure 3-9 - Comparison between ORGs and sectors of the number of STCAs per run

### 3.6. DISCUSSION OF RESULTS

ORG 2 with 70% 4-D FMS and datalink-equipped aircraft had statistically significantly more minimum separation infringements than ORG 0 (or ORG 1). However, the difference is small: a controller usually had the same number of infringements in both (see Figure 3-4). There were differences between the controllers, especially between sector 10 and sector 11 controllers.

A similar pattern emerged in the duration of infringements and in the number of 'serious' infringements. The reasons for these differences between controllers and sectors might have lessons for future PHARE trials, but it is beyond the scope of the present analysis to identify them or measure their significance.

Further analysis of the data, video and R/T recordings might allow more detailed lessons to be learned as to why particular minimum separation infringements occurred or might explain the individual reasons for the few, long infringements. However, if further analysis is undertaken, care is needed to avoid arguing from a few detailed observations (of video or R/T recordings) for significant changes to the PATs or other aspects of the system. There are many possible explanations for the differences between controllers and sectors and it would be difficult to distinguish some of these from the available data. They include: insufficient training; ineffective use of tools or other parts of the PD/1 system; problems with the facility (e.g. system speed of response, screen size or the R/T and phone system); and different sector route structures and shape.

There were few differences in the percentage time that controllers maintained the picture, although there is some evidence of a trend for the Planning and Tactical Controllers to differ. This is consistent with the differences between Planning and Tactical Controllers seen for a number of other measures (eg workload, see Figure 2-1 or Figure 2-2). Further analysis might

be able to identify whether the trend (at the 10% level of significance) for the Planning Controller to lose the picture (in ORG 2 (70%) as compared to ORG 0) is correlated with a corresponding increase in workload. However, it would be difficult with the data available to suggest whether any such correlation was causal in one direction or the other (ie whether increased workload led to loss of picture, or vice versa, or neither). The trend is consistent with observations (Annexes D and F) suggesting a lack of confidence in tool use. This may be a result of the relatively short training period.

There was no evidence of any difference between the ORGs of the frequency of STCAs. This suggests that the presence of the advanced tools did not significantly assist the controllers in planning ahead to avoid STCAs. Again, it would be difficult to explain whether the reason for this was insufficient training in the use of the tools; ineffective use of the tools or other parts of the PD/1 system; frustration with the user interface of the tools; or a lack of trust by the controllers in the predictions made by the tools.

## 4. QUALITY OF SERVICE

### 4.1. INTRODUCTION

The quality of service provided to the airlines is an extremely difficult parameter to measure. The two most common measures of quality of service, fuel used and delays incurred, could not be examined in PD/1. One of the reasons for this is that real-time ATC simulations used for research do not generally have a highly detailed representation of the aircraft's performance - thus prohibiting accurate examination of fuel usage. In addition, the PD/1 scenario examined only a relatively small proportion of the complete gate-to-gate flight, thus prohibiting an examination of aircraft delays.

For the purpose of this study, the quality of service to the airlines was examined by analysing the extent to which the aircraft were allowed to follow their desired flight plans. The following objective measures were therefore adopted as indicators of the quality of service offered to airlines by each ORG:

- percentage of time aircraft were at their requested cruise flight level (RCFL) and within one flight level of the requested level;
- the average time spent by aircraft in the sector.

These measures were used to test the following null hypotheses:

- the quality of service of ORG 0 is the same as that of ORG 1 ( $H_{0-1}$ );
- the quality of service of ORG 0 is the same as that of ORG 2 with 70% 4-D FMS and datalink-equipped aircraft ( $H_{0-2}$ );
- the quality of service of ORG 1 is the same as that of ORG 2 with 70% 4-D FMS and datalink-equipped aircraft ( $H_{0-3}$ );
- the quality of service of ORG 2 with 30% 4-D FMS and datalink-equipped aircraft is the same as that of ORG 2 with 70% 4-D FMS and datalink-equipped aircraft ( $H_{0-4}$ ).

Throughout the following analysis the Wilcoxon Signed Ranks test has been used. Results at the 10% level of significance have been included only to display trends and do not indicate that the null hypothesis is rejected.

The results presented in this section were derived in a different manner from the other results presented in this report and therefore require further explanation. When conducting this analysis, the results have to be examined on a *per aircraft* basis, as opposed to the per run basis of the other results examined herein. When conducting, for example, an analysis of the number of tactical interventions, it is possible to average the results per controller per run; however, when looking at the time spent in the sector (or percentage of time an aircraft is at its RCFL), it is more sensible to analyse the data on a per aircraft basis. By dint of the different route lengths in the two sectors, each individual aircraft must be examined for each organisation - and the results for that aircraft compared between organisations. Hence, over the eight weeks of the trials, the average time that each particular aircraft took to traverse each sector (or the percentage of time that each aircraft was at its RCFL) was compared between the ORGs. Therefore, there are many data points available in this analysis, leading to an ability to resolve a small difference as being a statistically significant result, even though it may be of no practical importance.

As in Sections 2 and 3, **P** signifies the probability of obtaining the observed difference, or a more extreme observation, by chance if the null hypothesis,  $H_0$  were indeed true, where  $H_0$  is the hypothesis that there is no difference between the ORGs. Note that by ORG a > ORG b, it is meant that the measures for ORG a were **significantly** greater than those for ORG b and this difference is unlikely to have occurred by chance.

The results are given in the following subsections and are discussed in Section 4.4.

#### 4.2. PERCENTAGE OF TIME AIRCRAFT AT REQUESTED CRUISE FLIGHT LEVEL

The percentage of time that the aircraft were at their requested cruise flight level (RCFL) may be considered as a measure of the quality of service, since it gives the proportion of time the aircraft were allowed to maintain their flight plan's optimum altitude. However, when analysing this parameter a number of additional factors must be considered.

There may be many cases where an aircraft is not allocated its requested cruise flight level, but the quality of service to the aircraft is not significantly reduced. If an aircraft is requesting, for example FL 310, but this cannot be granted for the complete flight due to tactical control reasons, it may well be preferable for FL 290 to be granted for the sector crossing - rather than FL 290 for a short period of time, followed by a climb to FL 310, followed by a descent to FL 290 for sector exit. Therefore, in addition to the results for the percentage of time the aircraft were at their requested cruise flight level, the percentage of time that the aircraft were within 2000' of their requested cruise flight level was also examined.

Finally, if the requested cruise flight level were maintained at the cost of breaking standing agreements, then this would not be a good measure of quality of service. However, where standing agreements were broken, it was found that aircraft were on average below the standing agreement flight level. Therefore the measure is valid, and perhaps conservative, as a standard of quality of service. An increase in the percentage of time can be interpreted as an increase in the quality of service to the airlines.

Table 4-1 below gives the results for the percentage of time aircraft were at their cruising flight level and Table 4-2 for when the aircraft were within 2000' of their requested cruise flight level. The null hypothesis  $H_{0,1}$  to  $H_{0,4}$  (Section 4.1) were examined using the Wilcoxon Signed Ranks test on a per aircraft basis.

ORGs compared	Sector	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	10	All	>0.15	No	Accept $H_{0-1}$
	11	All	>0.15	No	Accept $H_{0-1}$
ORG 0 vs ORG 2 (70%)	10	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
	11	All	0.0456	At 5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	10	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 1
	11	All	0.0188	At 2% level	ORG 2 (70%) > ORG 1
ORG 2 (30%) vs ORG 2 (70%)	10	Medium and High	>0.15	No	Accept $H_{0-4}$
	11	Medium and High	>0.15	No	Accept $H_{0-4}$

*Table 4-1 - Results for the percentage of time the aircraft were at their requested cruise flight level*

ORGs compared	Sector	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	10	All	0.0420	At 5% level	ORG 0 > ORG 1
	11	All	>0.15	No	Accept $H_{0-1}$
ORG 0 vs ORG 2 (70%)	10	All	0.00345	At 0.5% level	ORG 2 (70%) > ORG 0
	11	All	0.0493	At 5% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	11	All	0.00916	At 1% level	ORG 2 (70%) > ORG 1
	10	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 1
ORG 2 (30%) vs ORG 2 (70%)	10	Medium and High	>0.15	No	Accept $H_{0-4}$
	11	Medium and High	0.0993	At 10% level	[ORG 2 (70%) > ORG 2 (30%)]

*Table 4-2 - Results for the percentage of time aircraft were within 2000' of their requested cruise flight level*

Figure 4-1 below displays the percentage of aircraft which: always maintained their RCFL throughout sector 10 or throughout sector 11; never reached their RCFL, or flew at their RCFL for some, but not all, of their transit of sector 10 or 11. Figure 4-2 gives the same information for aircraft within 2000' of their RCFL. Both figures include data from all runs within each organisation, i.e. low, medium and high for ORGs 0, 1 and 2 (70%) and medium and high only for ORG 2 (30%).

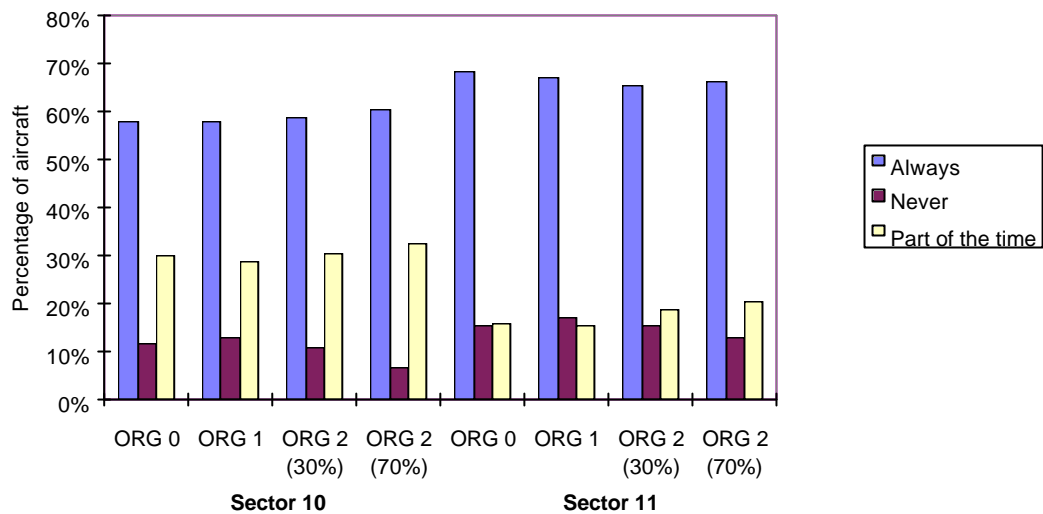


Figure 4-1 - Comparison between ORGs of the average percentage of time aircraft were at their requested cruising flight level

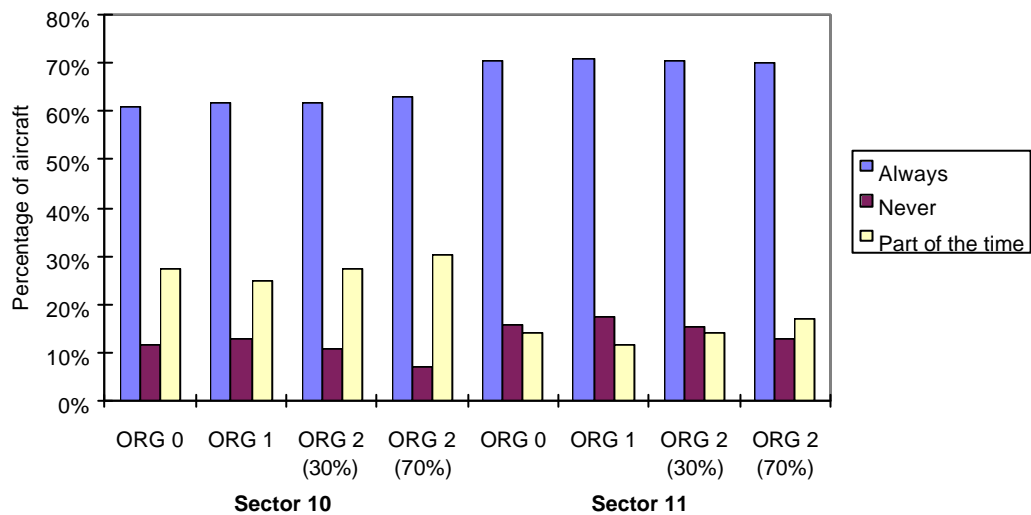


Figure 4-2 - Comparison between ORGs of the average percentage of time aircraft were within 2000' of their requested cruising flight level

As can be seen from the tables above, aircraft (in both sectors) in ORG 2 (70%) spend the highest percentage of time at their requested cruise flight level (or at least within one flight level of it) compared to both ORG 0 and ORG 1, by a statistically significant margin. This must be attributable to the introduction of 4-D FMS and datalink, because there is no significant difference between ORG 0 and ORG 1 in most cases, and in fact ORG 0 has a

significantly higher percentage of time than ORG 1 that aircraft are allowed to fly within 2000' of their cruise flight level in sector 10.

Although the difference between ORG 2 (30%) and ORG 2 (70%) is not significant there is a trend in sector 11 in favour of the 70% mix for aircraft within 2000' of their RCFL. This is shown in Table 4-2 as a result at the 10% level of significance.

In terms of overall figures, aircraft in ORG 0 spend 76.5% of their time within 2000' of their RCFL in sector 10, compared to 76.8% and 83.7% in ORGs 1 and 2 (70%) respectively. The figures for sector 11 are generally higher at 79.3%, 80.0% and 83.9% for ORGs 0, 1 and 2 (70%) respectively. Hence, although the differences are statistically significant, the actual scale of the difference is relatively small.

#### 4.3. AVERAGE TIME SPENT BY AIRCRAFT IN SECTOR

In order to reduce delays and to control fuel consumption, aircraft would prefer not to spend any greater time passing through the sector than was originally planned. Additionally, they would not like to pass through the sector at too high a speed, since this would increase fuel burn and probably incur delays further along the route (assuming that a previous delay was not being reduced).

To study this parameter in the PD/1 scenario is difficult, since there is relatively little time for delays to be incurred in either sector 10 or 11 and there are no holding patterns in either sector. However, it was decided to investigate the average time spent in each sector to see if significant differences did occur. If so, then the reason for such differences could be investigated.

Table 4-3 below gives the results for the average time spent by aircraft in the sector. The null hypothesis  $H_{0.1}$  to  $H_{0.4}$  (Section 4.1) were examined using the Wilcoxon Signed Ranks test on a per aircraft basis.

ORGs compared	Sector	Traffic Volume	P	Significant?	Conclusion
ORG 0 vs ORG 1	10	All	0.0967	At 10% level	[ORG 1 > ORG 0]
	11	All	>0.15	No	Accept $H_{0.1}$
ORG 0 vs ORG 2 (70%)	10	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
	11	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 0
ORG 1 vs ORG 2 (70%)	10	All	0.0234	At 0.1% level	ORG 2 (70%) > ORG 1
	11	All	<0.001	At 0.1% level	ORG 2 (70%) > ORG 1
ORG 2 (30%) vs ORG 2 (70%)	10	Medium and High	0.0134	No	Accept $H_{0.4}$
	11	Medium and High	0.0383	No	Accept $H_{0.4}$

Table 4-3 - Results for the average aircraft time spent in each sector

As can be seen from Table 4-3, there is a significant increase in the time spent in both sectors 10 and 11 from ORG 0 to ORG 1 and ORG 2 (70%). No significant difference is found

between ORGs 2 (30%) and ORG 2 (70%) or between ORGs 0 and 1, although there is a trend for ORG 1 aircraft to spend longer in sector 10 compared to ORG 0 aircraft. This is shown in Table 4-3 as a result at the 10% level of significance.

Figure 4-3 below illustrates the average time spent by aircraft in sectors 10 and 11. The data are averaged across high and medium traffic volumes only. Given the explanation in 4.1 of how the analysis was conducted on a per aircraft basis it should be noted that this figure is for illustration only and hides much of the variation between ORGs. However, the overall average times spent by aircraft in each sector as illustrated is a useful measure to consider, being one element of fuel efficiency from the airlines' perspective.

So, in terms of these relatively crude averages, the introduction of tools and 70% 4D FMS and datalink-equipped aircraft in ORG 2 (70%), as compared to the baseline ORG 0, decreases the average time aircraft spend in sector 10 by 0.04% (or 0.4 seconds). However, the pattern is reversed in sector 11 where the average time in sector increases by 4.9% (or 25.6 seconds).

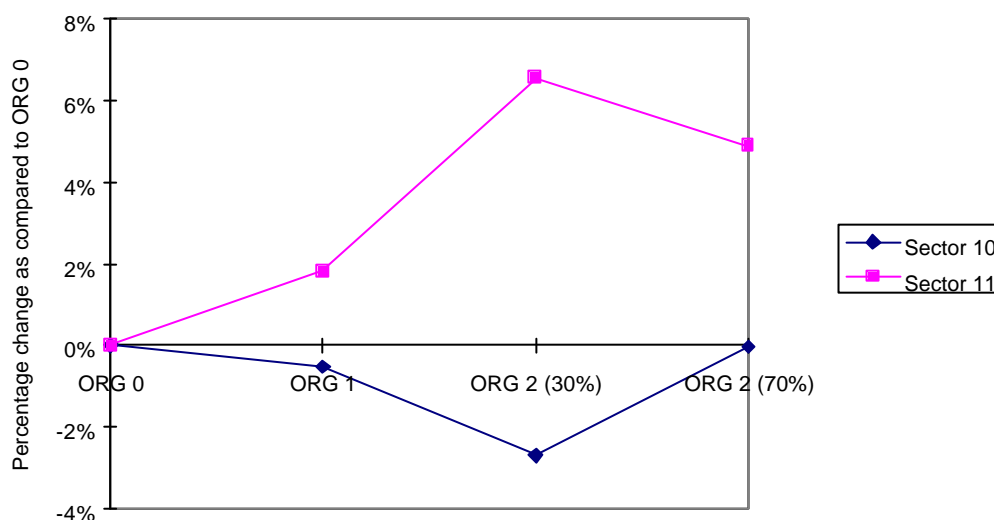


Figure 4-3 - Comparison between ORGs of the average time spent in sector by aircraft (shown as a percentage change over ORG 0)

#### 4.4. SUMMARY AND DISCUSSION OF RESULTS

In terms of quality of service to airlines, ORG 2 (70%) is significantly better than the baseline organisation, in that aircraft spent significantly longer at (or near) their requested cruise flight levels (RCFL). There was little difference, by this measure, between ORG 0 and ORG 1, so the effect must have arisen either due to the datalink or the 4-D FMS. The likely reason for this is that aircraft with 4D FMS negotiated and followed their preferred trajectories with more precision than 3-D FMS equipped aircraft. If this effect held operationally, it could provide a direct benefit to airlines, to justify fitting datalink and 4-D FMS equipment.

The illustrations of time in sector, Figure 4-1 and Figure 4-2 suggest a number of hypotheses relating to the effects of the different sector shapes, routes and standing agreements. These have not been formally tested, but include the following:

- In sector 11, for which many east-west crossing aircraft spend a relatively short time in the sector, more aircraft always fly within 2000' of their RCFL as compared to sector 10. However, more aircraft never fly within 2000' of their RCFL in sector 11 compared to sector 10, indicating that there it is more difficult for aircraft to reach within 2000' of their RCFL during the short transit time of sector 11.
- The standing agreements may also influence the results. sector 11 aircraft which are inbound to Amsterdam (or outbound) are subject to standing agreements at both the easterly and westerly sector boundaries, hence may never be able to get within 2000' of their CRFL. The longer sector 10 routes do not have any similar 'follow on' standing agreements and as shown in Figure 4-2 more aircraft in sector 10 than in sector 11 reach within 2000' of their CRFL for at least some of the time.

Rather more difficult to interpret is the significant increase in time spent in each sector which the advanced organisations represent, compared to the baseline. The picture is confused further by two effects: the averages showed slightly different results to the comparisons aircraft by aircraft; and the differences in time were small (0.4-25.6 seconds), but significant by virtue of there being many data points available. It might be that the effect was due to climbs or descents missed, but it is not appropriate to draw conclusions without more detailed investigation. It should also be noted that for most aircraft the transit of sector 10 or 11 is a small part of their journey, hence losses may be counter-balanced by changes elsewhere en route. Further analysis may continue looking at the differences in time spent in the sector between 3-D and 4-D aircraft. The gate to gate time will be measured in PD/3 and will give a much better indication of the quality of service provided to the airlines.

## 5. CONCLUSIONS

When considering the results of the PD/1 trials, it must be remembered that the introduction of computer assistance tools and 4D FMS and datalink equipped aircraft were not expected exactly to replace the functionality of the baseline scenario. Although the overall operational control objectives remained the same, the methods by which they were implemented were changed by the use of the tools and datalink: in particular, the Planning Controller, by doing more work (planning further ahead), was intended in the advanced organisations to reduce the Tactical Controller's conflict resolution workload.

The following summarises the results of the analysis. There is more discussion of the implications of the analysis in the main report, drawing on informal observations made throughout the development of PD/1 (Annex F).

*Workload* - The introduction of the advanced organisations (ie the PATs, the operational concept and the system) significantly increased subjective workload, and the subsequent introduction of 4D FMS and datalink brought it back to (or below) its original level for the Tactical Controller, but not for the Planning Controller. For those components of controller workload measured objectively, the full implementation of tools and datalink caused a significant decrease in all cases. (See Section 2.)

*Capacity* - Under the advanced organisations, the number of safe separation infringements increased, although the difference was small (usually by one infringement or fewer, per sector per run). There was a trend for the percentage of time the 'picture' was maintained by the Planning Controller to decrease, but the percentage was still 100% of the time in nearly 90% of cases. Thus there is little evidence for a change in airspace capacity. (See Section 3.)

*Quality of service* - The introduction of tools and datalink together cause the aircraft to spend more time in the sector but allow aircraft to fly at their requested cruise flight level significantly more of the time. By considering gate-to-gate flight, PD/3 may be able to draw conclusions about these quality of service measures with more confidence. (See Section 4.)

The following themes emerged from the analysis.

*Workload redistribution* - One of the goals of the introduction of computer assistance tools and datalink was to transfer some of the Tactical Controller's workload to the Planning Controller. The switch-over in subjectively-measured workload from the baseline to full implementation of tools and datalink suggest that this goal was met (see Figure 2-1 and Figure 2-2).

*Tool utilisation* - Some of the results are consistent with either the tools being used inefficiently or being under-used, in particular those results that are not the same for all levels of traffic volume. Possible reasons for this may be:

- Frustration experienced: This was a key TLX factor which increased significantly with the introduction of tools and datalink for the Planning Controller (see Figure 2-4). This frustration may stem from shortfalls in the functionality of the tools or a lack of understanding by the controllers of how the tools work.
- Time pressure: Time pressure for the Planning Controllers was significantly greater in ORG 2 (70%), especially in sector 11 (see Table 2-17). This may be due to difficulties with the tools or a result of the airspace considered: sector 11 is narrow and there is relatively short notice of westbound traffic.

- Confidence: The trend for a decrease in the percentage of time that the picture was maintained (Table 3-4) is consistent with observations (Annexes D and F) suggesting a lack of confidence in tool use. This may be a result of the relatively short training period.

Further research may be necessary to clarify how effectively the tools were used, and to determine whether the controllers found specific tools frustrating to use.

One possible interpretation of the workload changes is that the introduction of computer assistance tools implied a qualitative change in the work of the controllers: they now had to plan further ahead. This extra burden and the accompanying frustration of using unfamiliar and perhaps slightly inflexible tools was mitigated for the Tactical Controller, but not for the Planning Controller, by the introduction of 4-D FMS and datalink aircraft.

It should be noted that although the analysis results point towards proving the themes above, not all of the relevant null hypotheses were tested explicitly. Further examination of the data will be necessary to understand the workload results, for example to explore the effectiveness of the Tactical-Planning Controller teams. Only further research will answer the question of whether the Tactical Controller still subconsciously performed a planning role (duplicating the Planning Controller's work) and whether the balance of work between the Tactical and Planning Controller roles can be optimised.

## 6. ABBREVIATIONS

3-D	three dimensional (latitude, longitude and altitude)
4-D	four dimensional (latitude, longitude, altitude and time)
ATC	air traffic control
FL	flight level
FMS	flight management system
H <sub>0</sub>	null hypothesis
HIPS	highly interactive problem solver
ISA	instantaneous self assessment
NATS	National Air Traffic Services Ltd (UK)
ORG	organisation
PATs	PHARE advanced tools
PD/1	PHARE Demonstration 1
PHARE	Programme for Harmonised ATC Research in Eurocontrol
R/T	radio/telephony
RCFL	requested cruise flight level
STCA	short term conflict alert
TLX	task load index
WMW	Wilcoxon-Mann-Whitney; a statistical test for differences between sampled data
WSR	Wilcoxon Signed Ranks; a statistical test for differences between sampled data