EUROPEAN ORGANISATION
FOR THE SAFETY OF AIR NAVIGATION

EUROCONTROL

This Document is issued as EATM Method and Tool. The contents are not mandatory. They provide information and explanations or may indicate best practice.

A Tool for the Assessment of the Impact of Change in Automated ATM Systems on Mental Workload

Edition Number	:	1.0
Edition Date	:	31.08.2004
Status	:	Released Issue
Intended for	:	EATM Stakeholders

EUROPEAN AIR TRAFFIC MANAGEMENT

DOCUMENT CHARACTERISTICS

TITLE

A Tool for the Assessment of the Impact of Change in Automated ATM Systems on Mental Workload

Document Identifier		
	EATM Infocentre Reference:	040201-12
Document Identifier	**Edition Number:**	1.0
HRS/HSP-005-REP-03	**Edition Date:**	31.08.2004

Abstract

This document reports on the rationale and development process undertaken to construct a tool for assessing the impact of change in automated Air Traffic Management (ATM) systems on mental workload. The tool has been named '**A**ssessment of the **I**mpact on **M**ental workload (AIM)'. Findings and conclusions from a pilot validation study carried out to examine the preliminary validity of AIM are also provided.

This deliverable has been developed as part of the 'Solutions for Human-Automation Partnerships in European ATM (SHAPE)' Project conducted within the Human Factors Sub-Programme (HSP) of the EATM Human Resources Programme (HRS) by the Human Factors Team of the EUROCONTROL Human Factors Management Business Division (DAS/HUM) under the supervision of the EATM Human Resources Team (HRT).

The complete AIM Tool Set is available within the SHAPE Toolkit.

Keywords

Air Traffic Control (ATC)	Air Traffic Management (ATM) system	Automation
Mental workload	Computer assistance	Impact of automation
Human factors	Cognitive functions	Mental resources
Rating scales	Assessment of the Impact on Mental workload (AIM)	Multi-task performance
Processing capacity	Assessment tool	Subjective technique

Contact Persons	Tel	Division
Oliver STRAETER	+32-2-729 5054	Safety and Security Management Business Division (DAS/SSM)
Manfred BARBARINO	+32-2-729 3951	Human Factors Management Business Division (DAS/HUM)

Author

I. Low

STATUS, AUDIENCE AND ACCESSIBILITY

Status		Intended for		Accessible via	
Working Draft	☐	General Public	☐	Intranet	☐
Draft	☐	EATM Stakeholders	☑	Extranet	☐
Proposed Issue	☐	Restricted Audience	☐	Internet (www.eurocontrol.int)	☑
Released Issue	☑	*Printed & electronic copies of the document can be obtained from the EATM Infocentre (see page iii)*			

ELECTRONIC SOURCE

Path:	G:\Deliverables\HUM Deliverable pdf Library\	
Host System	**Software**	**Size**
Windows_NT	Microsoft Word 8.0b	

EATM Infocentre
EUROCONTROL Headquarters
96 Rue de la Fusée
B-1130 BRUSSELS

Tel: +32 (0)2 729 51 51
Fax: +32 (0)2 729 99 84
E-mail: eatm.infocentre@eurocontrol.int

Open on 08:00 - 15:00 UTC from Monday to Thursday, incl.

DOCUMENT APPROVAL

The following table identifies all management authorities who have successively approved the present issue of this document.

AUTHORITY	NAME AND SIGNATURE	DATE
SHAPE Project Leader	O. STRAETER	12. 07.2004
Chairman HRT Human Factors Focus Group (HFFG)	V.S.M. WOLDRING	12.07.04.
Manager EATM Human Resources Programme (HRS-PM)	M. BARBARINO	12. 07. 2004
Chairman EATM Human Resources Team / Programme Steering Group (HRT/PSG)	A. SKONIEZKI	19.07.04
Director ATM Programmes (DAP)	G. PAULSON	27.7.04

DOCUMENT CHANGE RECORD

The following table records the complete history of the successive editions of the present document.

EDITION NUMBER	EDITION DATE	INFOCENTRE REFERENCE	REASON FOR CHANGE	PAGES AFFECTED
0.1	01.04.2003		Working Draft	All
0.2	07.04.2003		Draft	All
0.3	2003		Draft 2 for submission to HFFG	All
0.4	24.12.2003		Proposed Issue for HRS-PSG Meeting in January 2004 (document configuration and editorial changes)	All
1.0	31.08.2004	040201-12	Released Issue (agreed on 28-29.01.2004) (final document configuration and editorial adjustments)	All

CONTENTS

APPENDICES

Page intentionally left blank

EXECUTIVE SUMMARY

This document describes the development of a human factors tool for the assessment of the impact on Mental Workload (MWL) of ATM systems. It also reports on the pilot study conducted to examine the preliminary validation of the tool.

This document has been developed as part of a larger project entitled ‘Solutions for Human-Automation Partnerships in European ATM (SHAPE)’ being carried out by the Human Factors and Manpower Unit (DIS/HUM) of EUROCONTROL, today known as the Human Factors Management Business Division (DAS/HUM).

SHAPE investigates specific human factors topics concerned with automation in ATM, namely trust (see EATM, 2003a,b,c), situation awareness (see EATM, 2003d), teamwork (see EATM, 2004a), experience and ageing (see EATM, 2003e, 2004b), future controller skill-set requirements (see EATM, 2004c), recovery from system failure (see EATM, 2004d), and a tool for the assessment of the impact of change in automated ATM systems on mental workload (covered by this report).

A consortium comprising of UK National Air Traffic Services Ltd (NATS) and Det Norske Veritas (DNV) conducted the work reported in this document.

Section 1, ‘Introduction’, outlines the background to the SHAPE Project, and the objectives and scope of the document.

Section 2, ‘Assessment of the Impact of Automation on Mental Workload’, recaps why a tool is required to assess the impact on MWL. It also states the objectives of the work package and briefly outlines the work carried out in this work package.

Section 3, ‘Overview of AIM’, provides an overview of the tool, its scope, a short description of the tool and what the AIM Tool consists of.

Section 4, ‘Rationale behind the Development Process of AIM and its Content’, describes how the ‘**A**ssessing the **I**mpact on **M**ental workload (AIM)’ Tool and its different components and versions were developed and constructed. It also explains the rationale behind the various components of the tool. The final product, that is the MWL assessment tool, is available separately.

Section 5, ‘Procedures for Using the AIM Tool’, details the data collection process required for AIM as well as the scoring process and interpretation of AIM. It also includes instructions for administering AIM.

Section 6, ‘Construct Validity, Internal Consistency Reliability and General Usability of AIM’, discusses the preliminary findings on the use of AIM, and its construct validity and internal reliability.

Section 7, ‘Key Achievements of the Work Package’, summarises the work package achievements, the tool, a discussion on the tool, recommendations for further work on AIM, and the future use of AIM.

References, Further Reading, a list of the Abbreviations and Acronyms used in this document and their full designations, as well as the names of those who contributed to this publication can be found at annex.

The literature reviews conducted in the work package and other theoretical background can be found in appendices to the document. Copies of the AIM measure and its recording forms, decision tree rating scales and instruction sheets are also appended (see Appendices F, L, M, N and O). A description of the pilot study and the preliminary results are also appended (see Appendix P).

1. INTRODUCTION

1.1 Purpose

The purpose of this document is to provide a human factors tool for assessing the impact of ATM system change on Mental Workload (MWL), which includes ATM computer-assistance tools and other forms of automation support.

1.2 Scope

The document is intended to provide a description of the development of a human factors tool for the assessment of the impact on MWL from ATM systems. It also reports on the pilot study conducted to examine the preliminary validation of the tool.

In addition, the deliverable aims to provide a resource in the form of a practical tool for the assessment of the impact on mental workload for EUROCONTROL project leaders and other project staff who are concerned with MWL assessment. The assessment of the impact on MWL tool is intended principally for deployment in real-time simulations of future ATM systems.

1.3 Background

The work on mental workload presented in this module is embedded in a larger project called ‘Solutions for Human-Automation Partnerships in European ATM (SHAPE)’. The SHAPE Project started in 2000 within the Human Factors Sub-Programme (HSP) of the EATMP Human Resources Programme (HRS) (see EATMP, 2000) conducted by the Human Factors and Manpower Unit (DIS/HUM) of EUROCONTROL, today known as ‘Human Factors Management Business Division (DAS/HUM)’.

SHAPE is dealing with a range of issues raised by the increasing automation in European ATM. Automation can bring success or failure, depending on whether it suits the controller. Experience in the introduction of automation into cockpits has shown that, if human factors are not properly considered, ‘automation-assisted-accidents’ may be the end result. SHAPE has identified the following seven main interacting factors which need to be addressed in order to ensure harmonisation between automated support and the controller:

<u>Trust</u>: The use of automated tools will depend on the controllers' trust. Trust is a result of many factors such as reliability of the system and transparency of the functions. Neither mistrust nor complacency are desirable. Within SHAPE guidelines were developed to maintain a correctly calibrated level of trust (see EATM 2003a,b,c).

Situation Awareness (SA): Automation is likely to have an impact on controllers SA. SHAPE developed a method to measure SA in order to ensure that new systems do not distract controllers' situation awareness of traffic too much (see EATM, 2003d).

Teams: Team tasks and performance will change when automated technologies are introduced (team structure and composition change, team roles are redefined, interaction and communication patterns are altered). SHAPE developed a tool to investigate the impact of automation on the overall team performance with a new system (see EATM, 2004a).

Skill set requirements: Automation can lead to both skill degradation and the need for new skills. SHAPE identifies new training needs, obsolete skills, and potential for skill degradation aiming at successful transition training and design support (see EATM, 2004c).

Recovery from system failure: There is a need to consider how the controller will ensure safe recovery should system failures occur within an automated system (see EATM, 2004d).

Workload: With automation human performance shifts from a physical activity to a more cognitive and perceptual activity. SHAPE develops a measure for MWL, in order to define whether the induced workload exceeds the overall level of workload a controller can deal with effectively (covered by this report).

Ageing: The age of controllers is likely to be a factor affecting the successful implementation of automation. Within SHAPE this particular factor of human performance, and its influence on controllers' performance, are investigated. The purpose of such an investigation is to use the results of it as the basis for the development of tools and guidance for supporting older controllers in successfully doing their job in new automated systems (see EATM, 2003e). An additional report on a questionnaire-investigation throughout the European Civil Aviation Conference (ECAC) area has also been produced (see EATM, 2004b).

These measures and methods of SHAPE support the design of new automated systems in ATM and the definition of training needs. It also facilitates the preparation of experimental settings regarding important aspects of human performance such as potential for error recoveries or impacts of human performance on the ATM capacity.

The methods and tools developed in SHAPE will be compiled in a framework in order to assist the user in assessing or evaluating the impact of new systems on the controller performance, efficiency and safety. This framework will be realised as a computerised toolkit called the ‘SHAPE Toolkit’ and is planned to be available in 2004.

2. ASSESSMENT OF THE IMPACT OF AUTOMATION ON MENTAL WORKLOAD

2.1 Introduction

The role and nature of the Air Traffic Control Officer (ATCO) tasks will almost certainly change as a result of the inclusion of increased automation within the ATM system. The ways in which automation impacts the controller may be quite varied and complex. All of these effects need to be addressed in advance and the human-automation partnership needs to be planned. By doing this early in the development lifecycle of the automation it will be possible to better gauge the impact that the automation will have on the controllers' performance and determine whether the system will in fact achieve its anticipated benefits. Understanding the impacts on the controller more fully and earlier also enables cost benefit trade-offs and different system options to be considered before too much development is done ensuring that the right system choices get made most effectively. Effective resolution of these Human Factors (HF) issues will be key to ensuring the successful design and implementation of this technology in ATM.

2.2 The Need for a Tool to Evaluate Impact of Automation on Mental Workload

Development of technical capabilities of automation has been rapid and opportunities for its implementation widespread. While the potential benefits of automation cannot be disputed, especially economic benefits, trade-offs and problems resulting from the use of automation exist. Two classes of problems have been suggested in the literature (Parasuraman & Riley, 1997; Wickens, Mavor & McGee, 1997; Woods, 1993; Woods, 1996). Some problems may be the result of how the automated device is implemented in practice. These are generally less serious and can be resolved by adequate, well-planned and structured training. The other class of problems that is more insidious may arise from unanticipated interactions between the components[1] of the work system in which the automated system is implemented. One of the problems in this class is the undesirable impact of automation on Mental Workload (MWL).

A taxonomy-based methodology such as the SHAPE Automation Framework described in EATM (2004c) can be used to analyse the automation and predict what impact the automated system might have on the functions and tasks of the controller. The particular value of the framework for predicting workload is that it will allow the prediction of the impact of the automation on the cognitive activities of the controller as well as the operational functions.

[1] The components of the work system consist of the human operator, the automated system and any other system in the workplace.

These predictions will, to some extent, allow the possible MWL impact of the automation on the controller to be identified. However, it is important to be able to confirm predictions made by the framework. It is also important to be able to describe and explain the impact the automated system has on the controller's workload before operational implementation, in order to be able to contribute to the design and validation of the automation system or prepare for the implementation process.

Hence, potential workload problems resulting from the use of automation needs to be examined and understood as the design and/or implementation of the automation progresses. There is a need for an evaluation technique suitable and sufficient for assessing the impact on MWL. The technique needs to take into account the functions being carried out and automated, as well as the type of automation. The technique also needs to consider the different dimensions of mental workload and attempt to indicate workload due to these different dimensions. This will enable more detailed diagnosis of workload issues in order to better inform the design of the automation. The technique required could be an entirely new tool for assessing MWL impact from automation, a modification of an existing tool to fit this purpose, or a combination of existing tools.

2.3 Work Package Objectives and Outline

The objective of the work package is to develop a tool appropriate for assessing MWL in investigations into the impact of automation on controllers. The requirement specifications for such a tool are as follows:

1. <u>Multi-dimensional</u>: Ability to evaluate the different dimensions of mental workload to allow the analyst a more detailed diagnosis of mental workload impacts.

2. <u>Practical and usable</u>: For easy and convenient application during real-time simulations of ATM and in a typical Human Factors Laboratory.

3. <u>Multi-scaled and diagnostic</u>: Consists of measurement sub-scales for each dimension of mental workload. This will enable a profile of MWL to be produced. In addition, the evaluation of workload can be further simplified or focused by using the appropriate sub-scale. The diagnosticity according to the profile of MWL will then enable better insight into the design of automation system with respect to MWL. Design recommendations and changes can be focused and targeted onto the part of the system responsible for the aspects of MWL affected.

4. <u>Situation/design sensitive</u>: Distinguishing load due to situational conditions and due to automated tools.

5. <u>Impact on spare processing capacity</u>: With automation resulting in increased sector capacity and traffic density, the time and mental capacity available to formulate a response to problems and emergencies is reduced. It is therefore important that the tool includes a capability to

indicate if the mental processing capacity of the controller has been affected by the system and thereby increasing the risk of overload.

The options for the work package were:

(i) to construct an entirely new mental workload assessment tool,
(ii) to adapt and modify an existing tool, or
(iii) to produce a tool by combining two or more existing tools.

The work package began with the construction of an approach to guide and structure the tool development process. The early stages of the work package included a literature review of the impact of automation on workload and the various workload concepts. The key concepts and core constructs for the proposed workload evaluation were defined. The literature search was not confined to the aviation industry. The decision was also made on a suitable theoretical model of mental workload in ATM.

A review of existing MWL assessment tools was also carried out. A simple system for assessing the suitability of each tool against defined criteria was constructed. This assessment system enabled the review of current methodologies, such as the Cooper-Harper Rating Scale, the Bedford Workload Scale, the Honeywell Cooper-Harper Workload Rating Scale, the Modified Cooper-Harper Rating Scale, the Subjective Workload Assessment Technique (SWAT), the Zachary/Zachlad Cognitive Analysis, the National Aeronautics and Space Administration Task Load Index (NASA-TLX), the Performance and Usability Modelling in ATM (PUMA) and the Instantaneous Self-Assessment (ISA). The system ensured that the assessment of suitability across methodologies is consistent and fair. As a result of the review, the decision was taken to construct a new tool for assessing the impact on MWL rather than modifying one or a combination of existing tools. The latter half of the work package consisted of the construction of the new tool and iterations of refinements following several review group meetings, a trial use of the tool and a small-scaled pilot study to examine the preliminary validity and reliability of the new tool.

The output is a tool for assessing MWL in investigations into the impact of automation on air traffic controllers.

Page intentionally left blank

3. OVERVIEW OF AIM

This section describes an overview of AIM; its scope, what it is able to do and what it consists of. The overview of AIM provides a useful structure to understand the development process and the rationale behind it. Readers may understand the development process and why certain activities were undertaken and how the product took shape if they had foresight of the tool from the overview of AIM.

3.1 Scope

AIM was designed to provide assessments of the impact on controller MWL from ATM system changes, as well as to meet the requirements specified in Section 2.3. AIM as a tool was designed to be:

1. Multi-dimensional: Ability to evaluate different dimensions of MWL. AIM allows the analyst to determine the MWL due to:

   - different cognitive functions,
   - the demands on different mental resource types.

   It also partitions the MWL assessment according to mental effort required and task difficulty. Mental workload impact due to differences in mental effort required may have different design implications from MWL impact due to task difficulty.

2. Multi-scaled and diagnostic: AIM consists of measurement sub-scales for each dimension of MWL. This will enable a profile of mental workload impact to be produced, that is a profile of MWL due to:

   - different cognitive function groups, e.g. multitasking workload, memory management workload, planning workload or decision-making workload;

   - the demands on different mental resource types, e.g. visual mental resources, spatial mental resource or verbal mental resources.

   In addition, the evaluation of mental workload can be further simplified or focused by using the appropriate individual cognitive function sub-scale. The diagnosticity according to the profile of MWL will then enable better insight into the design of automation system with respect to dimensions of MWL. Design recommendations and changes can be focused and targeted onto the part of the system responsible for the aspects of MWL affected.

3. Practical and usable: For easy and convenient application during real-time simulations of ATM and in a typical Human Factors Laboratory. AIM was designed to require minimal resources to administer. There is no special training required to administer AIM. The AIM Tool Set contains guidelines

for users to decide which version of AIM to use and how to score and interpret the different versions of AIM. The tool set also includes a computer-based tool which will automate the scoring of AIM.

4. Situation/design sensitive: Distinguishing mental workload as per situational conditions or design of automated tools.

5. Impact on spare processing capacity: With ATM system change or implementation of automation resulting in increased sector capacity and traffic density, the time and mental capacity available to formulate a response to problems and emergencies is reduced. It is therefore important that AIM includes a capability to indicate if the ATM system change may have a potential impact on the spare mental processing capacity of the controller and thereby increasing the risk of overload.

3.2 Short Description

AIM is a subjective MWL assessment tool. It requires the subject to rate the mental effort required for task performance and the difficulty of task performance. Both are rated on a seven-point rating scale. Two types of rating scales can be chosen according to experimental design: absolute and relative rating scales. Both allow either absolute or relative judgement of MWL.

The rating scales are embedded in decision trees containing questions that guide the subjects' use of the rating scales to make their MWL ratings. The MWL ratings are made on specific task items or defined cognitive function groups (e.g. multitasking, planning or decision-making tasks). The decision to use either the absolute or relative workload decision tree depends on the design of the trial or simulation in which AIM will be used and the simulation/trial equipment resources available (i.e. time, staff, simulation resources, etc.). Table 1 is a decision table that can be used as a guideline for deciding between absolute and relative workload decision trees.

Table 1: Guideline to decide between absolute and relative decision trees

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Availability of baseline design/equipment in trial/simulation
(current operational system or previous design prototypes will be used in the trial or simulated)</td></tr>
<tr><td>YES</td><td>NO</td></tr>
<tr><td rowspan="2">Controllers are experienced users of baseline design/ equipment (e.g. current operational equipment or previous design prototypes)</td><td>YES</td><td>Absolute workload decision tree</td><td>Relative workload decision tree</td></tr>
<tr><td>NO</td><td>Absolute workload decision tree</td><td>Not appropriate to use AIM</td></tr>
</table>

There are three versions of AIM that can be administered to collect the mental workload measurements. Each version varies in length and has different purposes and allows the assessment and diagnosis of different aspects of MWL. The decision to use one of the versions of AIM to assess the impact on MWL depends largely on the resources available for measurement and analysis. More importantly, the decision also depends on the objective of the MWL assessment.

Table 2 is a decision table that can be used as a guide for deciding which of the between different versions of AIM to use for different trial conditions. AIM-Hi contains nine items on which subjects will rate mental workload using one of the decision tree rating scale (chosen by the analyst or experimenter). AIM-Hi only allows a high level and global assessment of mental workload. AIM-Q contains 46 task items on which subjects will rate MWL. However, it allows the most detailed assessment of MWL. AIM-Cog consists of nine sub-tests, each of which can be administered individually. Each sub-test assesses the mental workload of a cognitive function sub-scale (e.g. planning, build and maintain situation awareness, memory management, etc.). The number of items in each AIM-Cog sub-test range between four and eleven items, on which subjects will rate MWL.

Table 2: Guideline to decide between the different versions of AIM

<table>
<tr><th colspan="2" rowspan="3"></th><th colspan="5">Objective of the Mental Workload (MWL) assessment</th></tr>
<tr><th rowspan="2">High-level overview of (overall MWL)</th><th colspan="4">Detailed understanding of the impact on MWL</th></tr>
<tr><th>Specific cognitive function group in mind</th><th>No specific cognitive function group in mind</th><th>Processing load on mental resource types</th><th>Impact on spare processing capacity</th></tr>
<tr><td rowspan="2">Limited resources (i.e. time available for measurement, analysis, staff, etc.)</td><td>Yes</td><td>AIM-Hi</td><td>AIM-Cog</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td>No</td><td>-</td><td>AIM-Cog</td><td>AIM-Q</td><td>AIM-Q</td><td>AIM-Q</td></tr>
</table>

3.3 What the Tool Consists of

This sub-section summarises the AIM Tool Set, that is the components constituting the AIM Tool Set. The AIM Tool Set can be contained either in a box folder or paper file folder, or in electronic folders on a 3.5’ (1.3 Mb) floppy diskette. AIM Tool Set consists of five sections. These and the documents or tool components which can be found in each section are listed below:

1. '<u>Description of the AIM Tool Set</u>': This section contains a document which describes AIM, its purpose and scope, and summarises the components available, as well as the sections in the tool set.

2. '<u>Guidelines for using AIM</u>': This section contains a document describing the guidelines for using AIM, i.e.:
   - which decision tree rating scale to use: absolute or relative,
   - which version of AIM to use: AIM-Q, AIM-Hi or one of the AIM-Cog,
   - preparation and administering AIM,
   - scoring and interpreting AIM (incl. the electronic 'AIM Scoring Tool').

3. '<u>AIM decision tree rating scales</u>': This section contains one laminated A5 card of each of the decision tree rating scales, and an A4 sheet for each decision tree rating scale with A5 cut-outs (suitable for laminating) of the decision tree rating scale. These A4 sheets are also available in electronic format for colour printing:
   - AIM absolute workload decision tree,
   - AIM relative workload decision tree.

4. '<u>AIM recording forms</u>': This section contains the recording forms for all three versions of AIM. For each version of AIM there is a form for both absolute and relative workload judgements:
   - AIM-Q recording forms,
   - AIM-Cog recording forms,
   - AIM-Hi recording forms.

5. '<u>AIM instruction sheets (for subjects)</u>': This section contains the instruction sheets which should be distributed to subjects with the recording forms for all three versions of AIM. For each version of AIM there is an instruction sheet for both absolute and relative workload judgements:
   - AIM-Q instruction sheets,
   - AIM-Cog instruction sheets,
   - AIM-Hi instruction sheets.

4. RATIONALE BEHIND THE DEVELOPMENT PROCESS OF THE AIM TOOL AND ITS CONTENT

4.1 Approach for the Development

The use of models[2] of the human user has become an important research and design tool in the study of human factors (HF) in ATM. The construction a model in HF for any purpose is usually achieved using a modelling technique. Modelling techniques may be characterised as having five elements (Timmer & Long, 1996). These five elements can be used to evaluate different techniques, compare different techniques against one another and also to develop new techniques. This five-element approach was adapted and modified for use as a framework to guide the development of a Mental Workload (MWL) evaluation technique for automation impact. It is proposed that in order to develop a tool for assessing the impact on MWL, the development process needs to take into account the five elements:

1. A purpose and scope: To describe, explain or predict MWL. In particular, the development processes need to establish which aspect or aspects of MWL.

2. A theoretical model: A theoretical representation and explanation of some aspect of MWL. This will not only provide a structural description of MWL but also an explanation of the processes and/or factors that determine the aspects of MWL decided above (in 1).

3. A methodology and representational format: The tool must have a representational format. The methodology consists of how the tool will measure and assess MWL. The representational format consists of: (i) its actual physical representation, i.e. pen/paper-based or executable as a computer program, (ii) its content (task-based items, questionnaire-based, decision tree, subjective rating scales, etc.) and (iii) its structure, i.e. single scale or multiple sub-scales.

4. A data collection process: Procedures for collecting the data to assess the workload, including the material and resources required to collect the data.

5. A scoring system and interpretation: Procedures to score the data, obtain the estimates of workload for statistical analyses and interpret the resulting estimates of the workload.

Each element represents a phase in the development process and determines what the goals and activities of that development phase are and what the output or decisions are as a consequence of the activities undertaken. Each of the following sub-sections (4.2 to 4.4) will begin by describing what goals and activities undertaken in the phase of the development process as well as

[2] A model is a representation of some respect of the human user that is constructed for a purpose.

describe what the output or decisions are as a result. Section 5 will describe the procedures set up and materials developed for the data collection process and scoring and interpretation system of AIM Tool.

4.2 Purpose and Scope

The goal in this phase was to decide on the purpose and specify the scope of AIM. The goal was not only to decide on the appropriate concept of MWL but to specify 'what' aspects of workload to measure and 'where' (and thus 'how') to measure the MWL.

The main activity consisted of two literature reviews. One was a literature review of the impact of automation on workload to decide what the scope of such a tool should be. The literature search was not confined to the aviation industry. A summary of a literature review conducted on the impact of automation on MWL can be found in Appendix A. In summary, the implication from the literature review on the impact of automation on workload was that it is essential that:

1. The concept of MWL is described well so that a decision can be made on what mental workload constructs (i.e. what aspects of mental workload) will be measured.

2. AIM should provide measurements on different dimensions of MWL, rather than provide a single global measurement of MWL. These dimensions can be along the different mental resources (e.g. spatial, verbal) required for task performance in an automated system.

3. AIM is able to provide a profile of MWL according to the types of cognitive functions in ATM.

The other was on the various mental workload concepts in the current literature to define the key concepts and core constructs for the proposed tool. The reason for this is that mental workload is a multidimensional and multi-faceted concept. A review of literature on the concept of MWL was necessary in order to examine and understand the various facets of MWL. Only then can the decision be made objectively on what aspects of MWL should AIM measure.

In summary, concepts of mental workload can be described as:

- The *mental effort* required to perform the tasks and the amount of *processing load* on the various mental resources required by task performance.

- A consequence of either *task difficulty* or the availability of mental resources (that is increased or decreased spare *processing capacity*).

- The manifestations of mental workload (that is how MWL can be observed). These are *performance decrements* or *task inefficiencies*.

Figure 1 illustrates the aspects of MWL and where each aspect is assumed to occur in relation to the human-automation work system, i.e. in the mental processing activities or physiological responses, in task performance or in the Human-Machine Interactions (HMI) with the automated system.

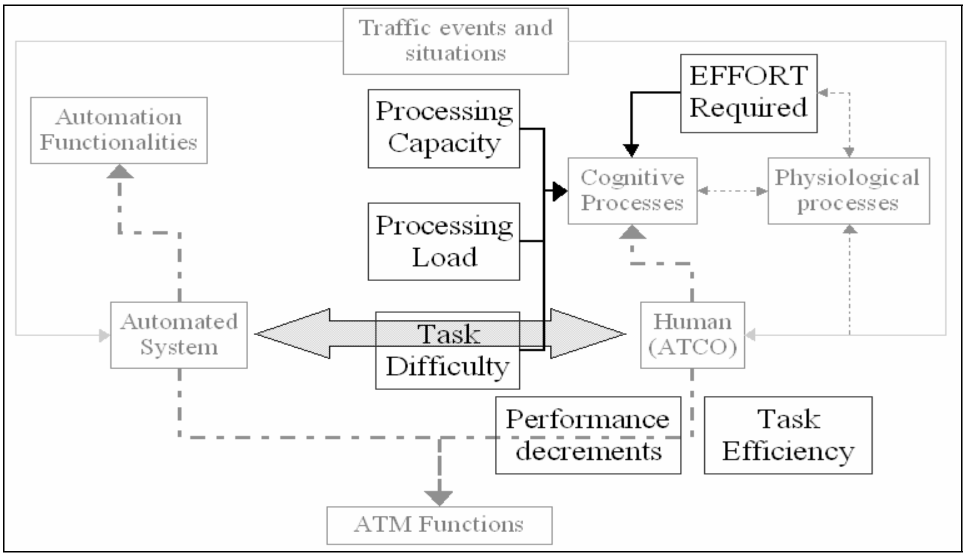


Figure 1: What aspects of mental workload can be measured and where

As a result of the literature reviews on research in automation impact on MWL and concepts of MWL, and a consultation process[3], the purpose and scope of AIM were decided as follows:

- to describe impact on mental workload and provide some diagnostic explanation of changes in:
  - MWL levels,
  - type of MWL,
  - distribution of MWL:
    - among the cognitive functions,
    - among the mental resources;
- focus on the impact on critical cognitive activities and likelihood of overload situations;
- 'what' aspects of mental workload will be assessed: mental effort, task difficulty, processing load (on mental resources) and spare processing capacity;
- 'where': mental workload due to (i) cognitive functions, (ii) interaction with automated system (iii) traffic and any other significant factors.

[3] Several meetings with HF specialists from EUROCONTROL Headquarters and Experimental Centre (EEC) were held as part of the consultation process.

4.3 Theoretical Model

The goal was to decide on the theoretical model of AIM. The main activity here consisted of a literature review of the various theories of workload. A theory of workload was selected based on the decisions and outcomes from Phase 1. The theory will support the construction of the technique and provide a basis for explaining the workload impact (i.e. ‘why’).

Several theoretical models were considered. Based on the purpose and scope of the tool, as specified above, each theoretical model was reviewed against:

- Its suitability: Does its concept of mental workload account for the aspects of mental workload that the AIM Tool seeks to measure?
- Its utility to the purpose and scope: Does it explain mental workload due to mental resource types and/or spare processing capacity?
- Its relevance to ATC: Does it account for mental workload due to multi-task performance and how; is it able to explain differential mental workload due to multi-task performance such as time-sharing and parallel task performance?

The review concluded that the Multiple Resource Theory (MRT) model of MWL (Wickens, 1980; Wickens, 1991) was suitable as the theoretical model for AIM as is explained below. A brief description of the MRT can also be found in Appendix B.

The MRT was chosen because of:

- Its suitability with respect to how it describes MWL. Its concept of MWL account for the aspects of MWL that the AIM Tool seeks to measure.
- Its utility to the purpose and scope of AIM. The MRT provides a structural definition for different mental resources. It makes a distinction between MWL due to these different mental resources. It provides an explanation on how the variation in demands on different mental resource types can affect MWL. In addition, it explains how spare processing capacity may be affected by MWL demands from different mental resource types.
- Its relevance to ATC. The cusp of competent and efficient ATC involves multi-task performance. MRT is able to describe and explain MWL due to multi-task performance. It is able to explain why changes in task difficulty and/or task structure (i.e. how the task is carried out) in one task may affect MWL in multi-task performance. More importantly, it is able to explain the phenomenon of *time-sharing,* in which two tasks, both of non-trivial difficulty, can be performed concurrently with no performance decrement, even though each can be shown to interfere with other activities.

4.4 Methodology and Representational Format

4.4.1 Introduction

The main activity here consisted of selecting a suitable format for AIM as well as the construction of the content and structure for the tool. A review of existing tools was carried out first, followed by a series of activities concerning the actual construction of the tool.

A review of existing MWL assessment techniques and tools was conducted after the purpose and the scope was defined and the MRT was selected as the preferred theoretical model of MWL. The purpose of the review of existing tools was to decide between:

(i) constructing an entirely new mental workload assessment tool for the purpose and scope defined in Section 4.2;

(ii) adapting and modifying an existing tool for the purpose and scope defined;

(iii) constructing a new technique by combining two or more existing tools for the purpose and scope defined.

A summary of the review on mental workload assessment techniques can be found in Appendix C. The outcome of the review of existing tools can be found in Appendix D.

Based on the outcome of the review of existing mental workload assessment tools, the decision was made to construct AIM as an entirely new assessment tool rather than develop AIM as a modification or combination of existing tools.

Most of the tools were unsuitable as they did not measure MWL with respect to the type of cognitive function workload nor were they able to diagnose MWL according to mental resource types, with the exception of PUMA. However, some of these tools reviewed had positive aspects that could be used in the development of AIM. These are described below:

1. Although PUMA, a mental workload modelling tool, was unsuitable as it was difficult to implement and is resource intensive, its workload algorithm based on the MRT and related resource channels classification and conflict matrix were useful materials for the construction of the new tool.

2. The structure of NASA-TLX was useful, that is a sub-scale and score for each aspect of workload. However, unlike NASA-TLX, the new tool will focus on partitioning mental workload further into cognitive function groups (for e.g. MWL due to planning tasks, decision-making tasks or multitasking) and mental resources (for e.g. MWL due to visual resource demands or spatial resource demands).

3. SWAT (and Zachary/Zachlad Cognitive Analysis) uses specific tasks as the item content in the tool, on which subjects rate the mental workload demands. Using specific tasks as the items in the tool makes the items to be rated easy and familiar to the subjects (i.e. controllers). Such items also allow the subjects to focus their mental workload judgements on specific task performance and situations. This may be easier rather than gross global judgements over a work period. It is easier to make subjective mental workload judgements of X, Y and Z than make a global judgement of mental workload without the consideration of specific task performance, especially if the mental workload experience for each of the X, Y and Z task is different.

4. The decision tree rating scale format in Cooper-Harpers, Modified Cooper-Harpers and Bedford workload tools was a useful and easy way to guide respondents' in their subjective ratings of mental workload.

4.4.2 Methodology

In order to satisfy the second requirement specification (in Section 2.3) which is to have a tool that is easy and convenient to apply during real-time simulations of ATM and in a typical Human Factors Laboratory, the decision was made to adopt a subjective assessment technique for AIM. This will allow to tool to be simple to use, easily portable and non-site specific, easy to implement with minimum equipment requirements and specialised expertise and resources. In addition, the subjective assessment technique is easy to implement in a real-time ATM simulation as well as in prototyping simulations of future ATM automated systems.

Although the decision was made to assess four different aspects of mental workload (in Section 4.2), it was decided that subjective ratings would only be made for mental effort required and task difficulty. The reasons were:

1. To keep the tool simple for use in simulations. The tool will be designed such that evaluations of processing load on mental resources and spare processing capacity will use the ratings on mental effort required and task difficulty by implementing the theoretical assumptions of MRT (which was the chosen theoretical model for the tool - see Section 4.3). This will be discussed further in later sections on item content of the tool and interpretation of the scores from the tool.

2. The difficulty of introspection with respect to what mental resources were used. It is unfair to expect controllers completing the tool to understand and clearly identify what the mental resources are and to rate the demand on these mental resources.

3. The likelihood that the tool will be designed to be administered at the end of a simulation run/trial, so that there will be no intrusion into task performance during the run/trial. This will make it difficult for controllers to meaningfully rate processing capacity for the entire duration run/trial.

4. One of the implications from the review of existing tools suggested that it was reasonably easy and meaningful for subjects to make subjective ratings on the amount of mental effort invested and on how difficult task performance was. These are common rating methods used in many of the existing tools.

Following from the review of existing tools, it was decided that a decision tree rating scale format would be used to guide the subjects' ratings. Hence, the rating scales for mental effort and task difficulty will be embedded within a decision tree.

Two versions of a decision tree rating scale were produced. Each decision tree contained questions and two rating scales, one for mental effort and one for task difficulty. The first version was an absolute workload decision tree allowing subjects to make ratings of their experience of MWL. The second version was a relative workload decision tree allowing subjects to make comparative ratings of their MWL experience. This means that subjects can rate how much required the mental effort was in the simulation/trial compared to the current operational situation. A short description of the construction of the rating scales and decision trees (both versions) can be found in Appendix E.

4.4.3 Representational format

As the idea was to keep the tool easy-to-use and requiring no specialised resources to administer it, it was decided that it would consist of a pen-and-paper test. It has to be noted, however, that this choice does not rule out electronic or computerised forms of the tool in future.

The item content of AIM consists of specific ATC-related cognitive tasks. MWL ratings (on mental effort and task difficulty) are made on these tasks. Three taxonomies were constructed for the list of cognitive tasks. The list of cognitive tasks consisted of 46 items. Appendix F describes how the output from the SHAPE work package on skill set requirements (see EATM, 2004c) was used to construct the list of 46 task items for AIM. The development of the three taxonomies is also described in the same appendix. The 46 task items can be found in Appendix G.

The first taxonomy categorises the task items into nine cognitive function groups. These are:

1. Multitasking
2. Direct attention to information sources
3. Take account of and process external information
4. Memory management
5. Build and maintain SA
6. Planning
7. Decision-making
8. Diagnosing and problem solving
9. Team awareness

These nine cognitive function groups in the taxonomy and the associated task items make up the nine sub-scales of AIM, allowing the assessment of MWL due to specific cognitive functioning. The key for matching the task item to the cognitive function groups can be found in Appendix H.

The second taxonomy for the AIM task items categorises the task items according to the types of mental resource required. The mental resource types are:

1. Encoding
2. Central Processing
3. Response
4. Verbal
5. Spatial
6. Visual
7. Auditory
8. Motor

The taxonomy of mental resources allows the diagnosis of mental workload according to the demand on different mental resource types. The key for matching the task item to the mental resource types can be found in Appendix I.

The third taxonomy for the 46 AIM task items categorises the items according to the degree of resource competition (or resource conflict). The categories are High, Medium and Low. The key for matching the task item to the mental resource types can be found in Appendix J. The degree of resource competition taxonomy allows the assessment of MWL to be diagnosed according the impact on spare processing capacity. The assumption is that the greater the MWL demand, the greater the impact on spare processing capacity. However, spare processing capacity is further affected and reduced if there is a high degree of resource competition between the processing load placed on the mental resources. Hence, the need to consider the degree of resource competition in the tasks to infer the impact on spare processing capacity.

5. PROCEDURES FOR USING THE AIM TOOL

5.1 Data Collection Process

AIM is designed to be administered at the end of a work period. As it was decided that AIM was to be a paper-based tool, there needed to be a data collection form to administer the tool and instruction sheets for both the subject and the user or experimenter.

5.1.1 AIM recording forms

The following three versions of AIM recording were produced to provide users a choice about how detailed the MWL assessment will be and hence a choice of lengths of recording forms and data collection time. For each version there are two types of recording forms, one for the absolute workload decision tree and another for the relative workload decision tree:

1. AIM-Q: Includes all 46 task items to be rated by the subject using the decision tree rating scale (either the absolute or the relative version). These items can be categorised later, after the data collection process, into the nine cognitive function groups. They can also be categorised later into the mental resource types. AIM-Q recording forms for both the absolute and the relative workload decision trees can be found in Appendix K.

2. AIM-Cog: Consists of nine sub-tests, one for each of the nine AIM cognitive function sub-scales, that is one for each cognitive function group allowing the assessment of MWL due to a specific targeted cognitive function group. These can be administered individually according to which cognitive functions are to be measured. For example, if a small scale trial were carried out on a new prototype of system change focused on assisting the controllers in decision-making, the decision-making sub-test could be administered for ease of administration and limitation of the resources required. Each of the nine sub-tests of AIM-Cog recording forms for both the absolute and the relative workload decision trees can be found in Appendix L.

3. AIM-Hi: This version is an abbreviated test with only nine items, intended to obtain a high-level and global assessment of the MWL due to the different cognitive function groups. However, it does not allow any further diagnosis into MWL due to mental resource types or specific task performance. Each cognitive function group is used as an item to be rated by the subject on mental effort required and difficulty. A simple but detailed explanation of each item is provided to the subject. This version looks a bit like NASA-TLX which provides subjects with descriptions of different dimensions of workload (such as frustration level, temporal demand or physical demand) and requires from them to make a global rating on each dimension. The difference is that AIM-Hi requires global ratings on each

cognitive function types. AIM-Hi recording forms for both the absolute and relative workload decision trees can be found in Appendix M.

5.1.2 Instructions for subjects

Instruction sheets for subjects were produced. These should be distributed to the subjects before the start of the trial or simulation to allow them time to read the instructions and gain some degree of familiarity with the AIM Tool. The instruction sheet should also be distributed with every administration of AIM. Three versions of instruction sheets were produced; one for each version of AIM. Similarly, for each version of the instruction sheet there are two different copies; one for the administration of AIM using the absolute workload decision tree and another for the administration of AIM using the relative workload decision tree. The instruction sheet options can be found in Appendix N.

5.2 Scoring System and Interpretation

The development of AIM required the construction of a scoring system and interpretation process.

5.2.1 Scoring system

AIM-Q is designed to measure the impact of automation on mental workload in detail. That is the mental effort required for task performance and difficulty of task performance. In addition, AIM-Q is designed to use the ratings on mental effort and task difficulty to evaluate the impact on mental workload due to different cognitive function groups and mental resource types, as well as to evaluate the impact on spare processing capacity.

In order to obtain a profile of MWL due to the different cognitive function groups and mental resource types, MWL scores for each function group and resource type needs to be computed. Similarly, to predict the impact on spare processing capacity, MWL scores for high resource competition task performance needs to be computed.

The data collected by the AIM Tool is interval data so averages can be computed and are meaningful for parametric statistical analyses and interpretation. The scoring system of AIM-Q uses the average ratings of a group of items as MWL scores (one each for mental effort and difficulty).

For example, the mental effort and difficulty MWL scores for multitasking workload are the averages of the mental effort and difficulty ratings, respectively on item numbers 1, 6, 8, 32, 33, 36 and 43. (The AIM-Q items which belong to the multitasking group can be found in Appendix H.) The mental effort and difficulty MWL scores for workload on response resources are the averages of the mental effort and difficulty ratings, respectively on item numbers 5, 8, 17, 18, 28 and 42. (The items that are in the response mental resource type can be found in Appendix I.)

The mental workload scores for mental effort and task difficulty are computed separately. One reason is to preserve the diagnosticity of the MWL scores. Mental workload impact in terms of the difficulty of task performance and mental effort have different implications on design recommendations for ATM system. The other reason is insufficient evidence that these two aspects of MWL are mutually exclusive and the relationship is additive.

Hence, individual subjects' ratings should preferably be scored accordingly before the data are statistically analysed. That is for each AIM-Q administered form the mental effort and task difficulty MWL scores for (i) each cognitive function group, (ii) each mental resource type and (iii) each resource competition category must be computed prior to statistical data analyses.

An electronic scoring tool for AIM-Q was constructed for to automate the scoring system. Instructions on how to use the 'AIM-Q Scoring Tool' can be found in Appendix O. The AIM-Q scoring tool can used to score individual completed recording forms. The tool is based on MS Excel. Hence, descriptive statistics and simple statistical analyses can be carried out in the tool immediately after the scoring of all individual completed forms. Alternatively, the data (i.e. computed scores) can be copied onto another MS Excel workbook and then exported to a statistical analysis programme (such as Statistical Package for the Social Sciences (SPSS™), Statistical Analysis System (SAS®), SYSTAT®, etc.).

Both AIM-Hi and AIM-Cog are designed as shorter versions of AIM-Q. There is no need to score individual subjects' ratings on AIM-Hi before data analysis. Subjects' ratings for the items on AIM-Hi can be entered directly into any data analysis software programme[4] (e.g. MS Excel, SPSS, etc.). Similarly, for AIM-Cog, the mental effort and difficulty ratings for the items on each sub-scale can be analysed directly in the way chosen by the analyst or entered directly into a data analysis software programme.

5.2.2 Interpretation

The differences between MWL scores of the categories in each of the AIM taxonomies can be interpreted as differences in amount of effort required or degree of difficulty. For example, Subject X's score on mental effort required for 'Planning' performance in condition Z is 3 and his/her score on mental effort required for 'Planning' performance in condition Y is 6. The interpretation is that subject X reported more mental effort required for planning in condition Y than condition Z. However, the data and scores in AIM do not allow interpretation of the degree of the difference. That is, in the example cited, it is not possible to infer that mental effort for planning in condition Y is twice as much as condition Z.

[4] However, it is recommended that the MWL ratings be entered in MS Excel initially, organised in a structure suitable for the specialised statistical analysis programmes and then exported into the statistical analysis programmes. For example, for within-subject comparisons in SPSS data for levels of within subject factors must be entered down under a column for each level.

Inferences about the impact on spare capacity can be made using the differences between mental workload scores on high resource competition task performance. The higher the MWL score on high resource competition task performance, the lower spare processing capacity available is. For example, Subject X's score on mental effort required for high resource competition tasks in condition Z is 3 and his/her score on mental effort required for high resource competition tasks in condition Y is 6. Then the inference is that subject X's spare processing capacity in condition Y is less than that in condition Z. However, again the data and scores in AIM do not allow interpretation of the degree of the difference. That is, in the example cited, it is not possible to infer that spare processing capacity in condition Z is twice as much as condition Y.

6. CONSTRUCT VALIDITY, INTERNAL CONSISTENCY RELIABILITY AND GENERAL USABILITY OF AIM

6.1 Introduction

This section reports on the preliminary findings and conclusions on the construct validity, internal consistency reliability and general usability of AIM, with respect to (i) using the decision tree rating scales, (ii) understanding the task items, (iii) rating the MWL on each task item, (iv) average time taken to complete AIM-Q and AIM-Hi, (v) the average time taken to score AIM-Q and AIM-Hi and (vi) diagnosticity of AIM.

Construct validity of a measurement tool can be defined as the degree to which inferences can legitimately be made from the measurements about the theoretical constructs that the tool measures (Trochim, 2000). That is how well does AIM measurements reflect MWL in ATC, does AIM measure MWL and it AIM sensitive to the differences in MWL and is AIM able to diagnose the impact on MWL. There are six types of construct validity: (i) face validity, (ii) content validity, (iii) concurrent validity, (iv) convergent validity, (v) discriminant validity and (vi) predictive validity. To demonstrate that AIM has overall construct validity it is essential to show that each type of construct validity has been ensure and satisfied. Sections 6.2 to 6.6 will report on the conclusion on each type of construct validity for AIM. Convergent and discriminant validities are presented and discussed in 6.5.

A pilot study was designed purposefully to investigate the preliminary construct validity and internal consistency reliability of AIM, as well as examine the general usability of AIM-Q and AIM-Hi. The types of construct validity investigated in the pilot study were concurrent, convergent, discriminant and predictive validity. The paradigm used to establish if there is evidence of validity is the degree to which observed findings or differences in the study matches the reasonable hypotheses that what can be found about mental workload or what differences can be expected between the conditions manipulated in the study. For example, if there is reasonable basis to expect that a true difference in mental workload exists between condition X and Y, then AIM should show that difference. However, if there is no reasonable basis to expect a difference in MWL between condition X and Y (i.e. the possibility that there is no real difference between the two conditions), then it is unfair to expect AIM to demonstrate a difference.

A short explanation of each of these four types of construct validity and how to demonstrate each of them is provided in Appendix P. The pilot validation study and the results of the study are also reported in Appendix P. Face validity and content validity are usually examined subjectively and analytically and will be discussed in Sections 6.2 and 6.3.

The internal consistency reliability of AIM is reported in 6.7 and the general usability of AIM is discussed in 6.8.

Section 7 will discuss further implications of the findings and conclusions, and will provide some recommendations on the future work on AIM.

6.2 Face Validity

Face validity refers to extent to which the users of the tool and other Subject Matter Experts (SMEs) perceive the tool to be measuring what it purports to measure. That is controllers and other SMEs perceive or feel that the items and rating scale is adequate for rating their subjective experience of mental workload and that the tool reflects MWL accurately. Although this the weakest way to try to demonstrate validity of a measurement tool, it can be useful for operational acceptance.

During the pilot study and the trial use of AIM-Hi in a prototyping simulation, controllers reported that the list of task items were sensible and comprehensible. They also found the decision tree and rating scales easy to use. The rating scales were found to be intuitive once the controllers were familiar with them. Controllers felt that they were able to use the decision tree and ratings scale to make judgements on their experience of mental workload.

6.3 Content Validity

Content validity refers to the extent to which the content of the measurement tool reflects the theoretical constructs it purports to measure and the extent to which the methodology of the tool is appropriate for the measurements. Developing the tool in a systematic way can ensure content validity and by making sure that the development of the tool is theoretically sound.

An approach was constructed to ensure that the development of AIM is systematic. The development of AIM also depended on the conclusions from various literature reviews and the implications of the findings from these literatures. The construction of the content of the tool was also based on these reviews as well as the MRT, which was selected as the theoretical framework for AIM. Frequent and regular reviews by domain HF experts were also held to ensure that the approach taken to develop and construct AIM was valid and sensible.

6.4 Concurrent Validity

Concurrent validity refers to the ability of the measurement tool to distinguish between two groups, which are theoretically different in the construct to be measured. That is AIM is able to discriminate between different MWL levels. To demonstrate this ability and ensure that AIM can distinguish between different workload levels, AIM should at least be able to discriminate between situations where there is clearly a difference in MWL levels. For example, between different traffic levels, different traffic and sector complexities and between different operational roles. The sensitivity of AIM can be demonstrated if AIM is able to discriminate between situations which are similar but may be different in terms of levels controller MWL. For example, amount of automated assistance, changes in operational procedure, increase in staff support.

Traffic levels and sectors were manipulated to create trial conditions where MWL levels are clearly different. In addition, the percentage of aircraft in the traffic sample which were datalink equipped created trial conditions which were highly similar, i.e. the only difference being the amount of automated assistance the controller was able to use. That is the impact of the automated assistance on controller MWL. Thus, in order to show concurrent validity, AIM should be able to sufficiently discriminate between the traffic levels, sectors and amount of automated assistance with respect to the MWL levels. Evidence of concurrent validity depends on the extent to which observed differences in AIM mental workload scores between traffic levels, sectors and amount of automated assistance match the expected differences in MWL for these factors.

There were 25 significant differences ($p<0.05$) found between traffic levels in AIM MWL scores. That is 70% of the expected differences between traffic levels were observed in the pilot study (i.e. observed differences which were statistically significant). This suggests that there is evidence that AIM was able to discriminate between high and low traffic levels, with respect to MWL levels.

All of the expected differences between East and West sectors were observed in the pilot study (i.e. observed differences which were statistically significant; $p<0.05$). This suggests that there is evidence that AIM was able to discriminate between East and West sectors, with respect to MWL levels.

There were nine significant differences ($p<0.05$) found between varying amounts of automated assistance and AIM MWL scores. That is 56% of the expected differences between varying amounts of automated assistance were observed in the pilot study (i.e. observed differences which were statistically significant). This suggests that there is evidence that AIM was able to discriminate between varying amounts of automated assistance.

Discussion on concurrent validity

Table 3 displays a summary of the percentage of observed differences found in the study, which match the expected differences. The table also displays in parenthesis the number of observed differences found in the study, which matched the expected differences as a fraction of the total number of expected differences.

Table 3: Summary of the percentage of observed differences found in the study, which match the expected differences

Traffic levels	Sector	Degree of automated assistance
70% (25/36)	100% (4/4)	56% (9/16)

The percentage of expected differences observed in the study was not as high as those for traffic levels and sector. There are a few issues that may have contributed to this. The first issue is the different impact that the automated assistance has on the different sector. The MWL scores reported by the controllers suggest that on East sector the MWL was larger for higher amount of automated assistance than that for lower amount of automated assistance. On the other hand, MWL scores reported by the controllers suggest that on West sector the MWL was smaller for higher amount of automated assistance than that for lower amount of automated assistance. This could result in the lack of significant differences between levels of automated assistance in the overall MWL comparison across sectors.

The next issue is that the impact on MWL between 50% and 95% of datalink-equipped aircraft is marginal. It is possible that the utility of the automated tools evens out at a certain percentage of datalink-equipped aircraft. It this is true, then the difference in the impact is marginal. Although significant differences were found in West sectors for both difficulty and mental effort MWL scores, only difficulty MWL score was significantly different in East sector. On the physiological measurements, a significant difference in pupil diameter[5] was found in East sector but no significant difference in MWL was found in Heart-Rate Variability (HRV) measurements. It may be prudent to examine this potential effect and investigate further the differences in mental workload between varying amounts of automated assistance from datalink technology by replicating this study using different percentages of datalink-equipped aircraft (e.g. 5%, 25%, 50%, 75%. and 95%).

Lastly, the controllers who volunteered to participate in the study did not have prior experience in using datalink technology. They were given a briefing and explanation on datalink technology and two training trials. The controllers expressed that this was insufficient and they had difficulties using the datalink technology because they were unfamiliar with the graphical user interface of the tools available in the system and what functions and assistance were available for them to use.

These issues needs to be considered and examined in further validation research on AIM and will be discussed further in Section 7.

Conclusion: The observed results and findings suggest that AIM has fairly good concurrent validity.

6.5 Convergent and Discriminant Validities

Convergent validity refers to the extent to which measures of the same constructs, which theoretically should be related to each other, are in fact observed to be related to each other (Trochim, 2000). The measures of mental workload from AIM should correlate highly or show convergence with

[5] Pupil diameter was measured only on East sector were

measures of MWL from other assessment tools that purport to measure mental workload.

Discriminant validity refers to the extent to which measures of other constructs are observed not to be related to each other (Trochim, 2000). The measures of mental workload from AIM should **not** correlate highly or show divergence from assessment tools that do not measure MWL.

In the pilot study conducted to investigate the preliminary validity of ATM only measurements of physiological activity were taken. No other subjective assessment tool was used because of the time constraints between measured trials and on the duration of the study. The three physiological measurements were:

(i) Heart Rate (HR) or Inter-Beat Interval (IBI): Heart rate or IBI has been used in other workload research studies. Several research studies have found that manipulation in MWL may not affect HR/IBI while manipulations in physical workload are likely to affect HR/IBI. That is HR/IBI as workload indicators are sensitive to physical workload changes rather than mental workload changes. However, HR/IBI are very sensitive also to other physiological reactions, bodily functions and motor responses (e.g. physical limb movements and vocal responses).

(ii) Heart-Rate Variability (HRV): HRV has been increasingly used in other workload research studies. Several research studies have found that manipulations in mental workload are more likely to affect HRV than HR/IBI. That is HRV is more sensitive than HR/IBI as an indicator of mental workload changes. Nevertheless, similar to HR/IBI, HRV measurements have been shown to be very sensitive to other physiological reactions and may be affected by these physiological responses bodily functions and movements.

(iii) Pupil diameter: Pupil diameter has been used in other research studies as an indicator of mental workload. Research findings on the sensitivity and utility of pupil diameter as a measurement of mental workload has been mixed, as pupillary responses vary according to other physiological reactions and environmental changes, apart from mental workload.

In the pilot validation study pupil diameter and HRV were used as convergent measures and IBI was used as a discriminant measure. Thus, in order to show convergent validity, AIM should correlate highly (with statistical significance) with pupil diameter measurements and HRV measurements and to show discriminant validity AIM should **not** correlate significantly with IBI (i.e. no statistically significant correlations). Evidence of convergent and discriminant validity depends on the extent to which observed correlation coefficients between AIM mental workload scores and pupil diameter, HRV and IBI match the expected correlations.

Table 4 provides a summary of the percentage of observed significant correlation found in the study, which matches the expected correlation. In brackets is the number of observed differences found in the study, which matched the expected differences as a fraction of the total number of expected differences.

Table 4: Summary of the percentage of observed significant correlation found in the study, which match the expected correlation

Convergence with Pupil diameter	Convergence with HRV	Divergence from IBI
78% (14/18)	33% (6/18)	83% (15/18)

In summary, fourteen significant correlation coefficients between AIM mental workload scores and pupil diameter were found, meaning that 78% of the expected correlations were observed in the pilot study (i.e. observed correlation coefficients which were statistically significant). On the other hand, only six significant correlation coefficients between AIM mental workload scores and HRV measurements were found, meaning that 33% of the expected correlations were observed in the pilot study. This suggests that there is evidence that AIM has good convergence with pupil diameter but poor convergence with HRV measurements.

In addition, only three significant correlation coefficients between AIM mental workload scores and IBI measurements were found, meaning that 83% of the expected correlations were observed in the pilot study (i.e. observed correlation coefficients that were **not** statistically significant). This suggests that there is evidence that AIM has good discriminance from IBI measurements.

6.6 Predictive Validity

Predictive validity refers to the ability of the measurement to predict something it should theoretically be able to predict. AIM was designed to be able to indicate whether the system change has an impact on the spare processing capacity of the controller. This means that, if the system change imposes an increase in mental workload on task performance which involves high resource competition/conflict, the spare processing capacity of the controller will potentially be reduced by the system change. To demonstrate this ability AIM should at least be able to predict in situations where there is clearly an impact on spare processing capacity.

In summary, five significant predictions on spare processing capacity were found. That is 83% of the expected predictions were observed in the pilot study (i.e. observed predictions which were statistically significant). This suggests that there is evidence that AIM has fairly good predictive validity.

6.7 Internal Consistency Reliability

Significant internal consistency reliability estimates ($p<0.01$) were found for AIM-Hi, AIM-Q and all of the AIM-Cog sub-tests. The results suggest that the items in AIM-Hi are consistent in their measurements. Similarly, the results suggest that the items in AIM-Q are consistent in their measurements. Finally, the results for the AIM-Cog sub-tests suggest evidence that the items in each sub-test are consistent in their measurement. The findings indicate good internal consistency reliability in AIM.

6.8 General Usability Issues

Controllers were given the instruction sheet on how to use the decision tree rating scale. They reported that instruction sheet explains adequately how to use the decision tree rating scale. They found the rating scales easy to use and very quickly became familiar with the decision tree. They reported that they did not have any difficulties understanding the ratings and their definitions. They appreciated the wide range of choice for rating their mental workload.

The task items in AIM-Q were very familiar tasks to the controllers and very recognisable. This made the rating of the mental workload on each task item easier than expected. Most of the task items and the verbs used in the task items were sufficiently easy to understand and translate, except for the words ‘Recognise’ and ‘Identify’. Controllers felt that these two words mean the same and hence, the task items are perceived to be similar.

The definitions provided for the AIM-Hi items were understandable and the controllers did not report any difficulties making global mental workload ratings on the cognitive function groups.

The controllers were initially given thirty minutes to complete AIM-Q. The first two occasions when AIM-Q was administered the controllers took between 25 and thirty minutes to complete AIM-Q. However, as they became more familiar with the AIM-Q items and the decision tree rating scales, the controllers took less time to complete AIM-Q. By the second half of the study (fifth measured trial onwards) the average time taken to complete AIM-Q was fifteen-twenty minutes. Similarly with AIM-Hi, controllers progressively took less time to complete as the study progressed. By the second half of the study the average time taken to complete AIM-Q was approximately five minutes.

After the data collection process, data from AIM-Q recording forms had to be scored before the statistical analyses. The average time taken to score each AIM-Q form using the ‘AIM-Q Scoring Tool’ was approximately five minutes. However, the time taken depended largely on the analyst skill in data entry in MS Excel and his/her familiarity with MS Excel.

In the pilot study AIM-Hi was used as a global assessment of mental workload even though each item on AIM-Hi was a cognitive function group, because there were not enough subject numbers to analyse mental workload from AIM-

Hi data according to the different cognitive function groups. However, if the sample size was sufficiently large, mental workload could be diagnosed according to the different cognitive function workload.

On the other hand, AIM-Q was used to diagnose mental workload assessment further according to its cognitive function workload profile as well as the profile of the demands on the different mental resource types. The diagnostic ability of AIM allowed the assessment of the distribution of mental workload impact amongst cognitive functions and mental resource types. Even though pupil diameter measurements showed similar significant differences in mental workload, the assessments were global and no further diagnosis about mental workload could be easily made from the pupil diameter measurements. That is without conducting an incident/task analysis of each measured trial and analysing the video recordings and the pupil diameter measurements to match incidents/tasks in each trial to the pupillary recordings.

In addition, the assessment of mental workload impact on high resource competition task performance allowed predictions on the potential impact on spare processing capacity.

7. KEY ACHIEVEMENTS OF THE WORK PACKAGE

This section summarises all the key achievements of the work package. This includes all the key products of the work package and the key findings of and conclusions from the pilot validation study. The section ends with a sub-section on further implications of the work package, and future work.

7.1 Key Products

The main product resulting from the work package is a tool for evaluating the impact on Mental Workload (MWL) from automated ATM systems. The tool is called **A**ssessment of the **I**mpact on **M**ental workload or AIM. The tool is able to provide assessments of MWL according to different dimensions for detailed diagnosis of mental workload, that is:

1. The ability to evaluate different dimensions of mental workload. AIM allows the analyst to determine the mental workload due to:
   - different cognitive functions,
   - the demands on different mental resource types.

2. AIM also partitions the mental workload assessment according to mental effort required and task difficulty. Mental workload impact due to differences in mental effort required may have different design implications from mental workload impact due to task difficulty.

3. AIM has measurement sub-scales for each dimension of mental workload. This will enable a profile of MWL impact to be produced. That is a profile of MWL due to:
   - different cognitive function groups, e.g. multitasking workload, memory management workload, planning workload or decision-making workload;
   - the demands on different mental resource types, e.g. visual mental resources, spatial mental resource or verbal mental resources.

4. In addition, the evaluation of MWL can be further simplified or focused by using the appropriate individual cognitive function sub-scale. The diagnosticity according to the profile of mental workload will then enable better insight into the design of automation system with respect to dimensions of mental workload. Design recommendations and changes can be focused and targeted onto the part of the system responsible for the aspects of MWL affected.

5. AIM was designed to require minimal resources to administer. There is no special training required to administer AIM. The AIM Tool Set contains guidelines for users to decide which version of AIM to use and how to score and interpret the different versions of AIM. The tool set also include

an computer-based tool which will automate the scoring of AIM. This makes AIM an easy and convenient tool to apply during real-time simulations of ATM and in a typical Human Factors Laboratory.

6. AIM includes an indicator of the extent to which mental workload report is attributable to the automated system or traffic/situational conditions. This allows the analyst to distinguish MWL due to situational conditions and due to the design of automated tools.

7. AIM includes a capability to indicate if the ATM system change may have a potential impact on the spare mental processing capacity of the controller and thereby increasing the risk of overload.

AIM is a subjective MWL assessment tool. It requires the subject to rate the mental effort needed for task performance and the difficulty of task performance. Both are rated on a seven-point rating scale. Two types of rating scales can be chosen according to experimental design: absolute and relative rating scales. Both types allow either absolute or relative judgement of mental workload.

The rating scales are embedded in decision trees containing questions that guide the subjects' use of the rating scales to make their mental workload ratings. The MWL ratings are made on specific task items or defined cognitive function groups (such as multitasking, planning or decision-making tasks).

The decision to use either the absolute or relative workload decision tree depends on the design of the trial or simulation in which AIM will be used and the simulation/trial equipment resources available (i.e. time, staff, simulation resources, etc.). Guidelines on deciding which decision tree rating scale to use are provided in the AIM Tool Set.

Three versions of AIM can be administered to collect the MWL measurements. This allows flexibility to the users in their assessment of the impact on MWL. Each version varies in length, and has different purposes and allows the assessment and diagnosis of different aspects of mental workload. The decision to use one of the versions of AIM to assess the impact on mental workload depends largely on the resources available for measurement and data analysis. More importantly, the decision also depends on the objective of the MWL assessment. Guidelines on deciding which version to use are provided in the AIM Tool Set.

Instructions sheet for subjects and guidelines on the preparation and administration of AIM are provided in the AIM Tool Set.

Guidelines on how to score AIM recording forms (all versions) and how to interpret AIM mental workload scores are provided in the AIM Tool Set. In addition, a computer-based tool was produced for scoring AIM-Q (the longest version of AIM) and is also provided in the AIM Tool Set.

The AIM Tool Set is available electronically or as a hard copy.

7.2 Key Findings and Conclusions

There is subjective evidence of face validity in AIM based on controllers' subjective reports after use of the tool. In addition, constructing an approach to ensure that the development of AIM has a theoretical basis and is systematic ensured content validity. The development of AIM also depended on the conclusions from various literature reviews and the implications of the findings from these literatures. The construction of the content of the tool was also based on these reviews, the MRT, accepted mental workload algorithms from PUMA and output from SHAPE work package on skill set requirements (see EATM, 2004c). Frequent and regular reviews by domain human factors experts were also held to ensure that the approach taken to develop and construct AIM was valid and sensible.

Overall, the preliminary validation of AIM provided sufficient evidence to conclude that AIM has fairly good construct validity. The pilot validation study found that AIM could discriminate between differences in overall mental workload levels in various conditions; different traffic levels, different sectors and different amount of automated assistance.

Reasonable hypotheses about true differences in mental workload distributions can be expected only in the traffic level conditions. That is there are real and actual differences between traffic levels in the mental workload levels among cognitive functions and in mental resource types. Hence, the ability of AIM to discriminate between different mental workload levels amongst cognitive functions and mental resources was explored further as it was fair to expect that these differences will be observed in the study. The study found sufficient evidence that AIM was able to diagnose mental workload between traffic levels according to these dimensions.

In the pilot study it was not possible to establish reasonable expectations about true differences in mental workload distributions between sectors. That is it was not possible to assume that there are actual differences between sectors in the mental workload levels among cognitive functions and in mental resource types. Without prior task analysis of both sectors to establish what the differences are between task performances required of either sector, there was no reasonable basis to expect what the actual differences in cognitive function workload and mental resource demand between sectors may be.

Similarly, it was not possible to establish reasonable expectations about true differences in all the mental workload distributions between different amount of automated assistance. That is it was not possible to assume that there are actual differences between different amount of automated assistance in the mental workload levels among the cognitive functions (except for decision-making) and in all of the mental resource types. It was reasonable to expect a difference in decision-making as almost all of the processes required for decision-making were predicted to be affected by datalink technology by the SHAPE Automation Framework (see EATM, 2004c). A methodological issue faced by the study is the basic assumption that there is a difference between the automated assistance conditions. That is there is a difference between

50% and 95% of datalink-equipped aircraft present in the sample. It is possible that the datalink technology used in the study has a marginal utility effect (i.e. beyond a certain percentage of datalink-equipped aircraft present, there is no added benefit to be gained from datalink technology). It was not possible within the scope of the study to rule out this out or examine where the ceiling may be if such an effect existed.

The convergence validity with pupil diameter was good and there is sufficient evidence to conclude that the discriminant validity was also good. However, the convergence with cardiac activity was poor. Convergence of AIM mental workload scores with HRV was found only in multitasking, direct attention and memory management functions and response, verbal and auditory mental resources. This is consistent with literature which suggests that cardiac monitoring as a workload measurement technique is more sensitive to physical workload and changes in physiological and muscular responses. The cognitive functions such as multitasking and directing attention usually involves overt actions like head, eye and limb movement. In addition, mental resources for responding, hearing and speaking involve physiological reactions and changes.

7.3 Further Implications and Future Work

Given the methodological difficulties in the pilot study and the constraints placed on the design of the pilot study, there is evidence of fairly good construct validity and internal reliability in AIM.

However, further validation work should be carried out with larger subject numbers and further replication of the validity investigations carried out in the pilot study. Future replications should include a wider range of percentages of datalink-equipped aircraft present, in order to examine the marginal utility effect of the datalink technology. In addition, future validation work should use different types of automated systems to explore the ability of the other cognitive function sub-scales to discriminate between the impact on mental workload in automated ATM systems. For the purpose of validation research, it may also be advisable to use a study design where the difference in impact on mental workload is compared with a baseline system.

Future validation work should include other subjective assessment tools as convergent and discriminant validity measures. In addition, future investigations could include other types of ATM system change.

Further work can also be done to examine the task item groupings and make potential changes to the task item grouping to strengthen further the validity of AIM. If and when the PUMA conflict matrix is revised, then task item grouping for high resource competition task performance should be revisited.

REFERENCES

Archer, R. & Lockett, J. (1997). WinCrew: A tool for analysing performance, mental workload and function allocation among operators. In: *Proceedings of ALLFN 97*, *Revisiting the allocation of function issues*, Vol. II, pp157-165, IEA Press. (WinCrew Web site: http://www.maad.com/MaadWeb/products/wincrew/wincrwma.htm).

Billings, C. E. (1991). Toward a human-centred aircraft automation philosophy. *International Journal of Aviation Psychology 1(4),* 261-270.

David, H. (EUROCONTROL unpublished summary report). *Classification of Techniques for the Measurement of Mental Workload (MWL)*.

Day, P.O., Hook M.K., Warren, C. & Kelly, C. (1993). The modelling of air traffic controller workload. In: *Proceedings Workload Assessment and Aviation Safety*, London, 27-28 April, Royal Aeronautical Society.

EATCHIP Human Resources Team (1997). *Model of the Cognitive Aspects of Air Traffic Control*. HUM.ET1.ST01.1000-REP-02. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATMP (2000). *Human Resources Programme – Stage 1: Programme Management Plan*. Ed. 1.0. Brussels: EUROCONTROL.

EATM Human Resources Team (2003a). *Guidelines for Trust in Future ATM Systems: A Literature Review*. HRS/HSP-005-GUI-01. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2003b). *Guidelines for Trust in Future ATM Systems: Measures*. HRS/HSP-005-GUI-02. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2003c). *Guidelines for Trust in Future ATM Systems: Principles*. HRS/HSP-005-GUI-03. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2003d). *The Development of Situation Awareness Measures in ATM Systems*. HRS/HSP-005-REP-01. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2003e). *Age, Experience and Automation in European Air Traffic Control*. HRS/HSP-005-REP-02. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2004a). *A Measure to Assess the Impact of Automation on Teamwork*. HRS/HSP-005-REP-07. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2004b). *Age, Experience and Automation in European Air Traffic Control*: *Survey in the ECAC Area*. HRS/HSP-005-REP-05. Ed. 1.0.. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2004c). *Impact of Automation on Future Controller Skill Requirements and a Framework for their Prediction*. HRS/HSP-005-REP-04. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

EATM Human Resources Team (2004d). *Managing System Disturbances in ATM: Background and Contextual Framework*. HRS/HSP-005-REP-06. Ed. 1.0.. Released Issue. Brussels: EUROCONTROL.

Eggemeier, F. T. & Wilson, F. G. (1991). Performance-based and subjective assessment of workload in multi-task environments. In: D. L. Damos (Ed.), *Multiple task performance.* London: Taylor & Francis.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In: P. A. Hancock & N. Meshkati (Eds), *Human mental workload.* Amsterdam: North Holland.

Endsley, M. R. & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task, *Ergonomics*, 42, 462-492.

Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation, *Human Factors,* 37, 381-394.

Ericsson, K. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87, 215-250.*

Gopher, D. (1986). In defence of resources. In: R. Hockey, A. Gaillard & M. Coles (Eds), *Energetics and human information processing.* Boston: Martinus Nijhoff.

Hamilton, I. (2000). Cognitive task analysis using ATLAS. In: J. M. Schraagen, S. F. Chipman & V. L. Shalin, *Cognitive Task Analysis*, Lawrence Erlbaum Associates, 215-236.

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

Harwood, K., Sanford, B. D. & Lee, K. K. (1998). Developing ATC automation in the field: It pays to get your hands dirty. *The Air Traffic Control Quarterly.*

Hicks, M. R. (1994). *AWAS (Aircrew Workload Assessment System) Analytical workload modelling tool - research and development programme*. Sowerby Research Centre, British Aerospace PLC.

Hill, S. G., Lavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L. & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors*, 34, 429-439.

Hopkins, V. D. (1995). *Human Factors in Air Traffic Control.* London: Taylor & Francis.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: implications for their use in the assessment of perceived mental workload, *Human Factors*, 33, 17-33.

O' Donnell, R. D. & Eggemeier, F. T. (1986). Workload assessment methodology. In: K. R. Boff, L. Kaufman & J. P. Thomas (Eds.) *Handbook of perception and human performance.* (Vol 2). New York: John Wiley.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: use, misuse, disuse, abuse. *Human Factors, 39(2),* 230-253.

Parasuraman, R., Molloy, R. & Singh, I. L. (1993), Performance consequences of automation induced complacency, *International Journal of Aviation Psychology*, 3, 1-23.

Reid, G. B. & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In: P. A. Hancock & N. Meshkati (Eds), *Human mental workload.* Amsterdam: North Holland.

Robinson, C. & Eberts, R. (1987). Comparison of speech and pictorial displays in a cockpit environment. *Human Factors, 29, 747-755.*

Sperandio, J. C. (1978). The regulation of working methods as a function of workload among air traffic controllers, *Ergonomics, 21, 193-202*.

Timmer, P. & Long, J. (1996). Integrating domain and worksystem models: an illustration from Air Traffic Management. In: A. Sutcliffe, D. Benyon & F. van Assche (Eds), *Domain knowledge for interactive system design.* London: Chapman & Hall.

Trochim, W. (2000). *The Research Methods Knowledge Base*. 2nd Ed. Cincinnati, OH: Atomic Dog Publishing.

Wickens, C. D., Mavor, A. & McGee, J. P. (Eds) (1997). *Flight to the Future: Human Factors in Air Traffic Control.* Washington, DC: National Academy Press.

Wickens, C. D., Mountford, S. J. & Schreiner, W. S. (1981). Multiple resources, task-hemispheric integrity and individual differences in time-sharing efficiency. *Human Factors, 22, 211-229.*

Wickens, C. D. & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30, 599-516.*

Wickens, C. D. (1980). The structure of attentional resources. In: R. Nickerson (Ed.), *Attention and Performance VIII.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Wickens, C. D. (1991). Processing resources and attention. In: D. L. Damos (Ed.), *Multiple task performance.* London: Taylor & Francis.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance.* 2nd Ed. New York: Harper Collins.

Wickens, C. D. (1999). Automation in air traffic control: Human performance issues. In: M. W. Scerbo & M. Mouloua (Eds.), *Automation Technology and Human Performance: Current research and trends,* 2-10. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wiener, E. L. (1985). Beyond the sterile cockpit. *Human Factors, 27,* 75-90.

Woods, D. D. (1993). The price of flexibility in intelligent interfaces. *Knowledge-Based systems,* 6*,* 1-8.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In: R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications,* 3-17. Hillsdale, NJ: Lawrence Erlbaum Associates.

FURTHER READING

Ackerman, P. L. & Wickens, C. D. (1982). Methodology and the use of dual- and complex-task paradigms in human factors research. In: *Proceedings of the Human Factors Society 26th Annual Meeting,* 354-358*.* Santa Monica, California: Human Factors Society.

EATCHIP Human Resources Team (1998). *Integrated Task and Job Analysis of Air Traffic Controllers – Phase 1: Development of Methods.* HUM.ET1.ST01.1000-REP-03. Ed. 1.0. Released Issue. Brussels: EUROCONTROL.

Gopher, D. & Donchin, M. (1986). Workload – An Examination of the concept. In: K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance.* (Vol 2). New York: John Wiley.

ABBREVIATIONS AND ACRONYMS

For the purposes of this document the following abbreviations and acronyms shall apply:

AIM	A tool for the Assessment of the Impact on Mental workload *(EATM(P), HRS, HSP, SHAPE)*
AIM-Cog	AIM-Cognitive *(EATM(P), HRS, HSP, SHAPE; a version of the AIM Tool)*
AIM-Hi	AIM-High-level *(EATM(P), HRS, HSP, SHAPE; a version of the AIM Tool)*
ATC	Air Traffic Control
ATCO	Air Traffic Controller / Air Traffic Control Officer *(US/UK)*
ATLAS	Air Traffic Land and Airborne System *(EU)*
ATM	Air Traffic Management
AWAS	Aircrew Workload Assessment System
BR	Blink Rate
CTAS	Controller TRACON Automation System
CWS	Controller Workstations
DAP	Director(ate) ATM Programmes *(EUROCONTROL Headquarters, SD)*
DAS	Director(ate) ATM Strategies *(EUROCONTROL Headquarters, SD)*
DAS/HUM *or just* HUM	Human Factors Management Business Division *(EUROCONTROL Headquarters, SD, DAS; formerly known as 'DIS/HUM' or just 'HUM')*
DIS	Director(ate) Infrastructure, ATC Systems and Support *(EUROCONTROL Headquarters, SDE)*
DIS/HUM *or just* HUM	Human Factors and Manpower Unit *(EUROCONTROL Headquarters, SDE; formerly stood for 'ATM Human Resources Unit'; today known as 'DAS/HUM' or just 'HUM')*
DNV	Det Norske Veritas

DSE	Display Screen Equipment
EATCHIP	European Air Traffic Control Harmonisation and Integration Programme *(later renamed 'EATMP' and today known as 'EATM')*
EATM(P)	European Air Traffic Management (Programme) *(formerly known as 'EATCHIP')*
ECAC	European Civil Aviation Conference
ECG	Electro-CardioGram
EEC	EUROCONTROL Experimental Centre *(France)*
EEG	Electro-EncephaloGram
EMT	Eye-Movement Tracking
ET	Executive Task *(EATCHIP)*
EU	European Union
FDP	Flight Data Processing
FDPS	FDP System
FPS	Flight Progress Strip
GUI	Guidelines *(EATCHIP/EATM(P))*
HFFG	Human Factors Focus Group *(EATM, HRT; formerly known as 'HFSG')*
HFSG	Human Factors Sub-Group *(EATCHIP/EATMP, HRT; today known as 'HFFG')*
HR	Heart Rate
HRS	Human Resources Programme *(EATM(P))*
HRT	Human Resources Team *(EATCHIP/EATM(P))*
HRV	Heart-Rate Variability
HSP	Human Factors Sub-Programme *(EATM(P), HRS)*
HTLA	High Traffic and Low Automation
HUM	Human Resources (Domain) *(EATCHIP, EATMP)*
IBI	Inter-Beat Interval
ISA	Instantaneous Self-Assessment

LTHA	Low Traffic and High Automation
LTLA	Low Traffic and Low Automation
MRT	Multiple Resource Theory
MS	MicroSoft
NASA	National Aeronautics and Space Administration *(US)*
NATS	National Air Traffic Services Ltd. *(UK)*
OTA	Observational Task Analysis
PD	Pupil Diameter
PUMA	Performance and Usability Modelling in ATM
REP	Report *(EATCHIP/EATM(P))*
SAS	Statistical Analysis System
SD	Senior Director, EATM Service Business Unit *(EUROCONTROL Headquarters; formerly known as 'SDE')*
SDE	Senior Director, Principal EATMP Directorate, *or, in short*, Senior Director(ate) EATMP *(EUROCONTROL Headquarters; now known as 'SD')*
SHAPE (Project)	Solutions for Human-Automation Partnerships in European ATM (Project) *(EATM(P), HRS, HSP)*
SME	Subject Matter Expert
SPSS	Statistical Package for the Social Sciences
ST	Specialist Task *(EATCHIP)*
SWAT	Subjective Workload Assessment Technique
TLX	Task Load Index *(NASA, US)*
TRACON	Terminal Radar Approach Control (facility)
URET	User Request Evaluation Tool
WAT	Workload Assessment Tool

Page intentionally left blank

CONTRIBUTORS

The contribution of the Members of the HRT Human Factors Sub-Group (HFSG), now known as the Human Factors Focus Group (HFFG), to this document during the group meetings, and further written comments, were much appreciated.

<u>Document configuration</u>

Carine Hellinckx EUROCONTROL Headquarters, DAS/HUM
(External contractor)

Page intentionally left blank

APPENDICES

Page intentionally left blank

APPENDIX A: A SUMMARY OF THE LITERATURE REVIEW CONDUCTED ON THE IMPACT OF AUTOMATION ON MENTAL WORKLOAD

The *raison d'être* for automation is to increase capacity and optimise workload of the human operator, whilst maintaining the safety integrity of the total system (i.e. hardware, software, operator and procedures). Different functions in ATM can be automated. Automated information integration and inference can reduce human working memory demands. Such systems can also provide superior integrated information to operators and thereby reduce excessive information processing workload. Automated ATM systems with functions that draw inferences about future events (prediction) can provide assistance to improve decision-making and may decrease thinking and interpretation workload (Harwood, Sanford & Lee, 1998).

On the other hand, some literature have suggested that workload may not be reduced by automated assistance provided, even under full automation (Billings, 1991; Wiener, 1985; Parasuraman, Molloy & Singh, 1993). These propose that there is a shift in the type of workload such as a greater monitoring load. When automation is implemented, operators may incur additional workload associated with input devices, errors, intrinsic ‘automation-management’ tasks and unexpected and new peripheral tasks needed to use the automation. New ATM tasks may also require more time thinking about and understanding the options or decisions generated by the automation. The workload analysis therefore needs to focus on the evaluation of the mental workload as a result of this shift towards knowledge-based thinking. Moreover, such knowledge-based thinking may interfere or compete for mental resources required for other ATM tasks.

Automated information integration and inference can reduce human working memory demands. A model of verbal report data suggests that what is consciously perceived and demands on working memory often influences subjective reports or ratings (Ericsson & Simon, 1980). This suggests that automation of information integration and inference may ease subjective workload or perceived workload.

ATM automated systems with functions that draw inferences about future events (prediction) can improve decision-making while reducing workload (Harwood, Sanford & Lee, 1998; Wickens, 1999). Two examples were cited. There were the User Request Evaluation Tool (URET) (Wickens, 1999) and the Controller TRACON Automation System (CTAS) (Wickens, 1999).

Apart from different functions that can be automated, there are also different levels of automation. In intermediate levels of automation, by keeping the human involved in system operations, better work system performance and lower workload can be achieved than in highly automated systems (Endsley & Kiris, 1995). For example, levels of automation which produced significantly lower subjective workload rating on NASA-TLX than other levels of automation

(Endsley & Kaber, 1999) were levels where decision-making functions were automated either joint decision-making between human and computer or fully automated decision-making. In the same study all the sub-scales of NASA-TLX correlated positively with the overall workload score, except performance. When performance rating was high, workload rating was low. Ratings of operator perceived successfulness in the task increased with increasing level of automation.

On the other hand, some literature have suggested that workload may not be reduced by automated assistance provided, even under full automation. They propose that there is a shift in the type of workload as a greater monitoring load is incurred (Billings, 1991; Wiener, 1985).

In short, the research findings on the impact of automation on workload appear inconsistent and inconclusive because of several methodological difficulties:

- There are many workload evaluation tools, which are common and easy to use in studies on automation. Different workload evaluation tools were used in the various studies. These simple but different workload evaluations may be measuring different aspects of workload or even different concepts of workload.

- Cognitive task performances in complex work environments such as ATM consist of several key cognitive function groups. Mental workload in such complex work environment is multi-dimensional. The demands placed on each of its functional dimensions contribute to the overall experience of mental workload. In addition, automation systems may have differential and specific impact on individual dimensions of MWL. Not all workload evaluation tools measure mental workload. Of those that do, very few or none, partition the mental workload into its different dimensions.

- The automation systems studied may have provided assistance for different cognitive functions. Hence, the impact of automation on workload is inconsistent and may vary, not only depending on tool or technique used but also on the type of cognitive functions being automated.

The research findings from the current literature and observed methodological difficulties have important implications on the development and design of a mental workload assessment tool such as AIM, namely how the tool was conceived, the development of the tool's methodology, what the tool should measure, what dimensions of MWL it should describe and explain, and finally the format and content of the tool (e.g. what sub-scales and how many sub-tests, etc.).

APPENDIX B: A SHORT DESCRIPTION OF THE MULTIPLE RESOURCE THEORY (MRT) MODEL OF MENTAL WORKLOAD

While there is evidence supporting the single resource theory (Kahneman, 1973; Gopher, 1986), it does not account for three phenomena. *Task insensitivity* and *structural alteration effects* describe the phenomenon in which changes in task difficulty and structure, respectively, in one task appear not to affect performance in another concurrent task. The other phenomenon is *time-sharing,* in which two tasks, both of non-trivial difficulty, are performed concurrently with no performance decrement, even though each can be shown to interfere with other activities (Wickens, 1991). Wickens (1980) analysed characteristics of task pairs and noted consistency of these three phenomenal effects along three structural dimensions in the information processing system.

The Multiple Resource Theory (MRT) postulates that there are different mental resources along three different dimensions (see Figure 2). The dimensions are stage (of cognitive processing), modality (of information input or output) and the processing code (of the information).

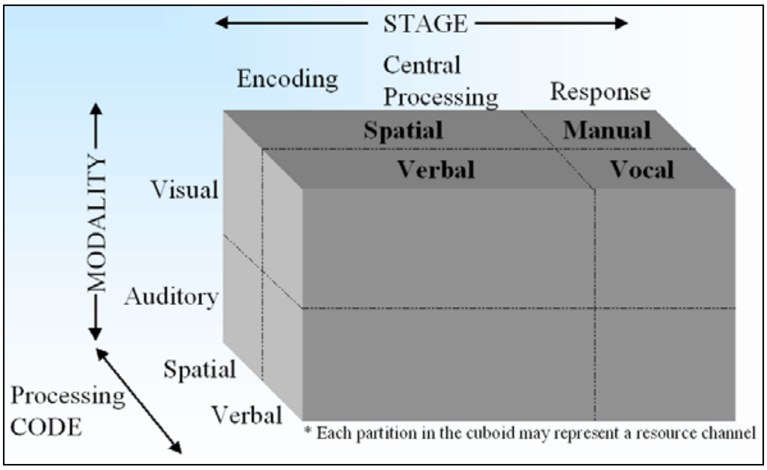


Figure 2: Diagrammatic representation of MRT

On the *stage-defined dimension*, resource channels used for *encoding and central processing* activities are separate from those used in *selection and execution of responses*. Manzey (1989) (cited in Wickens, 1991) found mutual interference and large performance trade-offs between a cursor positioning task and a switch-throwing task. However, no interference or performance trade-off was found when a mental arithmetic task was time-shared with either task. Instead, a mental arithmetic task was found to interfere with a Sternberg memory search task. In ATM the ATCO is able to acknowledge vocally each change in aircraft flight (a response demand) without disruption to his ability to maintain situation awareness of the airspace (perceptual-cognitive demand).

Separate resource channels are also defined along a *processing code dimension. Spatial* and *verbal* processes in perception, working memory and response demand separate resources. Examples of evidence of this can be found in research studies where verbal decision task performance was found to be disrupted more by vocal responses, while spatial decision task performance was disrupted more by manual responses (Wickens & Liu, 1988). When the verbal and spatial decision tasks were time-shared with a manual tracking task, the spatial decision-manual response task interfered most with the tracking task. Although there is evidence for separate spatial and verbal processing resources, cognitive activities may be ambiguous with respect to which resource channel they employ. Visual stimulus such as pictures, icons or geometric symbols, though non-verbal, may rapidly activate verbal codes (Robinson & Eberts, 1987). The ATCO can change the heading of an aircraft by arithmetic calculation (a verbal strategy) or imagining the vectors of the aircraft in the airspace (a spatial visualisation strategy). Travel routes can be learned as a series of verbal lists of instructions or spatial images (airspace maps of airways and air routes).

The third dimension is defined by the *perceptual modality*, in other words, visual versus auditory input. However, it is not clear if interference, say between two visual tasks, is a consequence of resource competition or due to visual scanning delays as the visual stimuli may be far apart and not within the fovea simultaneously.

The different channels describe how the resources can be accessed during performance. If two tasks place demands on the same channel, they have to compete with each other for resources and workload will increase. Different channels may also conflict when they access resources that are on the same dimension. For example, the ATCO may be deciding to change the altitude of an aircraft by calculating how far it is from another aircraft (verbal processing). At the same time, he may be visualising how to alter the heading of another aircraft (spatial processing). Although both tasks require separate code channels (verbal versus spatial), they conflict on the stage dimension (both are at the processing stage) and modality (both require visual input) channels. The MRT is useful to human factors practitioners for assessing time-sharing and multi-task performance in complex jobs such as ATM and predicting workload associated with time-sharing concurrent tasks.

Time-sharing efficiency as a skill

In addition to the demands on resources from the tasks, time-sharing two tasks also contributes to workload. However, efficient time-sharing skills can be cultivated by (i) an improvement in the skills of the single-task components, and (ii) developing a distinct skill in time-sharing that results explicitly and exclusively from multitasking experience. With practice, two tasks will demand fewer resources and become more data-limited (that is diverting more resources to the task does not affect performance). Improved dual task performance can be the result of better skills in the single task components and may be acquired via single task practice. Based on the MRT it has been suggested that operators select which tasks to perform ‘specifically to place demands upon qualitatively different capacities of processing resources’

(Wickens, Mountford & Schreiner, 1981). When separate analyses of dual task components were conducted and the results found suggested support for the prediction in Wickens *et al.* (1981). There is some evidence from studies which found that subjects trained in time-sharing strategies which emphasised resource and attention allocation performed better than control groups or in control tasks (Wickens, 1992).

Hence, efficient time-sharing skills may develop as a result of improved single task performance and a true skill in time-sharing. The latter skill is described as ‘knowing when to sample what from the display, when to make which response and how to integrate better the flow of information in the two tasks’ (Wickens, 1992). This proposition appears to be consistent with the skills developed in an experienced ATCO (Hopkins, 1995; Wickens, Mavor & McGee, 1997). Thus, for the experienced ATCO, the combined resource demand will be diminished and workload is moderated as time-sharing skills develop.

Page intentionally left blank

APPENDIX C: REVIEW OF MENTAL WORKLOAD ASSESSMENT TECHNIQUES

There are three workload-measurement groups: performance-based, subjective (i.e. self-report) and physiological measures. In addition to these three methods in workload assessment, one can also assess mental workload by using mental modelling techniques. Each group of technique will be briefly described in the following sections.

Performance-based assessment techniques

Performance-based assessment techniques utilise some aspect of the operator's performance or capability to perform the tasks or system functions, in order to provide an estimate of the workload. There are three common types of such techniques in the assessment of workload literature.

Primary-task methodology assesses mental workload by examining the level of operator performance as an indicator of workload levels, based on the hypothetical workload-performance relationship (Eggemeier, 1988; O' Donnell & Eggemeier, 1986) (see Table 5). As demands from a task or tasks increase, the proportion of total resources expended on workload associated with tasks also increases. Performance at criterion levels or better is associated with workload levels that fall within a low to moderate range, implying that the operator still has the capability to compensate for increasing workload through resource allocation. As resource expenditure approaches the upper limits of the first workload region, the operator begins to exhibit performance decrements, he is no longer able to perform the tasks at criterion level. In this second workload region, the operator is said to exceed the 'threshold for unimpaired performance' (Eggemeier & Wilson, 1991) and will show gradual degradation of performance. In the third workload region, performance is consistently poor and it is associated with high levels of workload.

Table 5: Relationship between performance and workload regions

MWL levels	First region (low–moderate)	Second region (high approaching overload)	Third region (very high or overload)
Performance quality	Criterion level or better	Performance degradation (below criterion level)	Consistently low performance

However, primary task measures have been found to be relatively insensitive to mental workload changes in the first or third regions. These measures will differentiate between first and second regions, and exhibit sensitivity to workload changes in the second region. Primary-task performance as a measure may reflect the overall effectiveness of man-machine interaction (O' Donnell & Eggemeier, 1986). A human operator may be able to compensate for increase in workload and still maintain good and above criterion level performance (i.e. first region). Hence, a constant performance in the first region does not necessarily reflect low operator workload. In addition, as a consequence of resource compensation in situations where workload increases, two operators may exhibit no obvious performance differences even though one of them may be performing at his/her capacity limits. It has been suggested that it may therefore be necessary to combine primary task measures with other workload measurement techniques.

Secondary task methodology measures the operator’s capability to perform an additional task or function concurrently with the primary task. Spare or reserve processing capacity not demanded by the primary task is assumed to be allocated to secondary task performance. An index of primary task workload is derived from the levels of concurrent task performance. There are several ways of implementing this technique:

- The first is to use the `loading task paradigm'. Subjects are instructed to maintain secondary-task performance, even if decrements in primary-task performance occur. The addition of the second task results in a total workload shift from the first region towards second region, so that primary-task performance measures can be used as indicators of workload.

- The second way to implement the technique is to use the `subsidiary task paradigm'. Subjects are instructed to maintain *primary*-task performance. As a result, secondary task performance is allowed to vary with difficulty. This is taken to indicate the spare capacity available in order to perform the secondary tasks.

However, it is difficult to conclude whether the resulting performance data is due to actual spare capacity or the intrusiveness of the unfamiliar secondary task, or even an artefact of the task priority defined by the experimenter (Wickens, 1992).

These paradigms also assume an undifferentiated capacity or mental resources available for all types of task performance. There is evidence against these and phenomenon such as efficient time-sharing of tasks and task insensitivity and structural alteration effects (described in Appendix B) demonstrate against the assumption of an undifferentiated mental resource.

Although there are ways to compensate in part for some of these methodological problems, primary and secondary task measures as a technique for assessing mental workload may not be suitable for our requirements. In dynamic and multitasking environments such as ATM, it is difficult to define primary and secondary task measures. Moreover, ATCOs have been found to be able to manage their workload and regulate their performances. They employ different adaptive or compensatory information processing strategies in response to increases or decreases in workload. For example, ATCOs decreased the amount of time they spent processing each aircraft in response to an unexpected increase in air traffic load, but when air traffic load was low, ATCOs paid more attention to peripheral details of the aircraft (Sperandio, 1978). Hence, an ATCO may change task strategies as well as resource allocation strategies accordingly to maintain performance at a desired level or a criterion level set by the authority.

Subjective assessment techniques

Subjective measures assess the subjective experience of the workload by the operators by requiring them to provide judgements of the workload or effort associated with performance of a task or system functions. Several studies have demonstrated the capability of these rating tools to reflect variation in demands across a variety of tasks (for a review see Eggemeier & Wilson, 1991). Many of these tools involve retrospective ratings of workload experienced and have been widely utilised for workload assessment in multi-task environments such as ATM. However, workload estimates from rating scales may constitute an 'averaged-out' measure of overall workload. This is of little diagnostic value in the evaluation of the cause of intensive workload in system design. On the other hand it has been suggested that diagnosticity can be better for multi-dimensional scales (Nygren, 1991; Hill, Lavecchia, Byers, Bittner, Zaklad & Christ, 1992). Common subjective workload assessment tools used in the literature include the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) and Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988).

Physiological measures

Changes in physiological reactions have been used as indicators of Mental Workload (MWL). These techniques depend on different physiological measures which are differentially sensitive to either global or specific arousal responses. Such techniques do not require an overt action on the part of the subject for the measurement. It is also a continuous measurement of workload. However, it does depend on the availability of expensive, specialised and often cumbersome equipment. Even with miniaturisation of some equipment, such techniques may be dependent on specialised expertise and sensitive to other extraneous 'noise' effects. Three main types of physiological measurements can be used as indicators of workload. Their feasibility and suitability in ATM research and the problems they present have been reviewed (David, EUROCONTROL unpublished summary report). Table 6 displays a list of physiological measures with a brief and simple interpretation of what each indicates about workload.

Table 6: List of physiological measures and their indicators of workload

Physiological technique	Measure	Workload indicators
Electro-cardiogram (ECG)	Heart Rate (HR) or Inter-Beat Interval (IBI)	An increase in HR or IBI indicates an increase in workload; largely affected by physical workload
	Heart-Rate Variability (HRV)	A decrease in HRV indicates an increase in MWL
Eye-movement Tracking (EMT)	Pupil Diameter (PD)	A decrease in PD indicates an increase in MWL
	Blink Rate (BR)	An increase in BR indicates an increase in strain or fatigue which may be inferred as an increase in workload.

Physiological technique	Measure	Workload indicators
Electro-encephalogram (EEG)	Activity on four main frequency bands: ▫ 8-12 Hz (alpha), ▫ 13-30 Hz (beta), ▫ 4-7 Hz (theta) and ▫ 1-3 Hz (delta)	Increases in brain activity indicate an increase in mental activity and hence MWL

Modelling mental workload in ATM

Assessment of mental workload using modelling techniques require a model of the interrelationships between the events in the work domain (i.e. air traffic events), task performance and mental workload. When assessing or predicting workload, the tool needs to model the ATM events or tasks that resulted in the workload. The ATM events and tasks also affect the strategies the ATCO employs to manage and regulate the performance. Common techniques used in mental workload modelling are task analysis and timeline analysis. Unlike primary and secondary task methodologies where the quality of performance is measured, models of workload are based on a task analysis of the ATM performance and are represented on a timeline format. Workload is assessed or predicted using a theory of mental workload (such as MRT) and its underlying assumptions about the performance of tasks and associated workload. For example, if the task analysis and timeline models a situation where the heading of two aircraft are managed simultaneously, the workload model, using say MRT, will predict high workload due to task demands and resource competition. Workload modelling tools are methodologies for carrying out task analyses and for making workload predictions based on task analyses using algorithms related to a particular theoretical model of mental workload. Examples of such tools are:

- ‘Performance and Usability Modelling in ATM (PUMA)’, a custom-built tool for NATS (Day, Hook, Warren & Kelly, 1993),
- ‘Aircrew Workload Assessment System (AWAS)’ (Hicks, 1994),
- ATLAS (Hamilton, 2000), and
- WinCrew (Archer & Lockett, 1997 - URL address: http://www.maad.com/MaadWeb/products/wincrew/wincrwma.htm).

APPENDIX D: REVIEW OF EXISTING MENTAL WORKLOAD ASSESSMENT TOOLS

The review of existing mental workload assessment tools was conducted in two stages:

- In Stage 1 a large selection of tools was assessed against the following criteria, defined from the requirement specifications in the SHAPE Project for such a tool (see Section 2.3):

  (i) Multi-dimensional: The tool accounts for multi-dimensions of mental workload.

  (ii) Practical and easy to use: In terms of preparation/resources required, cost of training required to use the tool.

  (iii) Multi-scaled and diagnostic: Multi-scales and ability to diagnose the origins of mental workload and inform design and recommendations for design changes in ATM system (automation, equipment, procedures, re-sectorisation, etc.).

  (iv) Sensitivity to differences due to traffic situation and ATM system design.

  (v) Theoretical background: Has a theoretical background and able to provide comprehensive explanations of mental workload.

  (vi) Critical cognitive activities: Mental workload due to critical cognitive activities are evaluated.

  (vii) Reliability and validity.

  (viii) Data format: Format of the data collected to assess mental workload.

- The selection based on Stage 1 review was then subjected to a second set of review criteria, defined by the purpose and scope given in Section 4.2:

  (i) *What* aspect of mental workload does the tool measure: Mental effort, task difficulty, processing load on mental resources, processing capacity;

  (ii) *Where* does the tool assume mental workload is likely to occur: Generic/global task performance, specific tasks/operations, performance degradation, HMI or internal physiological reactions; and

  (iii) Its assumption about *why* mental workload occurred: Single undifferentiated resources, multiple resources with resource conflicts and competition between different types of resources, attention bottlenecks in task performance.

The outcome from each stage will be reported in the following sections in this appendix. The tools reviewed in the first stage were as follows:

- Cooper-Harper Rating Scale
- Modified Cooper-Harper Rating Scale
- Bedford Workload Scale
- Crew Status Survey
- Honeywell Cooper-Harper Rating Scale
- NASA-TLX
- SWAT
- Zachary/Zachlad Cognitive Analysis
- PUMA
- ISA

The following tools were selected as a result of the Stage 1 review for Stage 2 review:

- Cooper-Harper Rating Scale
- Modified Cooper-Harper Rating Scale
- Bedford Workload Scale
- NASA-TLX
- SWAT
- PUMA

Stage 1 of the review

Technique	Description
Cooper-Harper Rating Scale	The Cooper-Harper Rating Scale is the current standard for evaluating aircraft handling qualities. It makes use of a decision tree that assesses adequacy for task, aircraft characteristics, and demands on the pilot to calculate and rate the handling qualities of an aircraft.
Criteria	**Review**
Multi-dimensional	Not: The Cooper Harper Scale deals explicitly with aircraft handling.
Practical and easy to use	Yes: ▪ Requires minimal training. ▪ A briefing guide has been developed. ▪ The task in question must be fully defined – to allow common reference when using the decision tree.
Multi-scaled and diagnostic	Yes: ▪ It is reported as being multi-scaled and diagnostic for aircraft handling factors such as the effect of: - wind gust, - aircraft pitch stability and - acceleration control. ▪ Sensitive to the psychomotor demands on an operator.
Situation/design sensitivity	Yes: Cooper-Harper ratings have been sensitive to variations in controls, displays and aircraft stability.
Theoretical background	No apparent theoretical model. However, the Cooper-Harper scale uses the following definition of pilot workload: ‘workload is the integrated mental and physical effort required to satisfy the perceived demands of a specific flight task’. The scale should only be used for workload assessment if handling difficulty is the major determinant of workload.
Consider all critical cognitive activities	Does not consider cognitive activities.
Reliability and validity	Have been found to be reliable in assessments of workload in aircraft handling, sensitive to variations in controls, displays and aircraft stability.
Method and data	Ordinal data. Requires non-parametric analysis. Decision tree with a scale which provides a rating from 1 (excellent) to 10 (major deficiencies).
Outcome	Although it does not satisfy all the criteria, as a tool it is easy to use and practical. It is selected for the second stage review as its methodology is useful and there is research evidence of its usefulness and sensitivity.

Technique	Description
Bedford Workload Scale	The Bedford Workload Scale is a modification of the Cooper-Harper Scale. It is intended to cater for a wider variety of pilot tasks (especially for systems that load the perceptual and communication channels). It was created by trial and error with the help of Royal Aircraft test pilots. It retained the decision tree and the four- and ten-rank ordinal structures of the Cooper-Harper Scale.
Criteria	**Review**
Multi-dimensional	Yes: The decision tree contains the terms, effort, spare capacity, attention and workload (insignificant, low, very high and extremely high).
Practical and easy to use	Yes: ▪ It has been reported that pilots found the scale easy, convenient and unobtrusive to use. ▪ Practice is required to become familiar with the scale.
Multi-scaled and diagnostic	Yes: May provide a good measure of spare capacity.
Situation/design sensitivity	No: There are question marks over the sensitivity of this scale. The scale was found not be to be sensitive to control configurations and counter measure conditions.
Theoretical background	The Bedford Workload Scale used the same definition of workload as the Cooper-Harper Scale: 'workload is the integrated mental and physical effort required to satisfy the perceived demands of a specific flight task'. In addition to the definition of workload the Bedford Workload Scale used the concept of spare capacity to aid in the definition of workload levels.
Consider all critical cognitive activities	The scale does mention, spare capacity, attention and effort.
Reliability and validity	Roscoe (1984) reported that pilot workload ratings and heart rates varied in similar manners. This may indicate a degree of convergent validity. There is evidence to suggest that the Bedford Workload Scale is a reliable and valid measure of workload (based on flight simulator data).
Method and data	Decision tree with a scale that provides a ten-point rating scale using the following end-points: workload insignificant to task abandoned. Requires non-parametric analysis, as it is not an interval scale.
Outcome	Although it does not satisfy all the criteria, as a tool it is easy to use and practical. It is selected for the second stage review as its methodology is useful and there is research evidence of its usefulness.

Technique	**Description**
Honeywell Cooper-Harper Rating Scale	The Honeywell Cooper-Harper Rating Scale uses a decision-tree structure to assess overall workload related to aircraft controllability.
Criteria	**Review**
Multi-dimensional	No: The scale concentrates upon the measurement of overall workload. Considers subjective workload and effort.
Practical and easy to use	Yes (see responses for Cooper-Harper Scale).
Multi-scaled and diagnostic	No: The scale requires participants to answer three questions, generally referring to a subjective evaluation of effort required to complete the task.
Situation/design sensitivity	Unknown.
Theoretical background	Unknown.
Consider all critical cognitive activities	No: Does not consider cognitive activities individually – concentrates upon eliciting a subjective measure of overall workload.
Reliability and validity	For a small sub-set of conditions the scale ratings correlated well with performance.
Method and data	Decision tree with a scale that provides a nine-point rating scale. The end-points of the scale are pilot effort not a factor for desired performance to control will be lost during some portion of required operation.
Outcome	Not selected for Stage 2 review.

Technique	Description
Modified Cooper-Harper Rating Scale	The Modified Cooper-Harper Rating Scale is a modification of the Cooper-Harper Scale, intended to produce an estimation of the workload associated with cognitive functions such as perception, monitoring, evaluation, communications and problem solving.
Criteria	**Review**
Multi-dimensional	No: The Modified Scale makes reference to mental effort.
Practical and easy to use	Yes: Same as Coopers Harper Rating Scale.
Multi-scaled and diagnostic	Not multi-scaled but diagnostic. The Modified Scale makes reference to the term mental effort. Have been used to estimate mental effort associated with cognitive functions such as perception, monitoring, evaluation, communications and problem solving. Assessments of the scale have focussed on perception, cognition, and communications. The system is designed for use in experimental conditions, therefore it may not be appropriate for situations requiring an absolute diagnosis of a subsystem.
Situation/design sensitivity	The Modified Scale is reported to be sensitive to: - communication load, - navigation load, - danger conditions, - flight conditions and - crew positions. It has been reported that the Modified Scale were as sensitive to task difficulty as SWAT but less sensitive than SWAT ratings to changes in tracking task difficulties.
Theoretical background	Same as Cooper Harper Rating Scale.
Consider all critical cognitive activities	Subjects are told to consider perception, cognition and communication.
Reliability and validity	Results suggest that the scale is a valid, statistically reliable indicator of overall mental workload. The scale is reported to provide consistent and sensitive ratings of workload across a range of tasks.
Method and data	Decision tree with a modified Cooper Harper Scale with a ten-point scale, ranging from very easy/highly desirable to impossible.
Outcome	Although it does not satisfy all the criteria, as a tool it is easy to use and practical. It is selected for the Stage 2 review as its methodology is useful (mental effort ratings made on instructed tasks such as communication, problem solving) and there is research evidence of its usefulness.

Technique	Description
Crew Status Survey	The Crew Status Survey provides measures of self-reported fatigue and workload. It also includes space for general comments.
Criteria	**Review**
Multi-dimensional	No: Self-report estimate of fatigue and overall workload.
Practical and easy to use	The Crew Status Survey is presented on cards – participants find it difficult to fill in the rating scale during periods of high workload. Verbal ratings have been found to work more effectively if; subjects can quickly scan a card copy of the rating to confirm the meaning of the rating and secondly if the subjects are not performing a conflicting verbal task. Fatigue and workload scale can be used independently.
Multi-scaled and diagnostic	No.
Situation/design sensitivity	The scales have been found to be sensitive to changes in task demand and fatigue. The fatigue and workload scales are independent of each other.
Theoretical background	Fatigue and strain, workload rated globally on a single scale even though it combines temporal demand, system demand, system management danger and acceptability in its definition.
Consider all critical cognitive activities	The survey does not consider the cognitive activities that are listed in SHAPE. The workload scale considers temporal demand, system demand, system management, danger and acceptability.
Validity and reliability	The survey is reported to have face validity – it was well received by pilots. The scales have been tested for test/re-test reliability. Both are now seven-point scales.
Method and data	The Crew Status Survey provides two scales that are used independently, both scales range from 1 to 7.
Outcome	Not selected for Stage 2 review.

Technique	Description
NASA-TLX	The NASA Task Load Index (TLX) is a multi-dimensional subjective workload rating technique. It provides an overall workload score based on a weighted average of ratings on six sub-scales; mental demand, physical demand, temporal demand, own performance, effort and frustration.
Criteria	**Review**
Multi-dimensional	Yes: Considers six dimensions: mental demands, physical demands, temporal demand, own performance, effort and frustration.
Practical and easy to use usable	Yes: ▪ Paper- or pencil-based rating scales. ▪ Requires two steps – the first is a rating task, the second is a pair-wise comparison of the six workload scales. ▪ Delays of fifteen minutes do not affect workload ratings.
Multi-scaled and diagnostic	Yes: TLX is thought to provide a sensitive indicator of overall workload as it differed among tasks of various cognitive and physical demands. The six sub-scales can be used to diagnose the sources of loading within a task.
Situation/design sensitivity	Yes: ▪ Reported to be a good measure of general workload. ▪ It is thought that NASA-TLX is sensitive to different levels of workload. ▪ It is thought that NASA-TLX and SWAT are essentially equivalent in terms of their sensitivity to task manipulations.
Theoretical background	In NASA-TLX workload is defined as the ‘cost incurred by human operators to achieve a specific level of performance’.
Consider all critical cognitive activities	No: Although mental demand is one of it’s six workload scales and is described as mental and perceptual activity, thinking, deciding, calculating, remembering, looking, searching and task complexity, it does not consider the activities separately.
Reliability and validity	High reliability. Used extensively in aviation research.
Method and data	All dimensions are rated on bipolar scales from 1 to 100, anchored at each end with a single adjective (for e.g. high/low, good/poor). An overall workload rating was determined from a weighted combination of scores on the six dimensions.
Outcome	Satisfied most of the criteria. Selected for Stage 2 review.

Technique	Description
SWAT	The Subjective Workload Assessment Technique (SWAT) combines ratings of three different scales (time load, mental effort load and psychological stress load) to produce an interval scale of mental workload.
Criteria	**Review**
Multi-dimensional	Yes.
Practical and easy to use	No: Although SWAT has been found to be relatively unobtrusive measure of workload, it requires some preparation prior to administration. SWAT requires two steps – the first is the development of scales and the second is event scoring. Scale development involves subjects ranking from lowest to highest 27 combinations of three levels of the three workload sub-scales. Event scoring, involves the subject providing a rating (1, 2, 3) for each sub-scale. The experimenter then maps the set of rankings to the SWAT score (0-100) calculated during the scale development.
Multi-scaled and diagnostic	Yes: Each of the scales, time, effort and stress can be examined individually as workload components: ▪ Time load scale considers amount of spare time available in planning, executing and monitoring a task. ▪ Mental effort load considers how much conscious mental effort and planning are required to perform a task. ▪ Psychological stress load considers amount of risk, confusion and anxiety associated with task performance.
Situation/design sensitivity	Yes: It has been reported that SWAT is sensitive to changes in task difficulty.
Theoretically sound	It is reported that SWAT provides a good cognitive model of workload that may be sensitive to individual differences. However, other sources have suggested that three dimensions of workload are not sufficient to assess workload. It is also reported that SWAT failed to detect resource competition effects in dual-task performance.
Consider all critical cognitive activities	No.
Reliability and validity	SWAT is reported to be a valid and sensitive and relatively unobtrusive measure of workload. It is also reported that SWAT ratings are less variable than the Modified Cooper-Harper Rating Scale ratings.
Method and data	SWAT produces a value between 0 and 100.
Outcome	Satisfied most of the criteria. Selected for Stage 2 review.

Technique	Description
Zachary/Zachlad Cognitive Analysis	The Zachary/Zachlad Cognitive Task Analysis Technique requires both operational Subject Matter Experts (SMEs) and 'cognitive scientists' to identify operator strategies for performing all tasks listed in a detailed cognitive tasks analysis. Following this a second group of SMEs then rates, using thirteen sub-scales, workload associated with performing each task.
Criteria	**Review**
Multi-dimensional	Unknown.
Practical and easy to use	No: Considerable effort given to developing a cognitive task analysis. The method requires two sets of SMEs, to develop the task timeline and to rate the associated workload.
Multi-scaled and diagnostic	Unknown.
Situation/design sensitivity	Unknown.
Theoretically sound	Unknown.
Consider all critical cognitive activities	Yes.
Reliability and validity	To date the method has only had limited application.
Method and data	A combination of mental workload modelling and subjective ratings on the model using SMEs.
Outcome	Not selected for Stage 2 review. However, its methodology is interesting and may be useful (task analysis to identify the cognitive task and workload ratings made on cognitive task models by SMEs). The concept in its methodology is similar to that of PUMA.

Technique	Description
PUMA	Performance and Usability Modelling in Air Traffic Management (PUMA) is a predictive workload modelling technique. It produces quantitative models from which predictions about the amount of effort or mental workload involved in ATM operations are made. The PUMA Analysis is conducted in a number of stages: ▪ The first stage involves the identification of ATCO cognitive activities during an air traffic scenario; this is followed by Observational Task Analysis (OTA). The OTA results in the identification of sequential tasks and action time lines. The analyst then works through the OTA with the ATCO to capture the ATCO's performance. ▪ The second stage involves building a generic model of each task. Cognitive activities that are required for each task are inferred from the OTA, video analysis and debrief interview. Each generified task is then refined using the Task Ordering Tool. The tasks are put together in order to build a model of task performance. ▪ The overall workload is then calculated using the Workload Assessment Tool (WAT).
Criteria	**Review**
Multi-dimensional	Yes. Mental demand as well as different mental resources.
Practical and easy to use	No: PUMA involves a number of time consuming and resource intensive stages. PUMA requires considerable SME input. It also requires analyst experience with ATM and PUMA procedures and hence cost of training to use the tool is substantial.
Multi-scaled and diagnostic	Yes: Considers different dimensions of mental workload and models the workload according to air traffic events, ATCO tasks, design configuration and operating procedures (existing and new).
Situation/design sensitivity	Using the information gained from the OTA PUMA has the ability to match intensive workload periods can be matched to ATM tasks and operations. PUMA is also reported to be sensitive to changes in air traffic events and the number of tasks being carried out. The analyst is thought to be able to trace workload peaks to particular design configurations. It is thought that the method should also enable the analyst to identify the effects of deviating from normal/standard procedure (including the task requirements of a new system) upon workload.

Criteria	Review
Theoretical background	Wicken's Multiple Resource Theory (MRT) is central to the PUMA method. The algorithm used by WAT is an implementation of MRT. The workload algorithm involves the concepts of multiple channels, concurrent task performance and task interference.
Consider all critical cognitive activities	Requires an extensive model of ATCO tasks and operating procedures, which include cognitive tasks and functions.
Reliability and validity	Workload algorithm has strong theoretically basis. The validity of the workload models is dependent upon the skills and experience of the analyst. Accurate workload modelling is dependent upon the quality of the OTA. The generification process involves a significant amount of subjective judgement by the analyst. The cognitive activities have to be inferred from the overt actions observed and the debriefs. It is reported that PUMA lacks a structured framework for modelling the cognitive activities in ATM. Pilot validation studies showed that PUMA evaluations dissociate from ISA systematically as expected.
Method and data	A mental workload modelling tool. PUMA produces a predictive workload against time/task rating. Workload peaks are identified as moderate to high workload levels relative to the overall shape of the workload model. It also logs channel interference, demand ratings and workload values generated by the algorithm for the entire duration of the scenario.
Outcome	Although it did not satisfy all of the criteria and is difficult to use and is resource intensive, it was selected for Stage 2 review as its theoretical background and workload algorithm are relevant to ATC. It has also been used in ATM and has been found useful.

Technique	Description
ISA	Instantaneous Self-Assessment (ISA) is a real-time subjective measure of mental workload. The method uses a recorder panel that is built into the ATC workstation. The panel consists of five buttons (numbered 1 to 5). The numbers correspond to ratings of very low workload to very high workload. The ISA recorder flashes two LED lights for thirty seconds every two minutes in order to prompt the ATCO to make a rating.
Criteria	**Review**
Multi-dimensional	No.
Practical and easy to use	ISA has been reported to be easy to use and unobtrusive. Responding to ISA prompts had no significant effects upon ATCO task performance. However, may require the installation of the LED and associated equipment.
Multi-scaled and diagnostic	No: However, ISA, as a real-time measure, may enable the identification of events or tasks that may have contributed or caused the perceived high workload.
Situation/design Sensitivity	ISA is thought to allow the identification of events and or tasks that may have caused or contributed to perceived high levels of workload.
Theoretically sound	Relies upon subjective evaluation of the workload situation by the ATCO. (Must include factors such as experience, training, individual differences, group polarisation, pride effect, etc.)
Consider all critical cognitive activities	No.
Reliability and Validity	ISA has been reported to be reliable and valid within an ATM environment. ISA ratings have been found to correlate with NASA-TLX ratings.
Method and data	ISA ratings of 1 to 5.
Outcome	Not selected for Stage 2 review.

Stage 2 of the review: what, where and why - workload technique review

The table below reviews each of the tools in Stage 2 according to the ‘what’, ‘where’ and ‘why’ criteria and the methodology employed. The outcome column indicates which of the tool's characteristics under the four criteria are relevant to SHAPE: ‘×’ stands for ‘not relevant’, ‘*P’* for ‘partially relevant’ and ‘ ’ for ‘relevant’.

Technique	WHAT	WHERE	WHY	Methodology	Outcome	
Cooper-Harper Rating Scale	Aircraft handling	Interface between the pilot and aircraft	Attributes causality to: ▪ a/c controls, ▪ a/c displays, ▪ a/c stability, ▪ a/c pitch, ▪ a/c speed control, ▪ wind gust.	Decision tree – rating provided of 1 to 10 Reported to be easy to follow	What	×
					Where	*P*
					Why	×
					Methodology	
Bedford Workload Scale	Processing capacity, in terms of spare capacity and general effort	Interface between the human and tasks	Attributes causality to: ▪ subjective perception of workload, ▪ effort, ▪ spare capacity, ▪ attention	Decision tree – rating provided of 1 to 10 Reported to be easy to follow	What	*P*
					Where	*P*
					Why	*P*
					Methodology	
Modified Cooper-Harper Rating Scale	Cognitive functions and mental effort	Interface between the human and systems/tasks	Have been able to attributes causality to: ▪ perception, ▪ monitoring, ▪ evaluation, ▪ communications, ▪ problem solving	Decision tree – rating provided of 1 to 10 Reported to be easy to follow	What	*P*
					Where	*P*
					Why	*P*
					Methodology	

<table>
<tr><th>Technique</th><th>WHAT</th><th>WHERE</th><th>WHY</th><th>Methodology</th><th colspan="2">Outcome</th></tr>
<tr><td rowspan="4"><u>SWAT</u></td><td rowspan="4">Mental workload; to some extent consider factors of:
▪ processing load,
▪ spare capacity.</td><td rowspan="4">Interface between the human and tasks</td><td rowspan="4">Attributes causality to three scales:
▪ time load
▪ mental effort load
▪ psychological stress</td><td rowspan="4">SWAT has two steps:
▪ development of scales and
▪ event scoring using the scales.
Produces a score of 0 to 100</td><td>What</td><td></td></tr>
<tr><td>Where</td><td>P</td></tr>
<tr><td>Why</td><td>✕</td></tr>
<tr><td>Methodology</td><td>P</td></tr>
<tr><td rowspan="4"><u>NASA TLX</u></td><td rowspan="4">▪ Processing load
▪ Effort required
▪ Time</td><td rowspan="4">Human and the task</td><td rowspan="4">Attributes causality to:
▪ mental demand,
▪ physical demand,
▪ temporal demand,
▪ performance,
▪ effort,
▪ frustration level</td><td rowspan="4">Rating technique using six scales</td><td>What</td><td>P</td></tr>
<tr><td>Where</td><td>P</td></tr>
<tr><td>Why</td><td>✕</td></tr>
<tr><td>Methodology</td><td></td></tr>
<tr><td rowspan="4"><u>PUMA</u></td><td rowspan="4">▪ Processing load on cognitive functions and resources
▪ Task difficulty</td><td rowspan="4">▪ Human and ATM tasks
▪ Events and HMI</td><td rowspan="4">Based on MRT:
▪ multiple resource types,
▪ concurrent task performance,
▪ task interference</td><td rowspan="4">Modelling:
▪ OTA and modelling,
▪ video analysis and debriefing</td><td>What</td><td></td></tr>
<tr><td>Where</td><td></td></tr>
<tr><td>Why</td><td></td></tr>
<tr><td>Methodology</td><td>✕</td></tr>
</table>

The conclusions from the Stage 2 review were as follows:

- Although PUMA appeared to satisfy the ‘what’, ‘where’ and ‘why’ criteria, it was considered unsuitable because it is a mental workload modelling tool and very difficult to implement. PUMA is resource intensive and requires substantial training of analyst before it can be used effectively. However, its workload algorithm based on the MRT and the related resource channel classification and conflict matrix were useful materials for the construction of the new tool.

- Although NASA-TLX was unsuitable, its structure was useful, offering a sub-scale and score for each aspect of workload. However, unlike NASA-TLX, the new tool will focus on partitioning mental workload further into cognitive function groups (for instance for. mental workload due to planning tasks, decision-making tasks or multitasking) and mental resources (for instance for mental workload due to visual resource demands or spatial resource demands).

- SWAT (and Zachary/Zachlad Cognitive Analysis) uses successfully specific tasks as the item content in the tool, on which subjects rate the mental workload demands. This concept, although not new, is useful for the construction of the content of the new tool.

- The decision tree rating scale format in Cooper-Harpers, Modified Cooper-Harpers and Bedford Workload Tools was a useful and easy way to guide respondents’ in their subjective ratings of mental workload and hence, worth considering for the new tool.

APPENDIX E: AIM DECISION-TREE RATING SCALES AND THEIR DEVELOPMENT

Following from the review of existing tools, it was also decided that a decision tree format would be used to guide the subjective ratings for effort and difficulty. Hence, the rating scale for mental effort and task difficulty will be embedded in the decision tree.

As a result of one of the meetings held as part of the consultation process, the decision was made to implement a seven-point interval rating scale (of 1-7) for both mental effort and task difficulty. The advantages are (i) greater choice of ratings and (ii) larger range of sensitivity to mental workload differences. The seven-point scales are incremental with point 1 being the lowest and point 7 being the highest. In addition, only alternate rating points are specifically defined for the subjects. The definitions of the rating scales can be found in Table 7. This was called the 'absolute rating scales'.

Table 7: Definitions of the absolute rating scales for mental effort and task difficulty

Rating point	Definition on task difficulty scale	Definition on mental effort scale
1	Easier than point 2	Less than point 2
2	'Easy'	'Little mental effort'
3	Harder than point 2 but easier than point 4	More effort than point 2 but less than point 4
4	'Neither easy nor difficult'	'Moderate mental effort'
5	Harder than point four but not as hard as point six	More effort than point four but less than point six
6	'Difficult'	'Large amount of mental effort'
7	Harder than point 6	More effort than point six

This allowed the words used to define the points to be kept to English words which are simple. In terms of word characteristics these are high-frequency and high-familiarity English words, which will be easy to understand for non-native English speakers. This will lower the risk of misunderstanding from translation to other European languages from English by non-native English speakers.

However, a rating scale of 1-7 implies that the mental workload ratings made will be an absolute rating of the controller's experience of workload. This will necessitate a baseline evaluation of mental workload, before changes in mental workload can be assessed. Not all simulations or research trials will have the luxury of time and resources to implement baseline measurements. Hence, another version of the seven-point interval ratings scale for mental effort and task difficulty was produced. The alternative version will enable subjects to make mental workload ratings of their workload experience in the simulation or trial relative to what they currently experience. That is controllers will rate how much more or less the mental effort was required in the simulation/trial compared to the current operational situation. Similarly, controllers will rate how much more or less difficult task performance was in the simulation/trial compared to the current operational situation. The points on the alternative rating scales range from -3 to 3. This was called the 'relative rating scales'. The definitions of the rating scales can be found in Table 8.

Table 8: Definitions of the relative rating scales for mental effort and task difficulty

Rating point	Definition on task difficulty scale	Definition on mental effort scale
- 3	Much easier	Large decrease in mental effort
- 2	Easier	Moderate decrease in mental effort
- 1	Slightly easier	Slight decrease in mental effort
0	No Change in difficulty	No change in mental effort
1	Slightly more difficult	Minimal increase in mental effort
2	More difficult	Moderate increase in mental effort
3	Much more difficult	Large increase in mental effort

The purpose of using the decision tree format was to guide the subjects through their ratings on mental effort and task difficulty on each tool item. Four simple questions were constructed for the absolute rating scales and five similar questions were constructed for the relative rating scales. Figure 3 displays the 'absolute workload decision tree' and the decision tree questions for the absolute rating scales and the rating scales embedded in the decision tree.

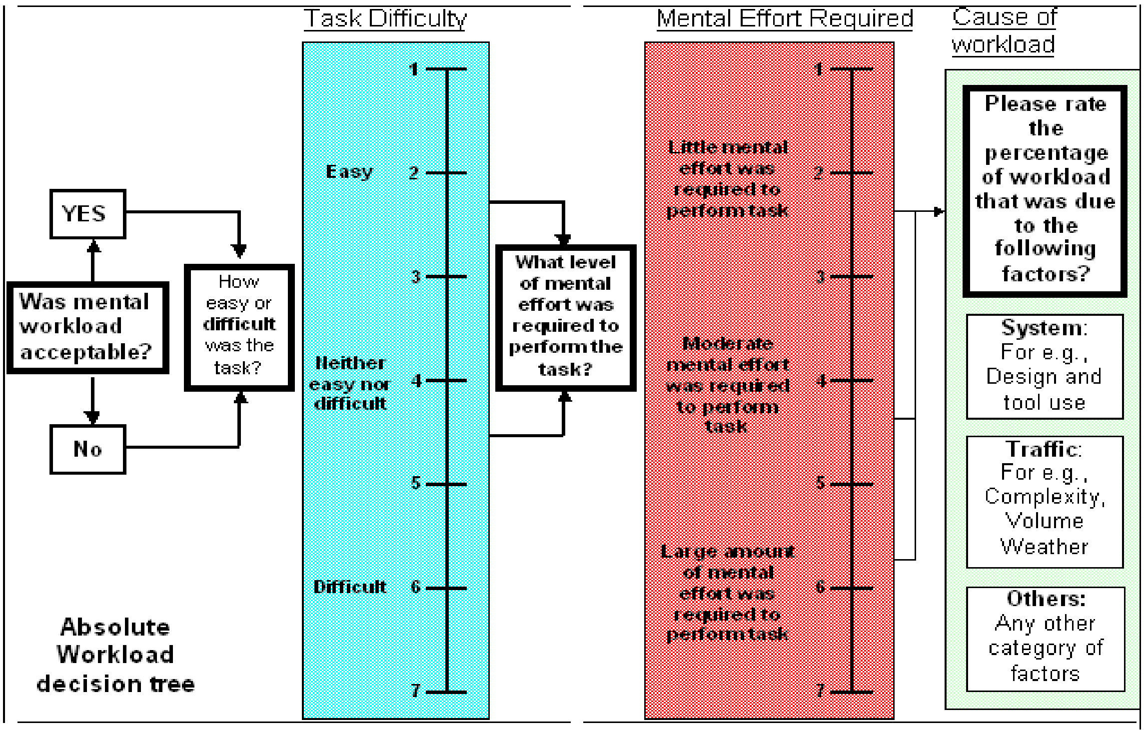


Figure 3: Absolute workload decision tree

The questions for the absolute rating scales were as follows:

1. 'Was mental workload acceptable?'. The response set available is binomial, 'Yes' or 'No'.

2. 'How easy or difficult was the task?'. The response set available is the absolute rating scale for task difficulty.

3. 'What level of mental effort was required to perform the task?'. The response set available is the absolute rating scale for mental effort.

4. 'Please rate the percentage of workload that was due to the following factors.' The response set included three factors: system, traffic and others. Subjects are required to decide what percentage of the workload experience can be attributed to each factor. This question was meant to elicit the source of the mental workload, in order to examine if the workload impact or differences are attributable more to the traffic situation and complexity, the system design or other unforeseen factors.

Figure 4 displays the 'relative workload decision tree' and the decision tree questions for the absolute rating scales and the rating scales embedded in the decision tree.

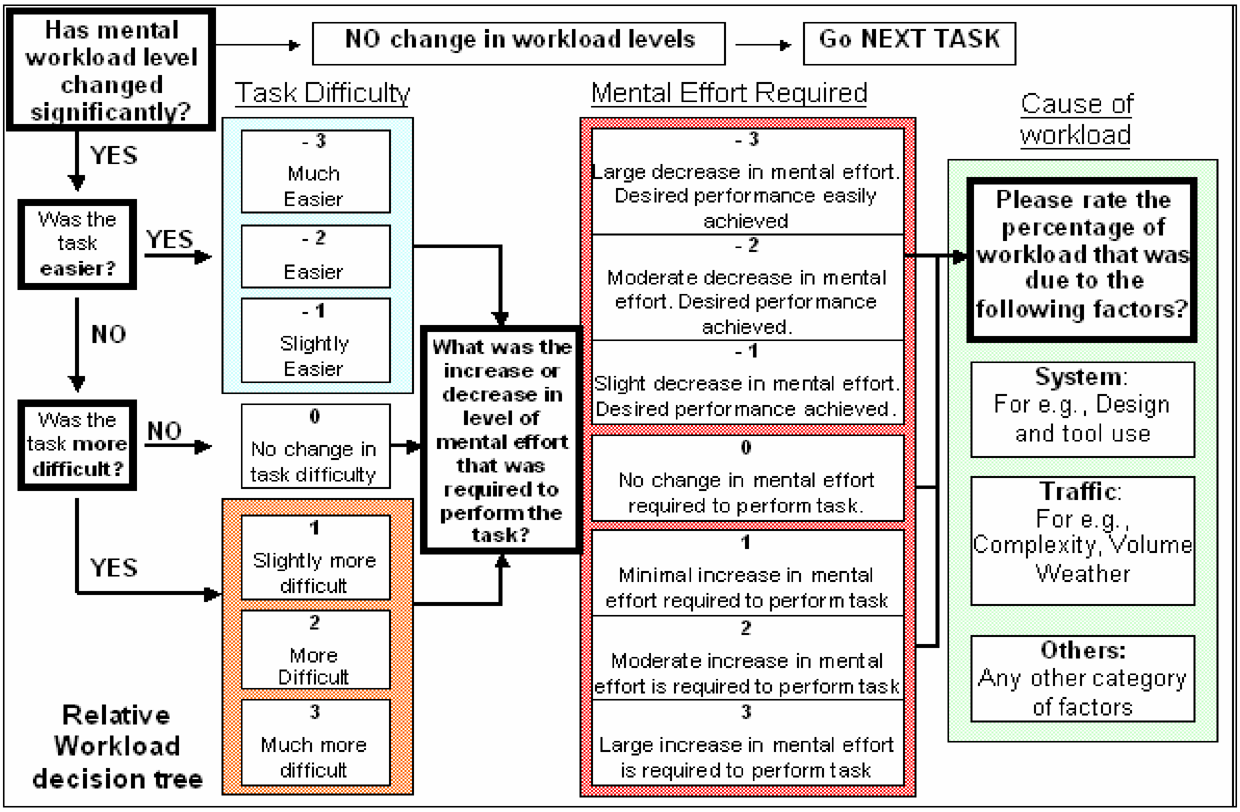


Figure 4: Relative workload decision tree

The questions for the relative rating scales were:

1. 'Has mental workload level changed significantly'. The response set available is binomial, 'No' (and subjects are instructed to go to the next item or 'Yes' (and subjects are led to Question 2).

2. 'Was the task easier?'. The response set available is binomial, 'Yes' (and subjects are confined to only points '-3', '-2' and '-1' of the relative rating scale for task difficulty and led to Question 4 after making their rating) or 'No' (and subjects are led to Question 3).

3. 'Was the task more difficult?' The response set available is binomial, 'Yes' (and subjects are confined to only points '3', '2' and '1' of the relative rating scale for task difficulty) or 'No' (and subjects confined to only point '0'). After making their decision, subjects are led to Question 4.

4. 'What was the increase or decrease in level of mental effort that was required to perform the task?'. The response set available is the relative rating scale for mental effort.

5. 'Please rate the percentage of workload that was due to the following factors.' The response set included three factors: system, traffic and others. Subjects are required to decide what percentage of the workload experience can be attributed to each factor.

APPENDIX F: CONSTRUCTION OF THE ITEMS FOR THE CONTENT OF AIM

Most of the existing subjective workload assessment tools required either a task specific workload rating or a non-task specific workload rating. The former require the subject to give a workload rating on each pre-defined task performance or specified tasks/operations (e.g. SWAT, Cooper-Harper rating scale). The latter require the subject to make a global workload rating for a work period, without defining what the task performance being assessed is or without specifying any particular task/operation (e.g. NASA TLX, Crew Status Survey). These non-task specific workload tools require subjects to make a global rating on one or more dimensions of workload, such as time pressure, frustration, physical demand and difficulty, mental effort/demand).

Non-task specific workload ratings will not provide the diagnosticity set by the requirement specification. They also do not provide much diagnosticity towards automation system design changes. In addition, they do not satisfy the criteria, which specify that mental workload need to be diagnosed according to which cognitive function groups are contributing to the mental workload, nor does it allow diagnosis of mental workload according to mental resources.

In order to satisfy these criteria, the item content of AIM needs to be task-specific and mental workload ratings need to be made on defined cognitive function groups or specified tasks (e.g. SWAT). A taxonomy of cognitive functions is needed. As part of the construction of the SHAPE Framework developed within the SHAPE work package on skill set requirements, a catalogue of cognitive functions and associated sub-functions was complied and validated by controllers (see EATM, 2004c). In validation exercises conducted, the controllers were also required to select the critical cognitive sub-functions for each operational function.

The selected critical cognitive sub-functions were taken as the task items for AIM. The list of 46 task items for the content of AIM can be found in Appendix G. Nine cognitive function groups were then derived based on the cognitive model used in the SHAPE framework (see EATCHIP, 1997). These were:

1. Multitasking
2. Direct attention to information sources
3. Take account of and process external information
4. Memory management
5. Build and maintain SA
6. Planning
7. Decision-making
8. Diagnosing and problem solving
9. Team awareness

The original SHAPE list of 21 cognitive functions were then classified into these nine groups. For example, the cognitive function group 'Diagnosing and problem solving' contained functions such as 'Diagnose perceived problem', 'Active problem solving' and 'Diagnose novel situations/problems'.

It was then possible to trace to which cognitive function group each of the 46 task items in the list belonged by using the original SHAPE cognitive function and sub-function list. For example, if Task A is a sub-function in 'Active Problem solving', then it belonged to the 'Diagnosing and problem solving' cognitive function group. The different cognitive function groups in the taxonomy and the associated task items made up the nine sub-scales of AIM, allowing the assessment of mental workload due to specific cognitive functioning. The key for matching the task item to the cognitive function groups can be found in Appendix H.

A second taxonomy for the AIM task items was constructed for assessing mental workload according to the different demand on mental resource types. The development of this taxonomy uses the MRT and the workload algorithm (based on MRT) used in PUMA. Each task item was then coded according to the types of mental resource it uses. The relationship between task item to mental resource type is not a one-to-one mapping. Each task item may belong to more than one mental resource type. The mental resource types are[6]:

1. Encoding
2. Central processing
3. Response
4. Verbal
5. Spatial
6. Visual
7. Auditory
8. Motor

The coding was carried out by two HF specialists, using existing PUMA task models[7] and validation exercise data[8] from work package on SHAPE Framework (see EATM, 2004c). The key for matching the task item to the mental resource types can be found in Appendix I. The type of mental resources taxonomy allowed for the assessment of mental workload to be diagnosed according the demand of the different types of mental resources.

The third and last taxonomy for the AIM task items was constructed for diagnosing the mental workload according to the degree of resource

[6] The types of mental resources are defined by the MRT. This is described in Appendix B.

[7] In PUMA, as part of the MWL modelling tool set, a database exists of common basic ATC tasks and the mental resource demand rating for each task.

[8] In SHAPE work package on skill set requirements several validation exercises were carried out with HF specialists. In one of the exercises several specialists were requested to select the cognitive processes required to carry out each of the cognitive sub-functions. This data provided more information about the mental resource type required for each task item. For example, if the item was the search and detection processes that belonged to the input and output stage of the cognitive model (i.e. ITA), the task item would be in the 'encoding resource' category.

competition (or resource conflict). The PUMA conflict matrix[9] contains interference coefficients for all possible conflicts between each type of mental resource (see Table 9). The coefficient values range from 0 to 1 and represent the extent of interference between the mental resources that are required for the task. Having identified the mental resources required for each task item in the second taxonomy, the development of this taxonomy uses resource conflict matrix in PUMA to rate each task item according to the degree of resource competition. The key for matching the task item to the mental resource types can be found in Appendix J.

The degree of resource competition taxonomy allowed for the assessment of mental workload to be diagnosed according the impact on spare processing capacity. The assumption is that the greater the mental workload demand, the greater the impact is on spare processing capacity. However, the spare processing capacity is further affected and reduced if there is a high degree of resource competition between the processing load placed on the mental resources.

Table 9: An example of a mental resource conflict matrix in PUMA

	Visual spatial monitor	**Visual spatial search**	**Visual verbal encoding**	**Auditory verbal encoding**	**Verbal processing**	**Spatial processing**	**Manual response**	**Verbal response**
Visual spatial monitor	Medium							
Visual spatial search	High	Low						
Visual verbal encoding	Low	Low	Low					
Auditory verbal encoding	Medium	Medium	Medium	High				
Verbal processing	Low	Medium	Low	High	High			
Spatial processing	Low	Medium	Low	Medium	High	High		
Manual response	Low	Low	Low	Medium	High	Medium	High	
Verbal response	Low	Medium	Low	High	High	High	High	High

9 The coefficients in the conflict matrix were constructed and determined empirically by Sowerby Research Centre, Bae, and are treated as confidential information. Permission was not granted to reproduce the conflict matrix showing the coefficients. The coefficients are available in licensed copies of PUMA. Instead, the cells in the matrix contain indications of whether the coefficient was high, medium or low.

Three versions of AIM were produced:

1. AIM-Q: Includes all 46 task items to be rated by the subject using the decision tree rating scale (either the absolute or relative version), enabling the assessment of mental workload due to all nine cognitive function groups.

2. AIM-Cog: Consists of nine sub-tests, one for each cognitive function group allowing the assessment of mental workload due to a specific targeted cognitive function group. For example, if a small-scale trial was carried out on a new prototype of a system change focused on assisting the controllers in decision-making, for ease of administration and minimal resource required, the decision-making sub-test can be administered individually. Each sub-test contains between four and eleven items.

3. AIM-Hi: This version is an abbreviated test with only nine items. The purpose is to obtain a high-level and global assessment of the mental workload due to the different cognitive function groups. However, it will not allow any further diagnosis into mental workload due to mental resource types or specific task performance. Each cognitive function group is a task item to be rated by the subject on mental effort required and difficulty. A simple but detailed explanation of each item is provided to the subject. This version is slightly resembles NASA-TLX which provides subjects with descriptions of different dimensions of workload (such as frustration level, temporal demand, physical demand, etc) and requires subjects to make a global rating on each dimension. The difference is that AIM-Hi requires global ratings on the nine cognitive function types.

APPENDIX G: TASK ITEMS FOR AIM

1. Prioritise tasks
2. Identify potential conflicts
3. Scan information displays
4. Apply previous experience
5. Share information / communicate with team members
6. Gather and interpret information
7. Scan Flight Progress Strip (FPS)
8. Divide attention (e.g. speaking and writing at the same time)
9. Choose solution
10. Evaluate options against traffic situation/conditions
11. Anticipate future traffic situation
12. Integrate information
13. Use mental or physical cues (e.g. cues, cocking strips, notes or mental tags) to remind oneself of actions required
14. Evaluate the consequences of the plan
15. Listen for relevant information
16. Manage and regulate workload
17. Ask for information
18. Perform actions before a/c arrives in sector or into area of responsibility
19. Prioritise and update currently useful and relevant knowledge in working memory
20. Extract relevant data for traffic assessment visual displays (level, time, route, speed)
21. Recognise conflict
22. Recognise the need to request assistance before workload exceeds capacity
23. Resolve conflict
24. Retrieve information from long-term memory

25. Scan radar or any ATC Display Screen Equipment (DSE) or Flight Data Processing System (FDPS) equipment
26. Check against traffic the feasibility and relevance of the request
27. Check external information and gather evidence
28. Update weather information
29. Formulate appropriate action or response
30. Formulate decision options
31. Anticipate team member’s needs/capability
32. Check information sources
33. Identify tasks which are highly similar (e.g. same instruction that needs issuing to several pilots)
34. Check order and priority of actions in plans
35. Develop new plan for the novel situation/problem
36. Evaluate importance of tasks
37. Monitor own capacity to cope with actual workload
38. Gather/interpret proactive information for team members
39. Recall and identify existing knowledge (rules, information) for an analogous situation
40. Assess impact on own and/or team’s workload and prioritise request
41. Scan reminders
42. Tidy up strip display and put in place mental reminders for next controller
43. Share time between tasks
44. Update ATC knowledge and assimilate into existing knowledge
45. Update team information
46. Verify information source/trust

APPENDIX H: KEY FOR THE COGNITIVE FUNCTION GROUPS

The numbers in the key refer to the item in Appendix G.

Multitasking

1
6
8
32
33
36
43

Direct attention to information sources and monitoring

3
6
7
15
17
25
32
41

Take account of, and process, external information

6
12
20
32

Memory management

12
13
19
24
28
34
44
45

Build and maintain SA

6
10
11
12
46

Planning

2
6
11
14
18
29

Decision-making

9
10
21
23
26
29
30
31
40

Diagnosing and problem solving

4
6
9
11
12
14
27
29
35
39
46

Team awareness

5
16
22
31
37
38
40
42
45
46

Page intentionally left blank

APPENDIX I: KEY FOR THE TYPES OF MENTAL RESOURCE

The numbers in the key refer to the item in Appendix G.

Encoding	Central processing	Response	Verbal	Spatial	Visual	Auditory	Motor
2	1	5	2	1	2	5	18
3	4	8	3	2	3	6	28
6	5	17	4	3	5	8	42
7	6	18	5	4	6	12	
8	8	28	6	6	7	15	
10	9	42	7	8	8	17	
11	10		8	9	10	18	
13	11		9	10	11	26	
15	12		12	11	12	27	
20	13		13	12	13	28	
21	14		14	13	20	31	
22	16		15	14	21	32	
25	18		16	16	22	33	
26	19		17	18	25	38	
27	20		18	19	26	45	
31	21		19	20	27	46	
32	22		20	21	31		
33	23		24	22	32		
37	24		27	23	33		
38	26		28	25	35		
40	27		31	26	36		
41	28		32	27	37		
45	29		33	29	38		
46	30		38	30	40		
	31		39	31	41		
	33		41	32	42		
	34		42	33	45		
	35		44	34	46		
	36		45	35			
	37		46	36			
	38			37			
	39			38			
	40			39			
	42			40			
	43			41			
	44			42			
	45			43			
	46			44			
				45			
				46			

Page intentionally left blank

APPENDIX J: KEY FOR THE DEGREE OF RESOURCE COMPETITION

The numbers in the key refer to the item in Appendix G.

High	Medium	Low
1	6	2
4	8	3
5	10	7
9	11	13
12	21	15
14	22	17
16	26	20
18	27	24
19	31	41
23	32	
25	33	
28	37	
29	38	
30	40	
34	42	
35	45	
36	46	
39		
43		
44		

Page intentionally left blank

APPENDIX K: AIM-Q RECORDING FORMS

For absolute workload decision tree

AIM Q

Participant number: ____________________ **Date:** ________ **Time:** ________ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
1. Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
2. Identify potenial conflicts	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
3. Scan information displays	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
4. Apply previous experience	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
5. Share information / Communicate with team members	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
6. Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
7. Scan FPS	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
8. Divide attention (e.g. Speaking and writing at the same time)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
9. Choose solution	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
10. Evaluate options against traffic situation / conditions	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
11. Anticipate future traffic situation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
12. Integrate information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Page 1 of 4

AIM Q

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
13. Use mental or physical cues (E.g. cues, cocking strips, notes or mental tags) to remind oneself of actions required	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
14. Evaluate the consequences of the plan	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
15. Listen for relevant information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
16. Manage and regulate workload	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
17. Ask for information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
18. Perform actions before a/c arrives in sector or into area of responsibility	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
19. Prioritise and Update currently useful and relevant knowledge in working memory	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
20. Extract relevant data for traffic assessment visual displays (level, time, route, speed)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
21. Recognise conflict	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
22. Recognise the need to request asistance before workload exceeds capacity	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
23. Resolve conflict	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
24. Retrieve information from Long Term Memory	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
25. Scan Radar or any ATC DSE (Display Screen Equipment) or FDPS equipment	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Page 2 of 4

AIM Q

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
26. Check against traffic the feasibility and relevance of the request	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
27. Check external information and gather evidence	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
28. Update weather information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
29. Formulate appropriate action or response	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
30. Formulate decision options	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
31. Anticipate team member`s needs / capability	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
32. Check information sources	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
33. Identify tasks which are highly similar (e.g. same instruction that needs issuing to several pilots)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
34. Check order and priority of actions in plans	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
35. Develop new plan for the novel situation / problem	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
36. Evaluate importance of tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
37. Monitor own capacity to cope with actual workload	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
38. Proactive information gathering/interpretation for team members	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Page 3 of 4

AIM Q

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
39. Recall and identify existing knowledge (rules, information) for an analogous situation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
40. Assess impact on own and/or team`s workload and prioritise request	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
41. Scan reminders	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
42. Tidy up strip display and put in place mental reminders for next controller	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
43. Time share between tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
44. Update ATC knowledge and assimilate into exisiting knowledge	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
45. Update team information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
46. Verify Information source/trust	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Page 4 of 4

For relative workload decision tree

AIM Q

Participant number: ____________________ **Date:** ________ **Time:** ________ **Sector:** ____

Weather Conditions: ____________________

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
1. Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
2. Identify potenial conflicts	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
3. Scan information displays	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
4. Apply previous experience	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
5. Share information / Communicate with team members	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
6. Information gathering and interpretation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
7. Scan FPS	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
8. Divide attention (e.g. Speaking and writing at the same time)	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
9. Choose solution	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
10. Evaluate options against traffic situation / conditions	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
11. Anticipate future traffic situation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
12. Integrate information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Page 1 of 4

AIM Q

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
13. Use mental or physical cues (E.g. cues, cocking strips, notes or mental tags) to remind oneself of actions required	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
14. Evaluate the consequences of the plan	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
15. Listen for relevant information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
16. Manage and regulate workload	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
17. Ask for information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
18. Perform actions before a/c arrives in sector or into area of responsibility	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
19. Prioritise and Update currently useful and relevant knowledge in working memory	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
20. Extract relevant data for traffic assessment visual displays (level, time, route, speed)	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
21. Recognise conflict	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
22. Recognise the need to request asistance before workload exceeds capacity	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
23. Resolve conflict	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
24. Retrieve information from Long Term Memory	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
25. Scan Radar or any ATC DSE (Display Screen Equipment) or FDPS equipment	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Page 2 of 4

AIM Q

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
26. Check against traffic the feasibility and relevance of the request	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
27. Check external information and gather evidence	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
28. Update weather information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
29. Formulate appropriate action or response	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
30. Formulate decision options	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
31. Anticipate team member`s needs / capability	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
32. Check information sources	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
33. Identify tasks which are highly similar (e.g. same instruction that needs issuing to several pilots)	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
34. Check order and priority of actions in plans	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
35. Develop new plan for the novel situation / problem	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
36. Evaluate importance of tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
37. Monitor own capacity to cope with actual workload	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
38. Proactive information gathering/interpretation for team members	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Page 3 of 4

AIM Q

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
39. Recall and identify existing knowledge (rules, information) for an analogous situation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
40. Assess impact on own and/or team`s workload and prioritise request	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
41. Scan reminders	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
42. Tidy up strip display and put in place mental reminders for next controller	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
43. Time share between tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
44. Update ATC knowledge and assimilate into exisiting knowledge	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
45. Update team information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
46. Verify Information source/trust	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Page 4 of 4

APPENDIX L: AIM-COG RECORDING FORMS

For absolute workload decision tree

AIM Cognitive sub-scales

Participant number: ________________ **Date:** _______ **Time:** _______ **Sector:** ______
Weather Conditions: ________________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Multitasking					
Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Divide attention (e.g. Speaking and writing at the same time)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Check information sources	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Identify tasks which are highly similar (e.g. same instruction that needs issuing to several pilots)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Evaluate importance of tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Time share between tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

AIM Cognitive sub-scales

Participant number: ________________ **Date:** _______ **Time:** _______ **Sector:** ______
Weather Conditions: ________________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Direct attention to information sources and monitoring					
Scan information displays	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Scan FPS	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Listen for relevant information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Ask for information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Scan reminders	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Scan Radar or any ATC DSE (Display Screen Equipment) or FDPS equipment	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Check information sources	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

AIM Cognitive sub-scales

Participant number: ____________ **Date:** ________ **Time:** ________ **Sector:** ______
Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Take account of and process external information					
Integrate information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Extract relevant data for traffic assessment visual displays (level, time, route, speed)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Check information sources	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

AIM Cognitive sub-scales

Participant number: ____________ **Date:** ________ **Time:** ________ **Sector:** ______
Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Memory Management					
Prioritise and Update currently useful and relevant knowledge in working memory	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Retrieve information from Long Term Memory	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Update weather information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Check order and priority of actions in plans	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Update team information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Integrate information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Use mental or physical cues (E.g. cues, cocking strips, notes or mental tags) to remind oneself of actions required	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Update ATC knowledge and assimilate into exisiting knowledge	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

AIM Cognitive sub-scales

Participant number: ____________ **Date:** ________ **Time:** ________ **Sector:** ______
Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Build and maintain situation awareness					
Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Anticipate future traffic situation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Integrate information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Verify Information source/trust	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Evaluate options against traffic situation / conditions	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

AIM Cognitive sub-scales

Participant number: ________________ **Date:** ______ **Time:** ______ **Sector:** ______

Weather Conditions: ____________________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.
Planning				
Identify potenial conflicts	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Information gathering and interpretation	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Anticipate future traffic situation	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Perform actions before a/c arrives in sector or into area of responsibility	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Evaluate the consequences of the plan	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Formulate appropriate action or response	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %

AIM Cognitive sub-scales

Participant number: ________________ **Date:** ______ **Time:** ______ **Sector:** ______

Weather Conditions: ____________________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.
Decision Making				
Recognise conflict	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Evaluate options against traffic situation / conditions	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Check against traffic the feasibility and relevance of the request	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Anticipate team member`s needs / capability	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Assess impact on own and/or team`s workload and prioritise request	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Formulate appropriate action or response	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Formulate decision options	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Choose solution	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Resolve conflict	Y / N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %

AIM Cognitive sub-scales

Participant number: ________________ **Date:** _______ **Time:** _______ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Diagnosing and Problem solving					
Apply previous experience	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Evaluate the consequences of the plan	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Formulate appropriate action or response	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Develop new plan for the novel situation / problem	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Anticipate future traffic situation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Integrate information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Choose solution	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Verify Information source/trust	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Recall and identify existing knowledge (rules, information) for an analogous situation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Check external information and gather evidence	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

AIM Cognitive sub-scales

Participant number: ________________ **Date:** _______ **Time:** _______ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Team awareness					
Manage and regulate workload	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Recognise the need to request asistance before workload exceeds capacity	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Monitor own capacity to cope with actual workload	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Assess impact on own and/or team`s workload and prioritise request	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Proactive information gathering/interpretation for team members	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Share information / Communicate with team members	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Anticipate team member`s needs / capability	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Update team information	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Verify Information source/trust	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	
Tidy up strip display and put in place mental reminders for next controller	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

For relative workload decision tree

AIM Cognitive Sub-scales

Participant number: ________________ **Date:** _______ **Time:** _______ **Sector:** _______

Weather Conditions: ____________________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Multitasking					
Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Information gathering and interpretation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Divide attention (e.g. Speaking and writing at the same time)	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Check information sources	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Identify tasks which are highly similar (e.g. same instruction that needs issuing to several pilots)	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Evaluate importance of tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Time share between tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	

AIM Cognitive Sub-scales

Participant number: ________________ **Date:** _______ **Time:** _______ **Sector:** _______

Weather Conditions: ____________________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Direct attention to information sources and monitoring					
Scan information displays	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Scan FPS	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Listen for relevant information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Ask for information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Scan reminders	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Scan Radar or any ATC DSE (Display Screen Equipment) or FDPS equipment	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Check information sources	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	
Information gathering and interpretation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %	

AIM Cognitive Sub-scales

Participant number: ____________ **Date:** ______ **Time:** ______ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Take account of and process external information				
Integrate information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Extract relevant data for traffic assessment visual displays (level, time, route, speed)	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Check information sources	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Information gathering and interpretation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

AIM Cognitive Sub-scales

Participant number: ____________ **Date:** ______ **Time:** ______ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Memory Management				
Prioritise and Update currently useful and relevant knowledge in working memory	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Retrieve information from Long Term Memory	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Update weather information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Check order and priority of actions in plans	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Update team information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Integrate information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Use mental or physical cues (E.g. cues, cocking strips, notes or mental tags) to remind oneself of actions required	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Update ATC knowledge and assimilate into exisiting knowledge	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

AIM Cognitive Sub-scales

Participant number: ____________ **Date:** ______ **Time:** ______ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Build and maintain situation awareness				
Information gathering and interpretation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Anticipate future traffic situation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Integrate information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Verify Information source/trust	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Evaluate options against traffic situation / conditions	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

AIM Cognitive Sub-scales

Participant number: __________________ **Date:** _______ **Time:** _______ **Sector:** ________

Weather Conditions: ____________________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.			Please rate the increase or decrease in mental effort required to perform the			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Planning									
Identify potenial conflicts	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Information gathering and interpretation	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Anticipate future traffic situation	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Perform actions before a/c arrives in sector or into area of responsibility	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Evaluate the consequences of the plan	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Formulate appropriate action or response	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %

AIM Cognitive Sub-scales

Participant number: __________________ **Date:** _______ **Time:** _______ **Sector:** ________

Weather Conditions: ____________________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.			Please rate the increase or decrease in mental effort required to perform the			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Decision Making									
Recognise conflict	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Evaluate options against traffic situation / conditions	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Check against traffic the feasibility and relevance of the request	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Anticipate team member`s needs / capability	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Assess impact on own and/or team`s workload and prioritise request	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Formulate appropriate action or response	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Formulate decision options	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Choose solution	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %
Resolve conflict	Y N	-1 0 1	-2 2	-3 3	-1 0 1	-2 2	-3 3	System Traffic Others	% % %

AIM Cognitive Sub-scales

Participant number: ____________ **Date:** ________ **Time:** ________ **Sector:** ________

Weather Conditions: ____________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Diagnosing and Problem solving				
Apply previous experience	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Evaluate the consequences of the plan	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Formulate appropriate action or response	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Develop new plan for the novel situation / problem	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Information gathering and interpretation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Anticipate future traffic situation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Integrate information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Choose solution	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Verify Information source/trust	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Recall and identify existing knowledge (rules, information) for an analogous situation	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Check external information and gather evidence	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

AIM Cognitive Sub-scales

Participant number: ____________ **Date:** ________ **Time:** ________ **Sector:** ________

Weather Conditions: ____________

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Team awareness				
Manage and regulate workload	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Recognise the need to request asistance before workload exceeds capacity	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Monitor own capacity to cope with actual workload	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Assess impact on own and/or team`s workload and prioritise request	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Proactive information gathering/interpretation for team members	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Share information / Communicate with team members	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Anticipate team member`s needs / capability	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Update team information	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Verify Information source/trust	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Tidy up strip display and put in place mental reminders for next controller	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

APPENDIX M: AIM-HI RECORDING FORMS

For absolute workload decision tree

AIM-Hi

Participant number: ____________ **Date:** ______ **Time:** ______ **Sector:** ______

Weather Conditions: ____________

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Direct attention to information sources and monitoring This involves scanning information displays, radar or strip displays to gather information, listenting out for relevant information, checking information and monitoring the radar regularly	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Take account of and process information This involves interpreting information gathered and integrating the information. It also involves extracting the relevant data from various visual displays and auditory information.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Memory Management This involves prioritising and updating useful and relevant knowledge in working memory, checking the order and priority of actions in plans, retrieving relevant information from long term memory and using mental or physical cues to remind oneself of actions required.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Build and maintain situation awareness This involves, building a picture and an understanding of the current traffic situation, anticipating the future traffic situation, and regularly maintaining and updating awareness of the traffic sitaution. Also, involves verifying information about traffic situation.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %

AIM-Hi

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Planning Generate a position plan, formulate appropriate action or response and perform actions before a/c arrives into area of responsibility. Also, reviewing, checking the credibility and evaluating the consequences of the plan, adapt plan or generate new a plan, if necessary.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Decision Making Make decisions for control actions, to solve conflicts and deal with pilots' request. This involves formulating decision options, evaluating options against traffic situation, choosing solution and formulating appropriate action/response.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Diagnosing and Problem solving Identify the problem, find an explanation and generate a solution. This also involves checking information and gathering evidence to verify the explanation, apply previous experience to solve the problem or develop a new plan to solve a novel situation/problem.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Team awareness Manage and regulate workload, recognise the need to request or offer asistance before workload exceeds capacity. Also, involves assessing impact on team's workload, prioritise tasks, being proactive in helping team members, sharing information and communicate effectively.	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %

For relative workload decision tree

AIM - Hi

Participant number: ____________________ **Date:** ______ **Time:** ______ **Sector:** _______

Weather Conditions: ______________________

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Direct attention to information sources and monitoring This involves scanning information displays, radar or strip displays to gather information, listening out for relevant information, checking information and monitoring the radar regularly	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Take account of and process information This involves interpreting information gathered and integrating the information. It also involves extracting the relevant data from various visual displays and auditory information.	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Memory Management This involves prioritising and updating useful and relevant knowledge in working memory, checking the order and priority of actions in plans, retrieving relevant information from long term memory and using mental or physical cues to remind oneself of actions required.	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Build and maintain situation awareness This involves, building a picture and an understanding of the current traffic situation, anticipating the future traffic situation, and regularly maintaining and updating awareness of the traffic sitaution. Also, involves verifying information about traffic situation.	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

AIM - Hi

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Planning Generate a position plan, formulate appropriate action or response and perform actions before a/c arrives into area of responsibility. Also, reviewing, checking the credibility and evaluating the consequences of the plan, adapt plan or generate new a plan, if necessary.	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Decision Making Make decisions for control actions, to solve conflicts and deal with pilots' request. This involves formulating decision options, evaluating options against traffic situation, choosing solution and formulating appropriate action/response.	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Diagnosing and Problem solving Identify the problem, find an explanation and generate a solution. This also involves checking information and gathering evidence to verify the explanation, apply previous experience to solve the problem or develop a new plan to solve a novel situation/problem .	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Team awareness Manage and regulate workload, recognise the need to request or offer asistance before workload exceeds capacity. Also, involves assessing impact on team's workload, prioritise tasks, being proactive in helping team members, sharing information and communicate effectively.	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

APPENDIX N: INSTRUCTION SHEETS FOR SUBJECTS

For the administration of AIM-Q using the absolute workload decision tree

AIM Q Instructions (Absolute rating scale)

What you need

- Absolute Workload Decision Tree (the sheet covered in clear plastic)
- AIM-Q Recording Form (spreadsheet)
- This instruction sheet

Introduction

This exercise has been designed to find out your views on the workload placed on you by the system that you have just used.

It is important that you read all of the instructions before you complete this exercise.

1. Please look at the AIM-Q Recording Form. It looks like this...

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors
1. Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
2. Identify potenial conflicts	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
3. Scan information displays	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %

The Recording Form has two purposes.

- The first purpose is to provide you with a list of items we wish you to evaluate using the decision tree. These items are shown in the first column of the Recording Form (highlighted by the red box).

- The second purpose is to provide you with an area to record the decisions that you have made whilst using the decision tree (highlighted by the blue box). If the rating options in the blue box are not the same as the rating options on the recording form that you have been given, please notify one of the trial/simulation staff.

2. Please look at the Absolute Workload Decision Tree. (It is covered in clear plastic)

The Absolute Workload Decision Tree is designed to lead you through four questions relating to workload. The questions are highlighted by the use of this box

Your task is to use the Decision Tree to evaluate all of the task items listed in the first column of the Recording Form.

The Decision tree works from left to right.

Each question in the decision tree has a number of possible answers. These answers are shown in the coloured boxes (blue, red and green). For example, the blue box in the decision tree provides you with a list of answers to the questions 'How easy or difficult was the task?

To answer a question please select the most suitable answer from the list of possible answers shown in the coloured box that is next to the question.

When you have answered a question please record your answer in the Recording Form and then follow the arrows on the decision tree to the next question.

It is vital that ALL of your answers are recorded in the Recording Form.

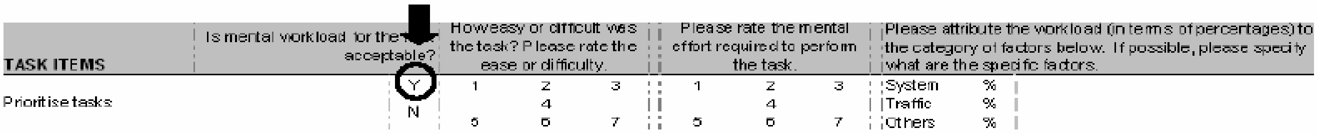
The next page shows an example of how to evaluate an item using the decision tree and how to record your answers on the Recording Form.

Assessment of the Impact on Mental workload (AIM)

Example of how to evaluate an item using the Absolute Workload Decision Tree

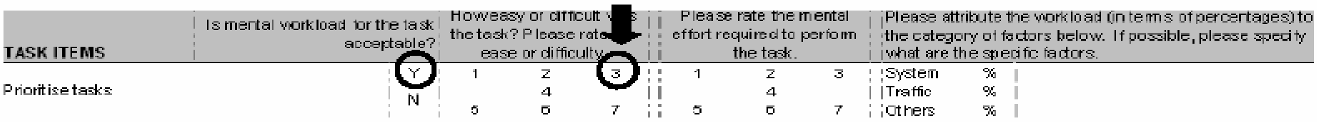
1. Select the first task from the AIM-Q Recording Form.
2. Work through the Absolute Workload Decision Tree from left to right answering each question for the selected task.

Question 1. Is the workload needed to perform this task acceptable? The possible answers are 'YES' and 'NO'. To record your answers please circle either Y or N on the Recording Form.



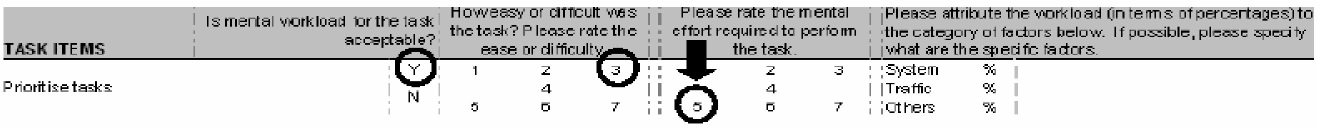
Question 2. How easy or difficult was the task? Please rate the difficulty of the task on a 7-point scale, shown in the blue box. The rating scale is incremental where 1 is the lowest rating and 7 the highest rating. 3 points on the scale (points 2, 4 and 6) are labelled with keywords. The other 4 points on the scale have been left blank but are defined as less than or more than the labelled point. To rate your answer look at the 3 keywords and decide which one is closest to your rating of 'Task Difficulty'. If your rating is identical to the keyword then please record the number in the Recording Form. If however, you believe that your rating of task difficulty is not suitably reflected by the keyword then consider the blank points around the keyword.

For example, if you believe that 'Prioritising tasks' was on the "Easy" side of the scale but not just 'Easy', you would select point 3 as your answer. Or, if you believe that 'Prioritising tasks' was very difficult to perform you would select point 7 as your answer. If you believed that the task was neither easy nor difficult, you would select 4 as your answer. Once you have selected your answer please enter it into the Recording Form.



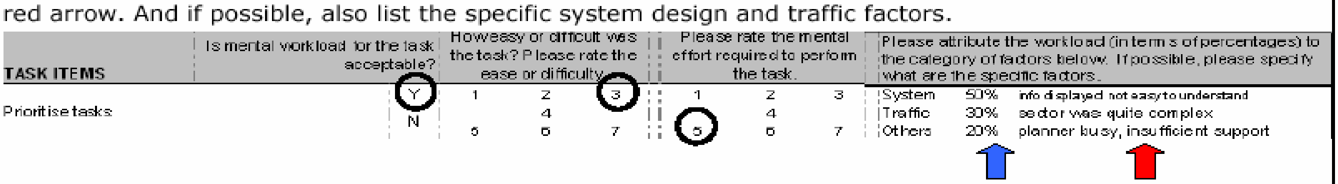
Question 3. What level of mental effort was required to perform the task? As in question 2 the possible answers are displayed on an incremental seven-point scale (shown in the red box), where 1 is the lowest rating and 7 the highest rating. 3 points on the scale (points 2, 4 and 6) are labelled with keywords. The other 4 points on the scale have been left blank but are defined as less than or more than the labelled point. To rate your answer look at the 3 keywords and decide which one is closest to your rating of 'Mental Effort Required'. If your rating is identical to the keyword then please record the number in the Recording Form. However, if you believe that your rating of mental effort required is not suitably reflected by the keyword then consider the blank points around the keyword.

For example, if you believe that you used a level of mental effort to perform the task that was greater than "moderate" but less than "large" you would select 5 as your answer. Alternatively, if you believed that you spent a very significant level of mental effort performing the task and believed this amount to be greater than large you would select 7 as your answer. Once you have selected your answer please enter it into the Recording Form.



Question 4. Please rate the percentage of workload that was due to the following factors. System, Traffic and Others. Enter your decision by writing the percentages into the Recording Form (as indicated by the blue arrow). It is important that these percentages add up to 100% in total.

If you believe that workload was due to "other" factors please list these factors in the box highlighted by the red arrow. And if possible, also list the specific system design and traffic factors.



Please repeat this process for each of the tasks listed in the Recording Form.

For the administration of AIM-Q using the relative workload decision tree

Assessment of the Impact on Mental workload (AIM)

AIM-Q Instructions (Relative Rating Scale)

What you need

- Relative Workload Decision Tree (the sheet covered in clear plastic)
- AIM-Q Recording Form (spreadsheet)
- These instructions

Introduction

This exercise has been designed to find out your views on the workload placed on you by the new system that you have just used.

It is important that you read all of the instructions before you complete this exercise.

1. Please look at the AIM-Q Recording Form. It looks like this...

TASK ITEMS	Has mental workload level for the task changed significantly in the new system ?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Identify potenial conflicts	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %
Scan information displays	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

The AIM-Q Recording Form has two purposes.

- Its first purpose is to provide you with a list of task items we wish you to evaluate using the decision tree. These items are shown in the first column of the Recording Form (highlighted by the red box).
- The second purpose is to provide you with an area to record the decisions that you have made whilst using the Relative workload decision tree (highlighted by the blue box). If the rating options in the blue box are not the same as the rating options on the recording form that you have been given, please notify one of the trial/simulation staff.

2. Please look at the Relative Workload Decision Tree. (It is covered in clear plastic)

The decision tree is designed to lead you through five questions relating to workload. The questions are highlighted by the use of this box

Your task is to use the Decision Tree to evaluate all of the task items listed in the first column of the Recording Form.

It is important when answering all of the questions in the decision tree that you base all of your answers on a comparison between the ATM system that you have just used and the system that you currently use at work in the operational environment. Unless you have been told otherwise by the trial/simulation staff. In which case, the trial/simulation staff will inform you of which system you will be comparing with.

Follow the arrows in the decision tree to lead you to the next question.

Each question in the decision tree has a number of possible answers. These answers are shown in the coloured boxes (blue orange, red and green). For example, the blue box in the decision tree provides you with a list of answers to the question, 'Was the task easier?'

To answer a question please select the most suitable answer from the list of possible answers shown in the coloured box that is next to the question.

When you have answered a question please record your answer in the recording form and then follow the arrows on the decision tree to the next question.

It is vital that ALL of your answers are recorded in the recording form.

The next page shows an example of how to evaluate at item using the decision tree and how to record your answers on the Recording Form.

Assessment of the Impact on Mental workload (AIM)

Example of how to use the Relative workload decision tree

1. Select the first task from the AIM-Q Recording Form.
2. Work through the Decision Tree answering each question for the selected task.

Question 1. Has mental workload changed significantly?

Consider the ATM system that you are currently using in the operational room. Compare the mental workload you experienced when using this new system to the current ATM system in your operational room. Has mental workload for the selected task changed significantly?

If your answer is 'NO' then by following the arrows on the decision tree you are instructed to move to the next task (the next task on the Recording Form). Please record your answer on the Recording Form by circling the 'N'.

If your answer is 'YES' then please record your answer in the Recording Form by circling the 'Y' and then move on to the next question by following the direction arrows.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Question 2. Was the task easier? If your answer is 'NO' please follow the arrows move onto question 3.

If your answer was 'YES', consider how much easier is it to perform the task in the new system. Please select the most appropriate answer from the blue box and circle the number on the Recording Form. Then follow the arrows to question 4.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Question 3. Was the task more difficult? If your answer is 'NO' please circle the 0 in the Recording Form and by following the arrows move onto question 4.

If your answer was 'YES', consider how much more difficult is it to perform the task in the new system. Please select the most appropriate answer from the orange box and circle the number on the Recording Form. Then follow the arrows to question 4.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Question 4. What was the increase or decrease in level of mental effort that was required to perform the task? The answers to this question are displayed in the red box. Select the most appropriate answer and enter you decision by circling the appropriate number in the Recording Form. Highlighted by the blue arrow in the diagram below.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Prioritise tasks	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System 50% Traffic 30% Others 20%	not easy to interpret info displayed sector was quite complex planner too busy to help

Question 5. Please rate the percentage of workload that was due to the following factors. System, Traffic and Others. Enter your decision by writing the percentages into the Recording Form, as indicated by the black arrow in the diagram above. It is important that these percentages add up to 100% in total.

If you believe that workload was due to other factors please list these factors in the box, as indicated by the red arrow. And if possible, also list the specific system design and traffic factors.

Please repeat this process for each of the tasks listed in the Recording Form.

For the administration of AIM-Cog using the absolute workload decision tree

Assessment of the Impact on Mental workload (AIM)

AIM-Cog Instructions (Absolute rating scale)

What you need

- Absolute Workload Decision Tree (the sheet covered in clear plastic)
- The correct AIM Cognitive sub-scale Recording Form (spreadsheet)
- This instruction sheet

Introduction

This exercise has been designed to find out your views on the workload placed on you by the system that you have just used.

It is important that you read all of the instructions before you complete this exercise.

1. Please look at the AIM-Cog Recording Form. It looks like this…

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.
Multitasking				
Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Information gathering and interpretation	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %
Divide attention (e.g. Speaking and writing at the same time)	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %

In the green box (in the figure above), is the name of the sub-scale of the recording form. Please ensure that the correct sub-scale has been given to you. If you have been given the incorrect recording form, please notify one of the trial/simulation staff.

The AIM-Cog Recording Form has two purposes.

- The first purpose is to provide you with a list of task items we wish you to evaluate using the decision tree. These items are shown in the first column of the Recording Form (highlighted by the red box).
- The second purpose is to provide you with an area to record the decisions that you have made whilst using the decision tree (highlighted by the blue box). If the rating options in the blue box are not the same as the rating options on the recording form that you have been given, please notify one of the trial/simulation staff.

2. Please look at the Absolute Workload Decision Tree. (It is covered in clear plastic)

The Absolute Workload Decision Tree is designed to lead you through four questions relating to workload. The questions are highlighted by the use of this box

Your task is to use the Decision Tree to evaluate all of the task items listed in the first column of the Recording Form.

The Decision tree works from left to right.

Each question in the decision tree has a number of possible answers. These answers are shown in the coloured boxes (blue, red and green). For example, the blue box in the decision tree provides you with a list of answers to the questions 'How easy or difficult was the task?

To answer a question please select the most suitable answer from the list of possible answers shown in the coloured box that is next to the question.

When you have answered a question please record your answer in the Recording Form and then follow the arrows on the decision tree to the next question.

It is vital that ALL of your answers are recorded in the Recording Form.

The next page shows an example of how to evaluate an item using the decision tree and how to record your answers on the Recording Form.

Assessment of the Impact on Mental workload (AIM)

Example of how to evaluate an item using the Absolute Workload Decision Tree

1. Select the first task from the AIM-Cog Recording Form.
2. Work through the Absolute Workload Decision Tree Decision Tree from left to right answering each question for the selected task.

Question 1. Is the workload needed to perform this task acceptable? The possible answers are 'YES' and 'NO'. To record your answers please circle either Y or N on the Recording Form.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Multitasking					
Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Question 2. How easy or difficult was the task? Please rate the difficulty of the task on a 7-point scale, shown in the blue box. The rating scale is incremental where 1 is the lowest rating and 7 the highest rating. 3 points on the scale (points 2, 4 and 6) are labelled with keywords. The other 4 points on the scale have been left blank but are defined as less than or more than the labelled point. To rate your answer look at the 3 keywords and decide which one is closest to your rating of 'Task Difficulty'. If your rating is identical to the keyword then please record the number in the Recording Form. If however, you believe that your rating of task difficulty is not suitably reflected by the keyword then consider the blank points around the keyword.

For example, if you believe that 'Prioritising tasks' was on the "Easy" side of the scale but not just 'Easy', you would select point 3 as your answer. Or, if you believe that 'Prioritising tasks' was very difficult to perform you would select point 7 as your answer. If you believed that the task was neither easy nor difficult, you would select 4 as your answer. Once you have selected your answer please enter it into the Recording Form.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Multitasking					
Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Question 3. What level of mental effort was required to perform the task? As in question 2 the possible answers are displayed on an incremental seven-point scale (shown in the red box), where 1 is the lowest rating and 7 the highest rating. 3 points on the scale (points 2, 4 and 6) are labelled with keywords. The other 4 points on the scale have been left blank but are defined as less than or more than the labelled point. To rate your answer look at the 3 keywords and decide which one is closest to your rating of 'Mental Effort Required'. If your rating is identical to the keyword then please record the number in the Recording Form. However, if you believe that your rating of mental effort required is not suitably reflected by the keyword then consider the blank points around the keyword.

For example, if you believe that you used a level of mental effort to perform the task that was greater than "moderate" but less than "large" you would select 5 as your answer. Alternatively, if you believed that you spent a very significant level of mental effort performing the task and believed this amount to be greater than large you would select 7 as your answer. Once you have selected your answer please enter it into the Recording Form.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Multitasking					
Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Traffic % Others %	

Question 4. Please rate the percentage of workload that was due to the following factors. System, Traffic and Others. Enter your decision by writing the percentages into the Recording Form (as indicated by the blue arrow). It is important that these percentages add up to 100% in total.

If you believe that workload was due to "other" factors please list these factors in the box highlighted by the red arrow. And if possible, also list the specific system design and traffic factors.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the workload (in terms of percentages) to the category of factors below. If possible, please specify what are the specific factors.	
Multitasking					
Prioritise tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System 50% Traffic 30% Others 20%	Not easy to interpret info display Sector was quite complex Planner too busy to help

Please repeat this process for each of the tasks listed in the Recording Form.

For the administration of AIM-Cog using the relative workload decision tree

Assessment of the Impact on Mental workload (AIM)

AIM-Cog Instructions (Relative Rating Scale)

What you need

- Relative Workload Decision Tree (the sheet covered in clear plastic)
- The correct AIM Cognitive sub-scale Recording Form (spreadsheet)
- This instruction sheet

Introduction

This exercise has been designed to find out your views on the workload placed on you by the new system that you have just used.

It is important that you read all of the instructions before you complete this exercise.

1. Please look at the AIM-Cog Recording Form. It should looks like this...

TASK ITEMS	Has mental workload level for the task changed significantly in the new	Is the task easier or more difficult? Please rate the change in difficulty.			Please rate the increase or decrease in mental effort required to perform the			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.		
Multitasking										
Prioritise tasks	Y	-1	-2	-3	-1	-2	-3	System	%	
	N	0			0			Traffic	%	
		1	2	3	1	2	3	Others	%	
Information gathering and interpretation	Y	-1	-2	-3	-1	-2	-3	System	%	
	N	0			0			Traffic	%	
		1	2	3	1	2	3	Others	%	
Divide attention (e.g. Speaking and writing at the same time)	Y	-1	-2	-3	-1	-2	-3	System	%	
	N	0			0			Traffic	%	
		1	2	3	1	2	3	Others	%	

In the green box (in the figure above), is the name of the sub-scale of the recording form. Please ensure that the correct sub-scale has been given to you. If you have been given the incorrect recording form, please notify one of the trial/simulation staff.

The AIM-Cog Recording Form has two purposes.

- Its first purpose is to provide you with a list of task items we wish you to evaluate using the Relative workload decision tree. These items are shown in the first column of the AIM-Cog Recording Form (highlighted by the red box).

- The second purpose is to provide you with an area to record the decisions that you have made whilst using the Relative workload decision tree (highlighted by the blue box). If the rating options in the blue box are not the same as the rating options on the recording form that you have been given, please notify one of the trial/simulation staff.

2. Please look at the Relative Workload Decision Tree. (It is covered in clear plastic)

The decision tree is designed to lead you through five questions relating to workload. The questions are highlighted by the use of this box

It is important when answering all of the questions in the decision tree that you base all of your answers on a comparison between the ATM system that you have just used and the system that you currently use at work in the operational environment. Unless you have been told otherwise by the trial/simulation staff. In which case, the trial/simulation staff will inform you of which system you will be comparing with.

Follow the arrows in the decision tree to lead you to the next question.

Each question in the decision tree has a number of possible answers. These answers are shown in the coloured boxes (blue orange, red and green). For example, the blue box in the decision tree provides you with a list of answers to the question, 'Was the task easier?'

To answer a question please select the most suitable answer from the list of possible answers shown in the coloured box that is next to the question.

When you have answered a question please record your answer in the Recording Form and then follow the arrows on the decision tree to the next question.

It is vital that ALL of your answers are recorded in the Recording Form.

The next page shows an example of how to evaluate at item using the decision tree and how to record your answers on the Recording Form.

Assessment of the Impact on Mental workload (AIM)

Example of how to use the Relative Workload Decision tree

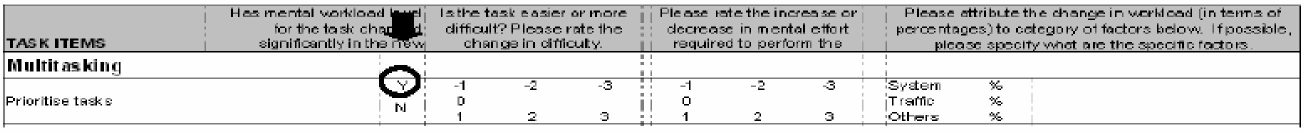
1. Select the first task from the AIM-Cog Recording Form.
2. Work through the Decision Tree answering each question for the selected task.

Question 1. Has mental workload changed significantly?

Consider the ATM system that you are currently using in the operational room. Compare the mental workload you experienced when using this new system to the current ATM system in your operational room. Has mental workload for the selected task changed significantly?

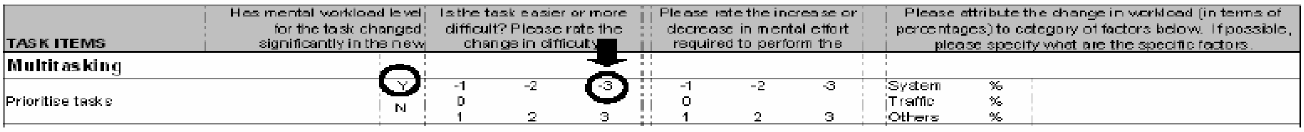
If your answer is 'NO' then by following the arrows on the decision tree you are instructed to move to the next task (the next task on the Recording Form). Please record your answer on the Recording Form by circling the 'N'.

If your answer is 'YES' then please record your answer in the Recording Form by circling the 'Y' and then move on to the next question by following the direction arrows.



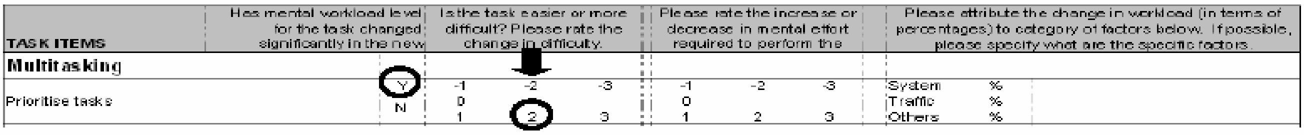
Question 2. Was the task easier? If your answer is 'NO' please follow the arrows move onto question 3.

If your answer was 'YES', consider how much easier is it to perform the task in the new system. Please select the most appropriate answer from the blue box and circle the number on the Recording Form. Then follow the arrows to question 4.

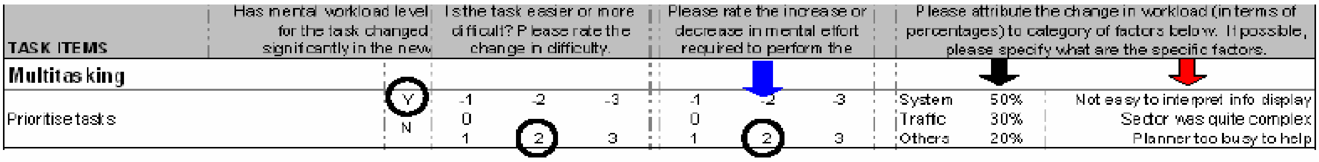


Question 3. Was the task more difficult? If your answer is 'NO' please circle the 0 in the Recording Form and by following the arrows move onto question 4.

If your answer was 'YES', consider how much more difficult is it to perform the task in the new system. Please select the most appropriate answer from the orange box and circle the number on the Recording Form. Then follow the arrows to question 4.



Question 4. What was the increase or decrease in level of mental effort that was required to perform the task? The answers to this question are displayed in the red box. Select the most appropriate answer and enter you decision by circling the appropriate number in the Recording Form. Highlighted by the blue arrow in the diagram below.



Question 5. Please rate the percentage of workload that was due to the following factors. System, Traffic and Others. Enter your decision by writing the percentages into the Recording Form, as indicated by the black arrow in the diagram above. It is important that these percentages add up to 100% in total.

If you believe that workload was due to other factors please list these factors in the box, as indicated by the red arrow. And if possible, also list the specific system design and traffic factors.

Please repeat this process for each of the tasks listed in the Recording Form.

For the administration of AIM-Hi using the absolute workload decision tree

Assessment of the Impact on Mental workload (AIM)

AIM Hi Workload Instructions (Absolute rating scale)

What you need

- Absolute Workload Decision Tree (the sheet covered in clear plastic)
- Aim-Hi Recording Form (spreadsheet)
- This instruction sheet

Introduction

This exercise has been designed to find out your views on the workload placed on you by the system that you have just used.

It is important that you read all of the instructions before you complete this exercise.

1. Please select the Recording Form. It looks like this…

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.	Please rate the mental effort required to perform the task.	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	Y N	1 2 3 4 5 6 7	1 2 3 4 5 6 7	System % Trafic % Others %
Direct attention to information sources and monitoring	Y	1 2 3	1 2 3	System %

The Recording Form has two purposes.

- The first purpose is to provide you with a list of task items we wish you to evaluate using the decision tree. These items are shown in the first column of the Recording Form (highlighted by the red box).
- The second purpose is to provide you with an area to record the decisions that you have made whilst using the decision tree (highlighted by the blue box). If the rating options in the blue box are not the same as the rating options on the recording form that you have been given, please notify one of the trial/simulation staff.

2. Please look at the Absolute Workload Decision Tree. (It is covered in clear plastic)

The Absolute Workload Decision Tree is designed to lead you through four questions relating to workload. The questions are highlighted by the use of this box

Your task is to use the Decision Tree to evaluate all of the task items listed in the first column of the Recording Form.

The Decision tree works from left to right.

Each question in the decision tree has a number of possible answers. These answers are shown in the coloured boxes (blue, red and green). For example, the blue box in the decision tree provides you with a list of answers to the questions 'How easy or difficult was the task?

To answer a question please select the most suitable answer from the list of possible answers shown in the coloured box that is next to the question.

When you have answered a question please record your answer in the Recording Form and then follow the arrows on the decision tree to the next question.

It is vital that ALL of your answers are recorded in the Recording Form.

The next page shows an example of how to evaluate an item using the decision tree and how to record your answers on the Recording Form.

Assessment of the Impact on Mental workload (AIM)

Example of how to evaluate an item using the Absolute Workload Decision Tree.

1. Select the first task from the AIM-Hi Recording Form.

2. Work through the Absolute Workload Decision Tree from left to right answering each question for the selected task.

Question 1. Is the workload needed to perform this task acceptable? The possible answers are 'YES' and 'NO'. To record your answers please circle either Y or N on the Recording Form.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.			Please rate the mental effort required to perform the task.			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	(Y)	1	2	3	1	2	3	System	%
			4			4		Traffic	%
	N	5	6	7	5	6	7	Others	%

Question 2. How easy or difficult was the task? Please rate the difficulty of the task on a 7-point scale, shown in the blue box. The rating scale is incremental where 1 is the lowest rating and 7 the highest rating. 3 points on the scale (points 2, 4 and 6) are labelled with keywords. The other 4 points on the scale have been left blank but are defined as less than or more than the labelled point. To rate your answer look at the 3 keywords and decide which one is closest to your rating of 'Task Difficulty'. If your rating is identical to the keyword then please record the number in the Recording Form. If however, you believe that your rating of task difficulty is not suitably reflected by the keyword then consider the blank points around the keyword.

For example, if you believe that 'Prioritising tasks' was on the "Easy" side of the scale but not just 'Easy', you would select point 3 as your answer. Or, if you believe that 'Prioritising tasks' was very difficult to perform you would select point 7 as your answer. If you believed that the task was neither easy nor difficult, you would select 4 as your answer. Once you have selected your answer please enter it into the Recording Form.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.			Please rate the mental effort required to perform the task.			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	(Y)	1	2	(3)	1	2	3	System	%
			4			4		Traffic	%
	N	5	6	7	5	6	7	Others	%

Question 3. What level of mental effort was required to perform the task? As in question 2 the possible answers are displayed on an incremental seven-point scale (shown in the red box), where 1 is the lowest rating and 7 the highest rating. 3 points on the scale (points 2, 4 and 6) are labelled with keywords. The other 4 points on the scale have been left blank but are defined as less than or more than the labelled point. To rate your answer look at the 3 keywords and decide which one is closest to your rating of 'Mental Effort Required'. If your rating is identical to the keyword then please record the number in the Recording Form. However, if you believe that your rating of mental effort required is not suitably reflected by the keyword then consider the blank points around the keyword.

For example, if you believe that you used a level of mental effort to perform the task that was greater than "moderate" but less than "large" you would select 5 as your answer. Alternatively, if you believed that you spent a very significant level of mental effort performing the task and believed this amount to be greater than large you would select 7 as your answer. Once you have selected your answer please enter it into the Recording Form.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.			Please rate the mental effort required to perform the task.			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	(Y)	1	2	(3)		2	3	System	%
			4			4		Traffic	%
	N	5	6	7	(5)	6	7	Others	%

Question 4. Please rate the percentage of workload that was due to the following factors. System, Traffic and Others. Enter your decision by writing the percentages into the Recording Form (as indicated by the blue arrow). It is important that these percentages add up to 100% in total.

If you believe that workload was due to "other" factors please list these factors in the box highlighted by the red arrow. And if possible, also list the specific system design and traffic factors.

TASK ITEMS	Is mental workload for the task acceptable?	How easy or difficult was the task? Please rate the ease or difficulty.			Please rate the mental effort required to perform the task.			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.		
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	(Y)	1	2	(3)	1	2	3	System	50%	info displayed not easy to understand
			4			4		Traffic	30%	sector was quite complex
	N	5	6	7	(5)	6	7	Others	20%	planner v. busy, insufficient support

Please repeat this process for each of the tasks listed in the Recording Form.

For the administration of AIM-Hi using the relative workload decision tree

Assessment of the Impact on Mental workload (AIM)

AIM Hi Instructions (Relative rating scale)

What you need

- Relative Workload Decision Tree (the sheet covered in clear plastic)
- AIM-Hi Recording Form (spreadsheet)
- This instruction sheet

Introduction

This exercise has been designed to find out your views on the workload placed on you by the new system that you have just used.

It is important that you read all of the instructions before you complete this exercise.

1. Please look at the AIM-Hi Recording Form. It looks like this...

The AIM-Hi Recording Form has two purposes.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.			Please rate the increase or decrease in mental effort required			Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Multitasking To be able to carry out several mental tasks at the same time. This includes prioritising tasks, dividing attention between tasks such as speaking and writing at the same time, switching attention between different tasks	Y	-1	-2	-3	-1	-2	-3	System	%
		0			0			Traffic	%
	N	1	2	3	1	2	3	Others	%

- Its first purpose is to provide you with a list of items we wish you to evaluate using the decision tree. These items are shown in the first column of the Recording Form (highlighted by the red box).

- The second purpose is to provide you with an area to record the decisions that you have made whilst using the decision tree (highlighted by the blue box). If the rating options in the blue box are not the same as the rating options on the recording form that you have been given, please notify one of the trial/simulation staff.

2. Please look at the Relative Workload Decision Tree. (It is covered in clear plastic)

The decision tree is designed to lead you through five questions relating to workload. The questions are highlighted by the use of this box

Your task is to use the Decision Tree to evaluate all of the items listed in the first column of the Decision Recording Form.

It is important when answering all of the questions in the decision tree that you base all of your answers on a comparison between the ATM system that you have just used and the system that you currently use at work in the operational environment. Unless you have been told otherwise by the trial/simulation staff. In which case, the trial/simulation staff will inform you of which system you will be comparing with.

Follow the arrows in the decision tree to lead you to the next question.

Each question in the decision tree has a number of possible answers. These answers are shown in the coloured boxes (blue orange, red and green). For example, the blue box in the decision tree provides you with a list of answers to the question, 'Was the task easier?'

To answer a question please select the most suitable answer from the list of possible answers shown in the coloured box that is next to the question.

When you have answered a question please record your answer in the Recording Form and then follow the arrows on the decision tree to the next question.

It is vital that ALL of your answers are recorded in the Recording Form.

The next page shows an example of how to evaluate at item using the decision tree and how to record your answers on the Recording Form.

Assessment of the Impact on Mental workload (AIM)

Example of how to use the Relative workload decision tree

1. Select the first task from the Recording Form.
2. Work through the Relative workload decision tree answering each question for the selected task.

Question 1. Has mental workload changed significantly?

Consider the ATM system that you are currently using in the operational room. Compare the mental workload you experienced when using this new system to the current ATM system in your operational room. Has mental workload for the selected task changed significantly?

If your answer is 'NO' then by following the arrows on the decision tree you are instructed to move to the next task (the next task on the Recording Form). Please record your answer on the Recording Form by circling the 'N'.

If your answer is 'YES' then please record your answer in the Recording Form by circling the 'Y' and then move on to the next question by following the direction arrows.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Multitasking	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Question 2. Was the task easier? If your answer is 'NO' please follow the arrows move onto question 3.

If your answer was 'YES', consider how much easier is it to perform the task in the new system. Please select the most appropriate answer from the blue box and circle the number on the Recording Form. Then follow the arrows to question 4.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Multitasking	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

Question 3. Was the task more difficult? If your answer is 'NO' please circle the 0 in the Recording Form and by following the arrows move onto question 4.

If your answer was 'YES', consider how much more difficult is it to perform the task in the new system. Please select the most appropriate answer from the orange box and circle the number on the Recording Form. Then

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.
Multitasking	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System % Traffic % Others %

follow the arrows to question 4.

Question 4. What was the increase or decrease in level of mental effort that was required to perform the task? The answers to this question are displayed in the red box. Select the most appropriate answer and enter you decision by circling the appropriate number in the Recording Form. Highlighted by the blue arrow in the diagram below.

TASK ITEMS	Has mental workload level for the task changed significantly in the new system?	Is the task easier or more difficult? Please rate the change in difficulty.	Please rate the increase or decrease in mental effort required to perform the	Please attribute the change in workload (in terms of percentages) to category of factors below. If possible, please specify what are the specific factors.	
Multitasking	Y N	-1 -2 -3 0 1 2 3	-1 -2 -3 0 1 2 3	System 50% Traffic 30% Others 20%	not easy to interpret info displayed sector was quite complex planner too busy to help

Question 5. Please rate the percentage of workload that was due to the following factors. System, Traffic and Others. Enter your decision by writing the percentages into the Recording Form, as indicated by the black arrow in the diagram above. It is important that these percentages add up to 100% in total.

If you believe that workload was due to other factors please list these factors in the box, as indicated by the red arrow. And if possible, also list the specific system design and traffic factors.

Please repeat this process for each of the tasks listed in the Recording Form.

APPENDIX O: AIM-Q SCORING TOOL

Mental workload ratings on individual AIM-Q recording forms should preferably be scored before the data are entered into a data analysis software programme for statistical analyses or similar, namely for each AIM-Q form administered the mean mental effort and task difficulty scores for (i) each cognitive function group, (ii) each mental resource type and (iii) each resource competition category should be computed prior to further statistical data analyses.

For example, subject X has completed a AIM-Q form after a work period in condition Y. Subject X's mental effort ratings are:

- Item 2 = 6
- Item 6 = 6
- Item 11 = 6
- Item 14 = 5
- Item 18 = 7
- Item 29 = 6

To obtain the MWL score on mental effort required by subject X in condition Y for ‘Planning’, the mean of the mental effort ratings on task items in the ‘Planning’ category of the cognitive function group taxonomy is computed. That is items 2, 6, 11, 14, 18 and 29. Thus, subject X's score on mental effort required for ‘Planning’ performance in condition Y is 6.

The ‘AIM-Q Scoring Tool’ is an electronic scoring tool constructed to automate the scoring system. The tool is based on a MS Excel Workbook. It contains the following five worksheets, each being individually named:

1. ‘Enter difficulty ratings’: Task difficulty ratings on all task items from each completed form are entered into this worksheet.

2. ‘Enter effort ratings’: Mental effort ratings on all task items from each completed form are entered into this worksheet.

3. ‘Cog function grp’: This worksheet displays the cognitive function group (i.e. multitasking, planning, decision-making, etc.) scores (mental effort and task difficulty) for each completed form.

4. ‘Mental resources’: This worksheet displays the scores (mental effort and task difficulty) on each mental resource type (i.e. spatial resources, verbal resources, encoding resources, etc.) for each completed form.

5. ‘Resource competition’: This worksheet displays the scores (mental effort and task difficulty) on each resource competition category (i.e. high, medium and low) for each completed form.

Each worksheet is described in further detail in the following paragraphs:

'Enter Difficulty ratings': Task difficulty ratings for each completed form are entered into cells along a row. Figure 5 shows an example of this worksheet and how data is entered.

With reference to Figure 5, from columns G onwards are the task item numbers. Each of these column heading is colour coded indicating the page number on which the item can be found. There is a group column heading in the row above the task item headings indicating the page numbers on AIM-Q.

The first six columns are for entering information about each completed form. For example, the subject name or identification index, the trial/simulation run number or identifier, condition or factor level, etc. Columns A-C are left blank for user's input. Columns E and F are pre-designated for trial/simulation run number or identifier ('TrialID') and subject name or identification index ('SubjID'), respectively. Using the example in Figure 5, Row 3 contains task difficulty ratings of Subject 2 in Trial 1, which was condition 3 in the trial design. The difficulty ratings for each task items are entered along the row under the appropriate column heading (i.e. task item number).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1													Page 1 (of AIM-Q)									
2				Condition	TrialID	SubjID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
3				3	1	2	7	5	4	2			7	5	2	4	4	3	4	3		6
4				3	1	1	3	5	4	4		4	2		3	3	6	3	7	4		6
5				1	2	1	2	5	4	2		5	3	2	3	3	4	3	7	4	2	4
6				1	2	3	5	3	2	5		4	4	2	3	3	7	6	3	2	1	4
7				2	3	1	4	2	5	2		4	3	2	5	4	5	2	7	4	4	5
8				2	3	3	2	3	4	3		4	3	2	3	2	4	2	4	2	1	3
9				1	4	1	4	4	4	2		5	4	2	4	4	5	2	6	3	2	5
10				1	4	2	4	5	4	3		4	5	4	2	4	4	4	4	3	3	5
11				3	5	3	4	2	3	3		4	3	2	3	3	5	3	2	3	1	4
12				3	5	2	4	4	3	4		3	4	4	2	3	4	4	4	4	4	5
13				2	6	1	2	3	3	2	2	2	2	2	3	3	2	3	3	2	2	3
14				2	6	2	4	6	4	3		4	5	3	2	3	5	5	5	4		6
15				1	7	2	4	5	5	3		4	6	5	3	3	4	4	5	5	5	6
16				1	7	3	3	2	4	2		4	3	2	2	2	3	2	2	2	2	2
17				2	8	3	3	4	2	2	2	2	2	2	2	3	5	3	2	3	1	4
18				2	8	2	4	4	3	3		3	4	3	2	3	4	2	2	3	2	4
19				3	9	1	3	4	2	3	3	3	2	2	3	3	4	3	2	3	2	3
20				3	9	3	2	2	2	2	2	2	3	1	2	2	2	2	2	2	2	2

Figure 5: An example of the 'Enter Difficulty ratings' worksheet in the AIM-Q scoring tool

'Enter Effort ratings': The format of the 'Enter Effort ratings' worksheet is an exact replication of the 'Enter difficulty ratings' worksheet. Data is entered in exactly the same way, except that the data to be entered are the mental effort ratings on each task item on individual completed AIM-Q form.

For each AIM-Q form administered the mean mental effort and task difficulty scores for (i) each cognitive function group, (ii) each mental resource type and (iii) each resource competition category will be automatically computed. The computed scores are displayed on the 'Cog Function grp', 'Mental Resources' and 'Resource Competition' worksheets respectively. Figures 6, 7 and 8 show examples of computed score on these worksheets.

The computed scores and the associated data (such as SubjID, TrialID and codes on factor levels) can be copied and inserted into another MS Excel workbook. This allows the user to conduct further statistical analyses in MS Excel or export the data into a statistical analysis programme.

In order to conduct statistical analysis on MS Excel, make sure that the data analysis tool set for MS Excel is installed. To verify whether this tool set is installed, click on 'Tools' on the windows menu and look through the drop-down menu for 'data analysis…'. If the tool is absent, install it by clicking on 'Tools' and selecting 'Add-Ins…' on the drop-down menu. On the dialog box displayed, make sure the options 'Analysis ToolPak' and 'Analysis ToolPak - VBA' are selected.

For some statistical analysis programmes (such as SPSS) data can be copied and inserted directly into a blank database file. However, it is recommended that the MWL ratings be entered in MS Excel initially, organised in a structure suitable for the specialised statistical analysis programme and then exported into the statistical analysis programme. For example, for within-subject comparisons in SPSS, data for levels of within subject factors must be entered down under a column for each level.

1				Task Difficulty									Mental Effort								
2	Conditior	TrialID	SubjID	Multitasking	Direct Attention to information sources & monitoring	Take into account of and process information	Memory Management	Build and Maintain Situation Awareness	Planning	Decision Making	Diagnosing and Problem Solving	Team Awareness	Multitasking	Direct Attention to information sources & monitoring	Take into account of and process information	Memory Management	Build and Maintain Situation Awareness	Planning	Decision Making	Diagnosing and Problem Solving	Team Awareness
3	3	1	2	5.0	5.0	3.0	4.4	3.7	4.4	2.6	2.7	4.5	5.0	5.5	4.3	5.4	5.0	6.0	4.0	4.1	5.5
4	3	1	1	3.3	3.8	4.5	5.1	4.6	5.3	4.6	4.3	3.8	3.8	4.8	5.5	4.7	4.8	5.3	4.8	4.7	3.8
5	1	2	1	2.9	3.1	3.8	3.6	3.8	4.5	3.8	3.4	3.2	2.9	3.3	3.5	4.0	4.0	4.5	4.1	3.6	3.0
6	1	2	3	3.4	2.9	4.3	3.4	4.2	3.3	2.6	3.4	3.1	3.9	3.3	4.0	3.3	4.0	3.7	2.6	3.4	3.2
7	2	3	1	3.0	3.8	3.5	3.6	3.8	3.7	4.1	3.5	3.8	3.3	4.1	4.0	4.0	4.0	3.5	3.8	3.7	3.8
8	2	3	3	2.9	2.6	3.3	2.0	2.6	2.8	2.3	2.5	2.1	2.9	2.5	3.0	1.9	2.4	2.8	2.0	2.2	2.3
9	1	4	1	3.1	3.1	3.5	4.0	4.0	4.2	4.4	3.5	3.3	3.4	3.4	3.5	4.2	4.3	4.8	5.0	4.2	3.3
10	1	4	2	3.6	3.8	3.3	3.5	3.8	3.8	3.4	3.7	4.4	3.6	3.3	3.0	3.2	3.6	4.0	3.4	3.5	4.2
11	3	5	3	2.6	2.3	2.5	2.1	3.2	3.2	2.3	2.8	2.3	2.7	3.1	2.8	2.3	3.0	3.5	2.6	2.7	2.6
12	3	5	2	3.9	3.6	3.0	4.2	3.8	3.8	3.4	3.7	4.6	3.9	3.5	3.0	4.2	3.8	3.8	3.3	3.7	4.8
13	2	6	1	2.3	2.1	2.5	3.0	2.5	2.7	3.6	2.7	2.9	2.7	2.5	3.3	3.2	3.5	3.7	3.6	3.6	3.0
14	2	6	2	3.4	4.1	3.8	4.0	3.8	4.7	3.5	3.4	4.2	3.4	4.3	3.8	4.0	3.8	4.8	3.5	3.5	4.2
15	1	7	2	3.4	4.3	3.3	3.7	3.4	4.2	3.3	3.4	3.8	3.6	4.3	3.3	3.7	3.4	4.2	3.3	3.4	3.6
16	1	7	3	2.6	2.5	2.5	2.0	2.4	2.5	2.1	2.3	2.2	2.4	2.9	2.3	2.3	2.4	2.8	2.6	2.5	2.2
17	2	8	3	2.6	2.0	2.5	2.6	2.8	3.5	2.7	2.5	2.8	2.6	2.3	2.3	2.3	2.4	3.2	2.7	2.5	2.7
18	2	8	2	3.0	2.6	2.3	2.6	2.8	3.3	2.6	2.8	3.5	2.9	2.6	2.3	2.7	2.8	3.2	2.6	2.8	3.8
19	3	9	1	2.9	2.6	3.0	2.6	3.3	3.3	3.3	3.0	3.0	3.1	3.1	3.3	3.4	3.8	4.0	4.1	3.5	3.4
20	3	9	3	1.9	2.0	2.0	1.9	1.8	1.8	2.0	1.8	2.2	2.0	2.4	2.5	1.9	2.0	2.3	2.0	2.1	2.1

Figure 6: An example of the scores on each cognitive function group from individual completed AIM-Q forms

1				Difficulty								Mental Effort							
2	Condition	run	Subj	Encoding	Central Processing	Response	Verbal	Spatial	Visual	Auditory	Motor	Encoding	Central Processing	Response	Verbal	Spatial	Visual	Auditory	Motor
3	3	1	2	3.8	3.7	6.0	4.1	3.7	3.8	3.7	7.0	4.6	4.5	5.0	4.5	4.6	4.5	4.0	7.0
4	3	1	1	4.6	4.4	5.7	4.3	4.4	4.5	4.7	7.0	4.8	4.3	3.0	4.4	4.5	4.7	4.4	4.0
5	1	2	1	3.7	3.4	2.8	3.2	3.5	3.7	3.1	3.5	4.0	3.5	2.8	3.4	3.6	4.0	3.0	3.5
6	1	2	3	3.0	3.1	1.6	2.9	3.2	3.1	2.2	1.7	3.1	3.2	1.8	2.9	3.3	3.3	2.2	2.0
7	2	3	1	3.7	3.6	3.5	3.5	3.7	3.8	2.9	5.0	3.8	3.7	3.3	3.7	3.8	3.9	3.0	4.5
8	2	3	3	2.5	2.3	1.6	2.3	2.5	2.5	1.9	1.7	2.3	2.3	2.0	2.3	2.4	2.3	1.9	2.3
9	1	4	1	3.5	3.7	2.7	3.2	3.7	3.5	2.7	4.0	3.8	4.2	2.7	3.7	4.1	3.9	2.8	4.0
10	1	4	2	4.0	3.7	3.5	3.6	3.8	4.0	3.5	4.0	3.7	3.6	3.0	3.4	3.6	3.7	3.3	3.5
11	3	5	3	2.2	2.5	1.6	2.2	2.5	2.3	1.7	1.7	2.7	2.6	1.8	2.5	2.8	2.7	2.1	1.7
12	3	5	2	3.8	3.9	3.8	3.8	3.9	3.9	3.5	4.5	3.8	3.9	3.8	3.7	3.9	3.9	3.5	4.5
13	2	6	1	2.6	2.9	2.3	2.5	2.9	2.7	2.5	3.0	3.0	3.2	2.5	2.9	3.3	3.1	2.8	4.0
14	2	6	2	4.0	3.6	4.0	3.9	3.7	3.9	3.7	5.0	4.1	3.7	4.3	4.0	3.8	4.0	3.9	5.5
15	1	7	2	3.9	3.5	3.5	3.8	3.5	3.7	3.5	3.0	3.9	3.5	3.5	3.8	3.5	3.7	3.5	3.0
16	1	7	3	2.3	2.3	2.0	2.2	2.3	2.3	2.2	2.0	2.6	2.5	1.8	2.3	2.6	2.6	2.1	1.5
17	2	8	3	2.6	2.8	2.6	2.4	2.8	2.7	2.1	3.5	2.5	2.7	2.4	2.4	2.7	2.6	2.0	3.0
18	2	8	2	2.9	2.9	2.6	2.7	2.9	2.9	2.4	2.7	2.9	2.9	2.8	2.7	2.9	2.9	2.5	3.0
19	3	9	1	2.9	2.9	2.5	2.7	3.0	3.0	2.8	3.0	3.2	3.5	2.8	3.1	3.5	3.4	3.0	4.0
20	3	9	3	2.0	2.0	1.8	1.9	2.0	2.0	1.9	2.0	2.2	2.1	2.2	2.1	2.2	2.3	1.9	3.0

Figure 7: An example of the scores on each mental resource types from individual completed AIM-Q forms

1				Difficulty			Mental Effort		
2	Condition	run	Subj	High	Medium	Low	High	Medium	Low
3	3	1	2	3.9	3.5	4.4	4.8	4.4	4.9
4	3	1	1	4.6	4.4	5.7	4.2	4.3	4.7
5	1	2	1	3.7	3.4	2.8	3.4	3.5	3.9
6	1	2	3	3.0	3.1	1.6	3.5	2.8	3.2
7	2	3	1	3.7	3.6	3.5	3.6	3.5	4.1
8	2	3	3	2.5	2.3	1.6	2.4	2.1	2.3
9	1	4	1	3.5	3.7	2.7	4.4	3.4	3.9
10	1	4	2	4.0	3.7	3.5	3.4	3.9	3.3
11	3	5	3	2.2	2.5	1.6	2.9	2.4	2.8
12	3	5	2	3.8	3.9	3.8	3.9	3.9	3.4
13	2	6	1	2.6	2.9	2.3	3.4	3.0	2.7
14	2	6	2	4.0	3.6	4.0	3.6	3.9	4.5
15	1	7	2	3.9	3.5	3.5	3.3	3.4	4.6
16	1	7	3	2.3	2.3	2.0	2.8	2.3	2.7
17	2	8	3	2.6	2.8	2.6	2.9	2.5	2.4
18	2	8	2	2.9	2.9	2.6	2.8	3.1	2.7
19	3	9	1	2.9	2.9	2.5	3.7	3.3	2.9
20	3	9	3	2.0	2.0	1.8	2.2	2.1	2.4

Figure 8: An example of the scores on each resource competition category from individual completed AIM-Q forms

APPENDIX P: PILOT STUDY ON THE VALIDITY, INTERNAL CONSISTENCY RELIABILITY AND GENERAL USABILITY OF AIM

Description of the Pilot Study

The pilot study was designed purposefully to investigate four types of preliminary construct validity (concurrent, convergent, discriminant and predictive) and internal consistency reliability of AIM, and the general usability of AIM-Q and AIM-Hi. The pilot study was held in the Human Factors Laboratory (HF Lab) in EUROCONTROL Experimental Centre (EEC) in Brétigny, France, in the first week of December 2002.

The pilot study simulated two Controller Workstations (CWS) equipped with datalink technology (which is the same version and prototype used in the DOVE simulations held between 26th May and 10th June 2002). The CWS covered East and West sectors. The planner and executive roles for each sector were bandboxed, that is each controller at a CWS controlled one sector carrying out both planner and executive roles. The pilot study also used automatic feed sectors. Three controllers took part in the pilot study. Nine measured trials were conducted over three days.

Three controllers of different nationality (French, Belgian and Austrian) participated in the pilot study. They were competent and fluent in English, even though English was their second or third language.

Design of the Pilot Study

Three different traffic samples were used in the pilot study. These traffic samples were the same ones used in DOVE simulations held between 26th May and 10th June 2002. The traffic samples varied according to high or low traffic levels (high and low) and high or low percentage of aircraft in the sample which were datalink equipped. The automated tools in the datalink technology can be used for controlling and managing only these datalink-equipped aircraft. The percentage of datalink-equipped aircraft was an indication of the amount of automated assistance the controller was able to use. The traffic samples were:

- High Traffic and Low Automation (HTLA) (traffic level was 140% and percentage of datalink-equipped aircraft was 50%);
- Low Traffic and High Automation (LTHA) (traffic level was 120% and percentage of datalink-equipped aircraft was 95%);
- Low Traffic and Low Automation (LTLA) (traffic level was 120% and percentage of datalink-equipped aircraft was 50%).

Each subject was exposed to each sample twice but not in the same sector, that is, if the subject controller traffic in HTLA sample on the East sector on the first occasion, the next time he/she controlled traffic in HTLA sample will be on the West sector. In addition, East sector was thought to be more difficult and complex than West sector to manage and control.

Hence the factors that were manipulated in the pilot study were:

1. Traffic levels.
2. Sector difficulty.
3. Amount of automated assistance.

Each trial lasted sixty minutes. After 25 minutes of each trial, the simulator was paused and AIM-Hi was administered. The controllers were given ten minutes to complete AIM-Hi. The simulation resumed and continued for another 25 minutes. At the end of each trial AIM-Q was administered. The controllers were given 25 minutes to complete AIM-Q.

Physiological measurements were also recorded during the measured trials. These were pupillary and cardiac activity. Pupillary activity was only recorded on East sector and was calibrated and set up approximately ten minutes prior to the start of each trial. Cardiac monitoring was calibrated both sectors and set up at the beginning of the day. All of the controllers wore the monitors all day. Event markers were set on the monitors to enable the extraction of relevant cardiac data (i.e. cardiac activity during measured trials) at the end of each day.

In between each trial, controllers were given a 15-20 minute coffee break. They were not given any restriction on their activities during their lunch and coffee breaks. They were allowed to drink coffee and/or smoke cigarettes during breaks.

Results and Findings

Concurrent validity

Concurrent validity refers to the ability of the measurement tool to distinguish between two groups, which are theoretically different in the construct to be measured. That is AIM is able to discriminate between different mental workload levels. To demonstrate this ability and ensure that AIM can distinguish between different workload levels, AIM should at least be able to discriminate between situations where there is clearly a difference in mental workload levels. For example, between different traffic levels, different traffic and sector complexities and between different operational roles. The sensitivity of AIM can be demonstrated if AIM is able to discriminate between situations which are similar but may be different in terms of levels controller mental workload. For example, amount of automated assistance, changes in operational procedure, increase in staff support.

In the pilot study conducted to investigate the preliminary validity of ATM, three factors were manipulated:

(i) Traffic levels: high traffic levels (140%) vs. low traffic levels (120%).

(ii) Sectors: 'East' and 'West' Sector, where 'East' sector was thought by controllers to be more difficult and complex than 'West' sector.

(iii) Percentage of aircraft in the traffic sample which were datalink equipped: the automated tools in the datalink technology can be used for controlling and managing only these datalink-equipped aircraft. The percentage of datalink-equipped aircraft was an indication of the amount of automated assistance the controller was able to use.

Traffic levels and sectors were manipulated to create trial conditions where mental workload levels are clearly different. In addition, the percentage of aircraft in the traffic sample which were datalink equipped created trial conditions which were highly similar, i.e. the only difference being the amount of automated assistance the controller was able to use. That is the impact of the automated assistance on controller mental workload. Thus, in order to show concurrent validity, AIM should be able to sufficiently discriminate between the traffic levels, sectors and amount of automated assistance with respect to the mental workload levels. Evidence of concurrent validity depends on the extent to which observed differences in AIM mental workload scores between traffic levels, sectors and amount of automated assistance match the expected differences in mental workload for these factors.

Traffic levels: In comparisons between high and low traffic levels, all other factors such as sector and amount of automated assistance available, were kept consistent. It was reasonable to assume that mental workload levels between traffic levels will be different. It was assumed the types of functions that the controllers will engage in would be similar for both high and low traffic levels. In addition, the types of mental resources required for task performance will also be similar for high and low traffic levels. Hence, it was assumed that the distribution of mental workload among function groups and mental resource types is similar for both high and low traffic levels. The difference will be in the quantity of mental workload within each function groups and mental resource types as a result of more aircraft to manage and more of each type of tasks to carry out. A diagrammatic representation of this hypothetical assumption is displayed in Figure 9.

It was therefore reasonable to expect mental workload levels (mental effort and difficulty) for high traffic level to be larger than that for low traffic level, on all the cognitive function groups and mental resource types.

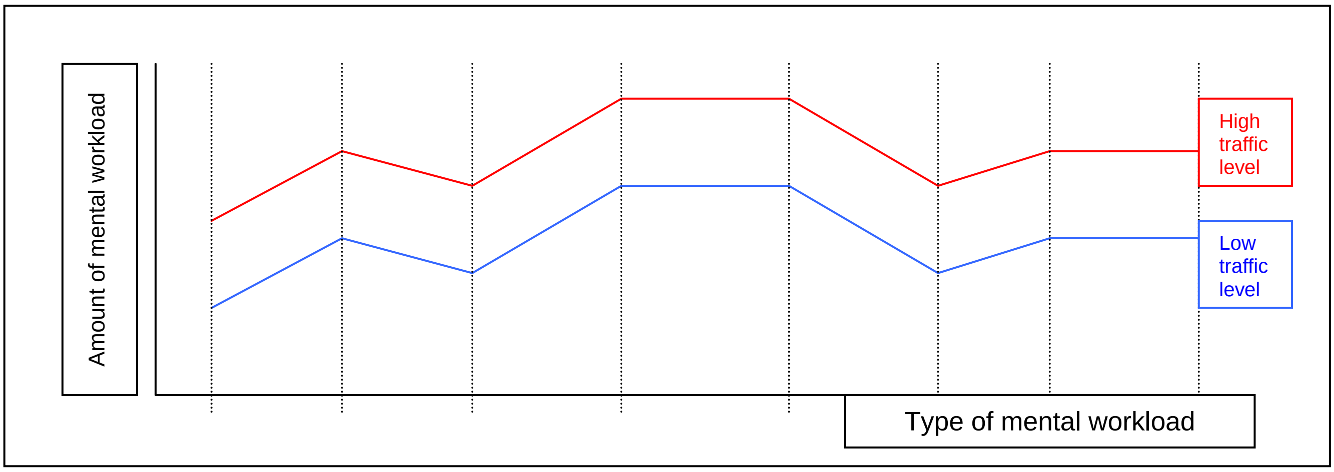


Figure 9: Same mental workload distribution in high- and low traffic levels but different amounts of mental workload

Based on the assumptions discussed in the paragraphs above about the differences in Mental Workload (MWL) between traffic levels, Table 10 displays the expected differences in AIM MWL scores when comparing mental workload between traffic levels. 'Team Awareness' was left out as controllers felt that the setup of the pilot study required very little teamwork, if any.

In the pilot validation study high traffic level was represented by traffic sample where the number of aircraft was 140% of the number of aircraft taken at a particular date. Low traffic level was represented by traffic sample where the number of aircraft was 120%. Table 11 displays the observed differences between traffic levels in AIM MWL scores found in the pilot validation study.

Overall difficulty MWL score on AIM-Q for high traffic level was significantly greater than for low traffic level. Similarly, on AIM-Hi overall difficulty MWL score for high traffic level was significantly greater than for low traffic level.

Table 10: Expected differences in AIM MWL scores between traffic levels in overall AIM-Q, cognitive function groups, mental resource types and AIM-Hi

Mental Workload (MWL) score	Difficulty	Mental effort
AIM-Hi MWL	High > Low	High > Low
Overall AIM-Q MWL	High > Low	High > Low
Cognitive function types		
Multitasking	High > Low	High > Low
Direct attention to information sources	High > Low	High > Low
Take account of and process external information	High > Low	High > Low
Memory management	High > Low	High > Low

Mental Workload (MWL) score	Difficulty	Mental effort
Build and maintain SA	High > Low	High > Low
Planning	High > Low	High > Low
Decision-making	High > Low	High > Low
Diagnosing and problem solving	High > Low	High > Low
Mental resource types		
Encoding	High > Low	High > Low
Central processing	High > Low	High > Low
Response	High > Low	High > Low
Verbal	High > Low	High > Low
Spatial	High > Low	High > Low
Visual	High > Low	High > Low
Auditory	High > Low	High > Low
Motor	High > Low	High > Low

Overall mental effort MWL score on AIM-Q for high traffic level was significantly greater than for low traffic level. Although overall mental effort MWL score on AIM-Hi for high traffic level was greater than for low traffic level, the difference was not significant.

The difficulty MWL scores on all cognitive functions groups were significantly larger for High traffic level than those were for low traffic level.

Mental effort MWL scores on planning, decision-making, and diagnosing and problem solving were significantly greater for high traffic level than those for low traffic levels. Although mental effort MWL scores for high traffic level on the other cognitive functions were greater than for low traffic level, the differences were not statistically significant.

Difficulty MWL scores for high traffic level on encoding, central processing, verbal, spatial and visual mental resources were significantly greater than for low traffic level. Although difficulty MWL scores for high traffic level on auditory resources were greater than for low traffic level, the difference was not statistically significant. No difference in difficulty MWL scores on response and motor resources were found between traffic levels.

Mental effort MWL scores for high traffic level on encoding, central processing, verbal, spatial and visual mental resources were significantly greater than for low traffic level. On the other hand, there was no difference in mental effort MWL score on auditory resource between traffic levels. Although mental effort MWL scores for high traffic level on response and motor resources were more little than for low traffic level, these differences were not statistically significant.

Table 11: Observed differences in AIM MWL scores between traffic levels in overall AIM-Q, cognitive function groups, mental resource types and AIM-Hi

Mental Workload (MWL) score	Difficulty	Mental effort
AIM-Hi MWL	140% > 120% (sig.)	*High > Low*
Overall AIM-Q MWL	140% > 120% (sig.)	140% > 120% (sig.)
<u>Cognitive function types</u>		
Multitasking	140% > 120% (sig.)	*140% > 120%*
Direct attention to information sources	140% > 120% (sig.)	*140% > 120%*
Take account of and process external information	140% > 120% (sig.)	*140% > 120%*
Memory management	140% > 120% (sig.)	*140% > 120%*
Build and maintain SA	140% > 120% (sig.)	*140% > 120%*
Planning	140% > 120% (sig.)	140% > 120% (sig.)
Decision-making	140% > 120% (sig.)	140% > 120% (sig.)
Diagnosing and problem solving	140% > 120% (sig.)	140% > 120% (sig.)
<u>Mental resource types</u>		
Encoding	140% > 120% (sig.)	140% > 120% (sig.)
Central processing	140% > 120% (sig.)	140% > 120% (sig.)
Response	-	*140% < 120%*
Verbal	140% > 120% (sig.)	140% > 120% (sig.)
Spatial	140% > 120% (sig.)	140% > 120% (sig.)
Visual	140% > 120% (sig.)	140% > 120% (sig.)
Auditory	*140% > 120%*	-
Motor	-	*140% < 120%*

In summary, 25 significant differences ($p<0.05$) between traffic levels in AIM MWL scores were found, that is 70% of the expected differences between traffic levels were observed in the pilot study (i.e. observed differences which were statistically significant). This suggests that there is evidence that AIM was able to discriminate between high and low traffic levels, with respect to mental workload levels.

<u>Sector difficulty</u>*:* In comparisons between East and West sectors, all other factors, such as traffic levels and amount of automated assistance available, were kept consistent. East sector was thought by controllers to be more difficult and challenging than West sector. It was therefore reasonable to assume that the overall mental workload levels between the sectors will be different. However, without conducting a detailed task model of each sector, it was not possible to assume that the distribution of mental workload among cognitive function will be similar for East and West sectors, or predict the

mental resources required for each sector and what the differences in mental resources demand could be reasonably expected. For example, it was not possible to assume that there is more multitasking workload than planning workload for both East and West. It may be that there is more multitasking than planning workload for East sector but the reverse in West Sector. Neither was it possible to assume without detailed sector task models that both sectors require Resource X to the same extent and the only difference between East and West is the mental workload demand on resource X. A diagrammatic representation of this hypothetical assumption is displayed in Figure 10.

Hence, within the scope of the pilot study and in the absence of sector task models, it was reasonable to expect only the overall mental workload levels (mental effort and difficulty) for East sector to be larger than that for West sector.

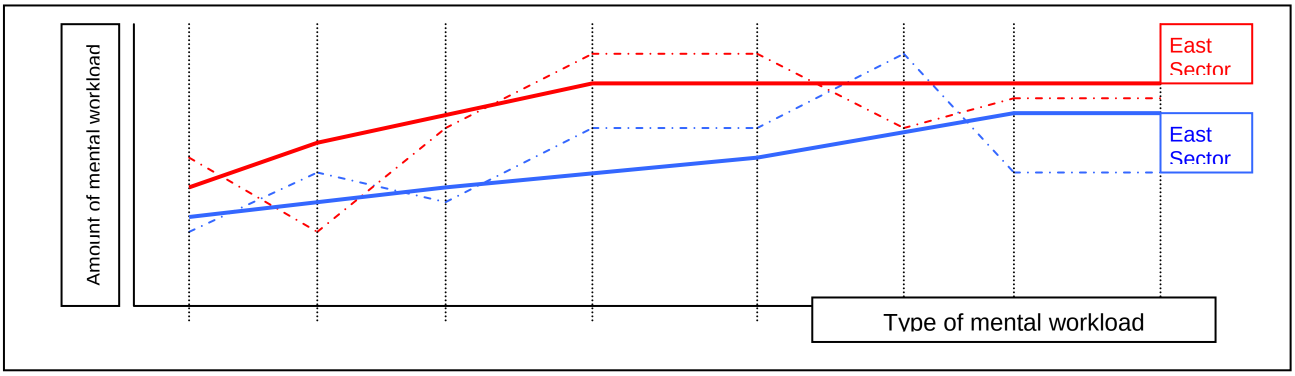


Figure 10: Different mental workload distribution (dotted lines) in East and West sectors but different amounts of overall mental workload between sectors (solid lines)

Based on the assumptions discussed in the paragraphs above about the differences in mental workload between East and West sectors, Table 12 displays the expected differences between East and West sectors in overall AIM-Q MWL scores and AIM-Hi MWL scores. Table 13 displays the observed differences between sectors in AIM MWL scores found in the pilot validation study.

Table 12: Expected differences between East and West sectors in overall AIM-Q MWL scores and AIM-Hi MWL scores

Mental Workload (MWL) score	Difficulty	Mental effort
AIM-Hi MWL	East > West	East > West
Overall AIM-Q MWL	East > West	East > West

Overall difficulty MWL scores on AIM-Q for East sector were significantly greater than for West sector. Similarly on AIM-Hi, overall difficulty MWL scores for East sector were significantly greater than for West sector.

Overall mental effort MWL scores on AIM-Q for East sector were significantly greater than for West sector. Similarly on AIM-Hi, overall mental effort MWL scores for East sector were significantly greater than for West sector.

Table 13: Observed differences between East and West sectors in overall AIM-Q MWL scores and AIM-Hi MWL scores

Mental Workload (MWL) score	Difficulty	Mental effort
AIM-Hi MWL	East > West (sig.)	East > West (sig.)
Overall AIM-Q MWL	East > West (sig.)	East > West (sig.)

In summary, all of the expected differences between East and West sectors were observed in the pilot study (i.e. observed differences which were statistically significant; $p<0.05$). This suggests that there is evidence that AIM was able to discriminate between East and West sectors, with respect to mental workload levels.

<u>Amount of automated assistance</u>*:* The amount of automated assistance available was determined by the percentage of datalink-equipped aircraft in the traffic sample. In comparisons between different amounts of automated assistance, all other factors such as traffic levels and sector were kept consistent. Given the scope of the pilot study and the knowledge about the automated system, there is a reasonable assumption that there will be a difference in the overall mental workload between different amounts of automated assistance. However, without conducting a detailed task model of how the automated system is used by the controllers, it was not possible to predict *a priori* the impact on mental resource types by the automated system and its use. Therefore, it was not possible to reasonably expect a difference in the mental workload demands on these mental resource types.

The sectors are different with respect to the complexities each present to the controllers, and the utility of the automated assistance to each sector may be different. Hence, within each sector mental workload will be compared between the different amounts of automated assistance. For East sector there is a difference in mental workload between different amounts of automated assistance. Similarly, for West sector overall mental workload between different amounts of automated assistance will also be different. No assumption was made about the direction of the differences (for example, ‘higher mental workload levels if the amount of automated assistance was larger’).

In addition, datalink technology - the automated system simulated in the pilot study - was analysed using the SHAPE Automation Framework (see EATM, 2004c). The framework predicted that most of the cognitive processes required for decision-making functions will be affected by datalink technology. These were:

- predicting,
- plan formulation,
- decide/select,
- evaluate.

There is therefore a reasonable assumption that the automated assistance may have an impact on mental workload demands on the decision-making function group, since most of the cognitive processes required for decision-making functions are impacted upon by the system. Hence, it was justifiable to expect differences in mental workload on decision-making between the different amounts of automated assistance.

Table 14 displays the expected differences between different amounts of automated assistance in overall AIM-Q MWL scores and AIM-Hi MWL scores.

Table 14: Expected differences between varying percentages of datalink-equipped aircraft in overall AIM-Q MWL scores and AIM-Hi MWL scores

<table>
<tr><th>Mental Workload (MWL) score</th><th>Difficulty</th><th>Mental effort</th></tr>
<tr><td>AIM-Hi MWL</td><td colspan="2" rowspan="2">Difference between different amounts of automated assistance</td></tr>
<tr><td>Overall AIM-Q MWL</td></tr>
<tr><td colspan="3"><u>East sector</u></td></tr>
<tr><td>AIM-Hi MWL</td><td colspan="2" rowspan="3">Difference between different amounts of automated assistance</td></tr>
<tr><td>Overall AIM-Q MWL</td></tr>
<tr><td>Decision-making</td></tr>
<tr><td colspan="3"><u>West sector</u></td></tr>
<tr><td>AIM-Hi MWL</td><td colspan="2" rowspan="3">Difference between different amounts of automated assistance</td></tr>
<tr><td>Overall AIM-Q MWL</td></tr>
<tr><td>Decision-making</td></tr>
</table>

In the pilot validation study the amount of automated assistance was by manipulating the percentage of datalink-equipped aircraft in the traffic sample. The higher the percentages the higher the amount of automated assistance available. The percentages were set at 95% or 50%. Table 15 displays the observed differences between varying percentages of datalink-equipped aircraft in overall AIM-Q MWL scores and AIM-Hi MWL scores found in the pilot validation study.

Table 15: Observed differences between varying percentages of datalink-equipped aircraft in overall AIM-Q MWL scores and AIM-Hi MWL scores

Mental Workload (MWL) score	Difficulty	Mental effort
AIM-Hi MWL	-	*95% < 50%*
Overall AIM-Q MWL	*95% > 50%*	95% < 50% (sig.)
East sector		
AIM-Hi MWL	95% > 50% (sig.)	*95% > 50%*
Overall AIM-Q MWL	95% > 50% (sig.)	-
Decision-making	95% > 50% (sig.)	-
West sector		
AIM-Hi MWL	95% < 50% (sig.)	95% < 50% (sig.)
Overall AIM-Q MWL	95% < 50% (sig.)	95% < 50% (sig.)
Decision-making	*95% < 50%*	95% < 50% (sig.)

Although overall difficulty MWL score on AIM-Q for higher amount of automated assistance (across both sectors) was larger than that for lower amount of automated assistance, the difference was not statistically significant. On AIM-Hi, no difference between the different amounts of automated assistance was found in overall difficulty MWL scores (across both sectors).

Overall mental effort MWL score (across both sectors) on AIM-Q for higher amount of automated assistance was significantly smaller than that for lower amount of automated assistance. Similarly on AIM-Hi, overall mental effort MWL score (across both sectors) for higher amount of automated assistance was smaller than that for lower amount of automated assistance, but the difference was not statistically significant.

On East sector overall difficulty MWL scores on AIM-Q and AIM-Hi for higher amount of automated assistance were significantly larger than those were for lower amount of automated assistance. No difference between the different amounts of automated assistance was found in overall mental effort MWL scores on AIM-Q. Although on AIM-Hi the overall mental effort MWL score for higher amount of automated assistance was larger than that for lower amount of automated assistance, the difference was not significant.

Furthermore, on East sector, decision-making difficulty MWL scores for higher amount of automated assistance was significantly larger than that for lower amount of automated assistance. However, no difference between the different amounts of automated assistance was found in mental effort MWL scores on decision-making.

On West sector, overall difficulty and mental effort MWL scores on AIM-Q for higher amount of automated assistance were significantly smaller than those were for lower amount of automated assistance. Similarly, overall difficulty and mental effort MWL scores on AIM-Hi for higher amount of automated assistance were also significantly smaller than those were for lower amount of automated assistance.

Furthermore, on West sector, decision-making difficulty MWL scores for higher amount of automated assistance was significantly smaller than that for lower amount of automated assistance. Decision-making difficulty MWL scores for higher amount of automated assistance was also smaller than that for lower amount of automated assistance, but the difference was not statistically significant.

In summary, nine significant differences ($p<0.05$) between varying amounts of automated assistance and AIM MWL scores were found, that is 56% of the expected differences between varying amounts of automated assistance were observed in the pilot study (i.e. observed differences that were statistically significant). This suggests that there is evidence that AIM was able to discriminate between varying amounts of automated assistance.

The percentage of expected differences observed in the study was not as high as for traffic levels and sector. There are a few issues that may have contributed to this. The first issue is the different impact that the automated assistance has on the different sector. The mental workload scores reported by the controllers suggest that on East sector the mental workload was larger for higher amount of automated assistance than that for lower amount of automated assistance. On the other hand, mental workload scores reported by the controllers suggest that on West sector the mental workload was smaller for higher amount of automated assistance than that for lower amount of automated assistance. This could result in the lack of significant differences in the overall mental workload comparison across sectors.

The next issue is that the impact on mental workload between 50% and 95% of datalink-equipped aircraft is marginal. It is possible that the utility of the automated tools evens out at a certain percentage of datalink-equipped aircraft. It this is true, then the difference in the impact is marginal. Although significant differences were found in West sectors for both difficulty and mental effort MWL scores, only difficulty MWL score was significantly different in East sector. On the physiological measurements a significant difference in pupil diameter[10] was found in East sector but no significant difference in mental workload was found in HRV measurements. It may be prudent to examine this potential effect and investigate further the differences in mental workload between varying amounts of automated assistance from datalink technology by replicating this study using different percentages of datalink-equipped aircraft (for example, 5%, 25%, 50%, 75% and 95%).

[10] Pupil diameter was measured only on East sector

Lastly, the controllers who volunteered to participate in the study did not have prior experience in using datalink technology. They were briefed and were given explanations on datalink technology and two training trials. The controllers expressed their feeling that this was insufficient and that they had difficulties using the datalink technology because they were unfamiliar with the graphical user interface of the tools available in the system and ignorant of what functions and assistance were available to them.

These issues needs to be considered and examined in further validation research on AIM, and they will be further discussed in Section 7.

Conclusion: The observed results and findings suggest that AIM has fairly good concurrent validity.

Convergent and discriminant validities

Convergent validity refers to the extent to which measures of the same constructs that theoretically should be related to each other are in fact observed to be related to each other (Trochim, 2000). That is the measures of mental workload from AIM should correlate highly or show convergence with measures of mental workload from other assessment tools that purport to measure mental workload.

Discriminant validity refers to the extent to which measures of other constructs are observed not to be related to each other (Trochim, 2000). That is the measures of mental workload from AIM should **not** correlate highly or show divergence from assessment tools that do not measure mental workload.

To demonstrate convergent validity, the correlation coefficients between mental workload scores from AIM and mental workload measures from other techniques or tools should be significantly high. These can be other subjective mental workload assessment tools (for example, NASA-TLX, ISA, etc.) or physiological indicators of mental workload (such as pupil diameter, heart rate variability, brain wave activity, etc.).

To demonstrate discriminant validity, the correlation coefficients between mental workload scores from AIM and dissimilar measures from other techniques or tools should be low (i.e. insignificant correlation coefficients). These can be other subjective assessment tools (for example, of situational awareness, fatigue, etc.) or physiological indicators of physical workload (such as inter-beat-interval or heart rate). Although there may be relationship between these measures and mental workload, discriminant validity is more powerful if correlations between AIM and these measures are insignificant, i.e. AIM measures are able to discriminate between mental workload and constructs with which it interacts.

In the pilot study conducted to investigate the preliminary validity of ATM, measurements of physiological activity were also taken. No other subjective assessment tool was used because of the time constraints between measured trials and on the duration of the study. The three physiological measurements were:

(i) Heart Rate (HR) or Inter-beat-interval (IBI): Have been used in other workload research studies. Several research studies have found that manipulation in mental workload may not affect HR/IBI while manipulation in physical workload is likely to affect HR/IBI, because HR/IBI as workload indicators are sensitive to physical workload changes rather than to mental workload changes. However, HR/IBI are very sensitive also to other physiological reactions, bodily functions and motor responses (e.g. physical limb movements and vocal responses).

(ii) Heart-Rate Variability (HRV): Has been increasingly used in other workload research studies. Several research studies have found that manipulation in mental workload is more likely to affect HRV than HR/IBI because HRV is more sensitive than HR/IBI as an indicator of mental workload changes. Nevertheless, similarly to HR/IBI, HRV measurements have appeared to be very sensitive to other physiological reactions and may be affected by these physiological responses bodily functions and movements.

(iii) Pupil diameter: Has been used in other research studies as an indicator of mental workload. Research findings on the sensitivity and utility of pupil diameter as a measurement of mental workload has been mixed, as pupillary responses vary according to other physiological reactions and environmental changes, apart from mental workload.

In the pilot validation study pupil diameter and HRV were used as convergent measures and IBI was used as a discriminant measure. Thus, in order to show convergent validity, AIM should correlate highly (with statistical significance) with pupil diameter measurements and HRV measurements and to show discriminant validity AIM should **not** correlate significantly with IBI (i.e. no statistically significant correlations). Evidence of convergent and discriminant validity depends on the extent to which observed correlation coefficients between AIM mental workload scores and pupil diameter, HRV and IBI match the expected correlations.

Table 16 displays the expected correlations between the AIM measures of different aspects of mental workload and the physiological indicators. Each number in parenthesis indicates the number of possible correlations. For example, there are eight sets of AIM MWL scores for each cognitive function group. Thus, eight correlation coefficients with pupil diameter are expected.

Table 16: Expected correlations between AIM measures of the different aspects of mental workload and the physiological indicators

	Pupil Diameter	*HRV*	*IBI*
Overall MWL	Highly correlated (1)	Highly correlated (1)	No. sig correlation (1)
Cognitive functions	Highly correlated (8)	Highly correlated (8)	No. sig correlation (8)
Mental Resources	Highly correlated (8)	Highly correlated (8)	No. sig correlation (8)
Aim Hi MWL	Highly correlated (1)	Highly correlated (1)	No. sig correlation (1)

Table 17 displays the observed correlation coefficients between the overall MWL scores from AIM-Q and AIM-Hi, the MWL scores from each cognitive function group, and the physiological indicators.

Overall MWL scores from AIM-Q correlated significantly ($p<0.005$) with pupil diameter measurements, but correlation with HRV measurements was not signiticant. In addition, overall MWL scores from AIM-Q did not correlate with IBI. The results suggest that overall MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Overall MWL scores from AIM-Hi correlated significantly ($p<0.10$) with pupil diameter measurements, but correlation with HRV measurements was not signiticant. In addition, overall MWL scores from AIM-Hi did not correlate with IBI. The results suggest that overall MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Table 17: Observed correlation coefficients between the overall AIM-Q and AIM-Hi MWL scores, each cognitive function group MWL score, and the physiological indicators

	Pupil Diameter	*HRV*	*IBI*
Overall AIM Q MWL	- 0.703 (.001)	n.s	No correlation
AIM Q Multitasking	- 0.427 (*.077*)	- 0.489 (.002)	No correlation
AIM Q Direct Attention	- 0.799 (.000)	- 0.399 (.016)	No correlation
AIM Q Take into account information	n.s	n.s	No correlation
AIM Q Memory Management	- 0.793 (.000)	- 0.307 (*.069*)	No correlation
AIM Q Build & maintain SA	n.s	n.s	No correlation
AIM Q Planning	- 0.678 (.002)	n.s	No correlation
AIM Q Decision Making	n.s	n.s	No correlation
AIM Q Diagnose & Problem solving	n.s	n.s	No correlation
Overall AIM Hi MWL	- 0.407 (*.094*)	n.s	No correlation

Multitasking MWL scores correlated significantly (p<0.10) with pupil diameter measurements and correlated significantly (p<0.005) with HRV measurements. In addition, multitasking MWL scores did not correlate with IBI. The results suggest that multitasking MWL scores showed convergence with pupil diameter and HRV measurements, and discriminance from IBI measurements.

Direct attention MWL scores correlated significantly (p<0.000) with pupil diameter measurements and correlated significantly (p<0.05) with HRV measurements. In addition, direct attention MWL scores did not correlate with IBI. The results suggest that direct attention MWL scores showed convergence with pupil diameter and HRV measurements, and discriminance from IBI measurements.

Take into account information MWL scores did not correlate significantly with pupil diameter and HRV measurements. Take into account information MWL scores did not correlate with IBI. The results suggest that take into account information MWL scores did not show statistically significant convergence with pupil diameter and HRV measurements, but showed discriminance from IBI measurements.

Memory management MWL scores correlated significantly ($p<0.000$) with pupil diameter measurements and correlated significantly ($p<0.10$) with HRV measurements. In addition, memory management MWL scores did not correlate with IBI. The results suggest that memory management MWL scores showed convergence with pupil diameter and HRV measurements, and discriminance from IBI measurements.

Build and maintain SA MWL scores did not correlate significantly with pupil diameter and HRV measurements. Build and maintain SA did not correlate with IBI. The results suggest that build and maintain SA MWL scores did not show statistically significant convergence with pupil diameter and HRV measurements, but showed discriminance from IBI measurements.

Planning MWL scores from AIM-Q correlated significantly ($p<0.005$) with pupil diameter measurements, but correlation with HRV measurements was not signiticant. In addition, overall MWL scores from AIM-Q did not correlate with IBI. The results suggest that overall MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Decision-making scores did not correlate significantly with pupil diameter and HRV measurements. Decision-making did not correlate with IBI. The results suggest that decision-making MWL scores did not show statistically significant convergence with pupil diameter and HRV measurements, but showed discriminance from IBI measurements.

Diagnosing and problem solving scores did not correlate significantly with pupil diameter and HRV measurements. Diagnosing and problem solving did not correlate with IBI. The results suggest that diagnosing and problem solving MWL scores did not show statistically significant convergence with pupil diameter and HRV measurements, but showed discriminance from IBI measurements.

Table 18 displays the observed correlation coefficients between each of the mental resource type MWL scores and the physiological indicators.

Table 18: Observed correlation coefficients between the each mental resource type MWL scores and the physiological indicators

	Pupil Diameter	*HRV*	*IBI*
Encoding	- 0.885 (.000)	n.s	No correlation
Central processing	- 0.712 (.001)	n.s	No correlation
Response	- 0.829 (.000)	- 0.383 (.021)	-
Verbal	- 0.896 (.000)	- 0.321(*.057*)	No correlation
Spatial	- 0.723 (.001)	n.s	No correlation
Visual	- 0.836 (.000)	n.s	No correlation
Auditory	- 0.959 (.000)	- 0.303 (*.072*)	-
Motor	- 0.730 (.001)	n.s	-

Encoding MWL scores from AIM-Q correlated significantly ($p<0.000$) with pupil diameter measurements, but correlation with HRV measurements was not significant. In addition, encoding MWL scores from AIM-Q did not correlate with IBI. The results suggest that encoding MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Central processing MWL scores from AIM-Q correlated significantly ($p<0.005$) with pupil diameter measurements, but correlation with HRV measurements was not significant. In addition, central processing MWL scores from AIM-Q did not correlate with IBI. The results suggest that central processing MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Response MWL scores correlated significantly ($p<0.000$) with pupil diameter measurements and correlated significantly ($p<0.05$) with HRV measurements. However, response MWL scores also correlated significantly with IBI. The results suggest that response MWL scores showed convergence with pupil diameter and HRV measurements. However, response MWL scores did not show discriminance from IBI measurements, instead showed convergence with IBI measurements.

Verbal MWL scores correlated significantly (p<0.000) with pupil diameter measurements and correlated significantly (p<0.10) with HRV measurements. In addition, verbal MWL scores did not correlate with IBI. The results suggest that verbal MWL scores showed convergence with pupil diameter and HRV measurements, and discriminance from IBI measurements.

Spatial MWL scores from AIM-Q correlated significantly (p<0.005) with pupil diameter measurements, but correlation with HRV measurements was not signiticant. In addition, spatial MWL scores from AIM-Q did not correlate with IBI. The results suggest that spatial MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Visual MWL scores from AIM-Q correlated significantly (p<0.000) with pupil diameter measurements, but correlation with HRV measurements was not signiticant. In addition, visual MWL scores from AIM-Q did not correlate with IBI. The results suggest that visual MWL scores from AIM-Q showed convergence with pupil diameter but not with HRV measurements, and discriminance from IBI measurements.

Auditory MWL scores correlated significantly (p<0.000) with pupil diameter measurements and correlated significantly (p<0.10) with HRV measurements. However, auditory MWL scores also correlated significantly with IBI. The results suggest that auditory MWL scores showed convergence with pupil diameter and HRV measurements. However, auditory MWL scores did not show discriminance from IBI measurements, instead showed convergence with IBI measurements.

Motor MWL scores correlated significantly (p<0.005) with pupil diameter measurements but correlation with HRV measurements was not signiticant. However, motor MWL scores also correlated significantly with IBI. The results suggest that motor MWL scores showed convergence with pupil diameter but not with HRV measurements. Moreover, motor MWL scores did not show discriminance from IBI measurements, instead showed convergence with IBI measurements.

In summary, fourteen significant correlation coefficients between AIM MWL scores and pupil diameter were found, that is 78% of the expected correlations were observed in the pilot study (i.e. observed correlation coefficients that were statistically significant). On the other hand, only six significant correlation coefficients between AIM MWL scores and HRV measurements were found, that is 33% of the expected correlations were observed in the pilot study. This suggests that there is evidence that AIM has good convergence with pupil diameter but poor convergence with HRV measurements.

In addition, only three significant correlation coefficients between AIM MWL scores and IBI measurements were found, that is 83% of the expected correlations were observed in the pilot study (i.e. observed correlation coefficients that were **not** statistically significant). This suggests that there is evidence that AIM has good discriminance from IBI measurements.

Predictive validity

Predictive validity refers to the ability of the measurement to predict something it should theoretically be able to predict. AIM was designed to be able to indicate if the system change has an impact on the spare processing capacity of the controller, that is if the system change imposes an increase in mental workload on task performance which involves high resource competition/ conflict, the spare processing capacity of the controller will be potentially reduced by the system change. To demonstrate this ability AIM should at least be able to predict in situations where there is clearly an impact on spare processing capacity, for example, between different traffic levels or different traffic and sector complexities.

In the validation study traffic levels and sectors were manipulated to create trial conditions where mental workload levels are clearly different. It is reasonable to predict that if traffic levels were high, spare processing capacity would be reduced, as there are more aircraft to manage in the time available. It is also reasonable to predict that the spare processing capacity when controlling East sector would be lower than when controlling West sector, given the same traffic levels and controllers. The reason being East sector is more difficult and complex than West sector.

In addition, the percentage of aircraft in the traffic sample which were datalink equipped was also manipulated, the only difference being the amount of automated assistance the controller was able to use. Although it was not possible to predict the direction of the impact, it was reasonable to predict that there will be an impact on spare processing capacity from the amount of automated assistance. Evidence of predictive validity depends on the extent to which observed predictions on spare processing capacity from AIM in situations of different traffic levels, sectors and amount of automated assistance match the expected predictions for these factors.

AIM uses the increase or decrease in MWL scores on high resource competition/conflict task performance to predict the impact on spare processing capacity, that is a reduction in spare processing capacity is predicted from an increase in MWL scores on high resource competition/ conflict task performance. Table 19 displays the expected predictions on spare processing capacity for changes in traffic levels, sectors and amount of automated assistance and the expected MWL scores. Each number in parenthesis indicates the number of possible predictions.

Table 19: Expected predictions on spare processing capacity for changes in traffic levels, sectors and amount of automated assistance

	Traffic level (high vs. low levels)	**Sector (East vs. West)**	**Automated assistance (95% vs. 50%)**
Spare processing capacity	Spare processing capacity in high traffic level is smaller than in low traffic levels => high resource competition MWL scores in high traffic > than in low traffic (2)	Spare processing capacity in East sector is smaller than in West sector => high resource competition MWL scores in East > than in West sector (2)	There is a difference in spare processing capacity between amounts of automated assistance => high resource competition MWL scores in 95% ≠ than in 50% (2)

Table 20 displays the observed predictions found on spare processing capacity for changes in traffic levels, sectors and amount of automated assistance in the pilot study.

Table 20: Observed predictions on spare processing capacity for changes in traffic levels, sectors and amount of automated assistance

Traffic level (high vs. low levels)		**Sector (East vs. West)**		**Automated assistance (95% vs. 50%)**	
MWL scores on high resource competition task performance					
Difficulty	**Mental effort**	**Difficulty**	**Mental effort**	**Difficulty**	**Mental effort**
High > Low traffic (p< .000)	High > Low traffic (p< .000)	East > West (p< .000)	East > West (p< .000)	-	95% < 50% (p< .011)

Difficulty and mental effort MWL scores on high resource competition task performance for high traffic level were significantly larger than that for low traffic level. Similarly, difficulty and mental effort MWL scores on high resource competition task performance for East sector were significantly larger than that for West sector. On amount of automated assistance, mental effort MWL scores on high resource competition task performance for higher automated assistance was significantly smaller than that for lower automated assistance.

In summary, five significant predictions on spare processing capacity were found, that is 83% of the expected predictions were observed in the pilot study (i.e. observed predictions that were statistically significant). This suggests that there is evidence that AIM has fairly good predictive validity.

Internal consistency reliability

Internal consistency reliability refers to how well the items in the tool reflect the same construct and yield similar results, that is how well the items in AIM reflect mental workload and how well items in the same sub-tests reflect the same aspect of mental workload.

Internal consistency reliability was measured using Cronbach's alpha to estimate the internal consistency within:

- AIM-Hi and its items,
- AIM-Q and its items,
- the AIM-Cog sub-tests and the items within each sub-test.

Table 21 displays the internal consistency reliability estimates for the three versions of AIM.

Table 21: Internal consistency reliability estimates for the three versions of AIM (statistical significance is indicated in parentheses)

Version of AIM	Cronbach's Alpha
AIM-Hi	0.8790 (p < .000)
AIM-Q	0.9361 (p < .000)
AIM-Cog	
Multitasking	0.6514 (p< .000)
Direct attention	0.7954 (p< .000)
Take into account	0.5460 (p< .005)
Memory management	0.6181 (p< .01)
Build and maintain SA	0.6239 (p< .000)
Planning	0.8136 (p< .000)
Decision-making	0.8916 (p< .01)
Diagnose and problem solving	0.8354 (p< .000)

Significant internal consistency reliability estimates ($p<0.01$) were found for AIM-Hi, AIM-Q and all of the AIM-Cog sub-tests. The results suggest that the items in AIM-Hi are consistent in their measurements. Similarly, the results suggest that the items in AIM-Q are consistent in their measurements. Finally, the results for the AIM-Cog sub-tests suggest evidence that the items in each sub-test are consistent in their measurement. The findings indicate good internal consistency reliability in AIM.

Page intentionally left blank