GRIFFITH UNIVERSITY

# Development and Proof of Concept of a Predictive Model of Flight Deck Cognitive Workload

## Author

Harris, D, Scott, S

## Published

2024

## Journal Title

Aviation Psychology and Applied Human Factors

## Version

Accepted Manuscript (AM)

## DOI

10.1027/2192-0923/a000267

## Rights statement

This accepted manuscript is distributed under the Creative Commons CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

## Downloaded from

https://hdl.handle.net/10072/432522

## Griffith Research Online

https://research-repository.griffith.edu.au

# Development and Proof of Concept of a Predictive Model of Flight Deck Cognitive Workload

Don Harris[1]

Steve Scott[1]

[1] Coventry University, Faculty of Engineering, Environment and Computing, Coventry, UK

**Abstract:**

Workload is the 'cost' (in information processing terms) of performing a flight task. Predictive models are valuable at the early stages of system design to identify potentially excessive cognitive demands before major system development commences. This paper describes the development of a task-based predictive workload method. It utilises timeline analysis supplemented by characterisations of the cognitive nature of the tasks using Multiple Resource Theory. This is followed by the identification of potential task conflicts and estimates of the influence of the task environment. Workload predictions were made for a manual approach and landing in a modern airliner. Predictions were compared against corresponding Bedford scale workload estimates. Results show the predictive technique produced comparable estimates of workload.

**Keywords:** Workload; Workload Modelling; Workload Prediction; Task Analysis

Don Harris

Faculty of Engineering, Environment and Computing

Coventry University

Coventry CV1 5FB

United Kingdom

don.harris@coventry.ac.uk

## Publication Ethics

Informed consent was obtained from all participants included in the study. Project was granted ethical approval by the Coventry University Research Ethics Committee (Project Number P106244)

## Funding

This project has received funding from the Clean Sky 2 Joint Undertaking (JU) under grant agreement No 945583. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Clean Sky 2 JU members other than the Union.

# Introduction

There is no universally accepted definition of mental workload, but it is generally acknowledged that it is the experience resulting from task demands related to competition for limited information processing resources. It is regarded as the information processing 'cost' of performing a given task (or tasks). It is intimately related to information processing theory and specifically to the finite capacity of cognitive resources in working memory (Harris, 2011). Workload is the cognitive corollary of task load (the imposition of tasks) and is one of the limiting human factors in the design of complex socio-technical systems.

Workload has three basic sources: the nature of the task or tasks (their cognitive demands; number and nature of tasks to be performed concurrently; time pressure, etc.); the task environment, and the skill/experience of the user (Johannsen, 1979). The design of flight deck equipment affects the sources of workload and can also mediate some aspects of the task environment. As a result, it is desirable to be able to predict the impact of new equipment on pilot workload at the early stages of the design process. Predictive workload models are particularly useful before a new system using new procedures is implemented. With a new design there is a requirement to foresee information processing conflicts and identify potential workload peaks where system demands exceed the operator's cognitive capacity. The demands imposed by the task, system and environment will inform issues such as the number of crew required, the scheduling of tasks and determine if further automated assistance is required to keep cognitive demands within acceptable bounds. Measures of workload (e.g. using workload scales and/or physiological measures) only verify if a design goal has been achieved; they do not assess if a design is likely to achieve its human performance objectives (see the 'V' development cycle: Reuzeau & Nibbelke, 2004).

Analytical approaches to workload estimation are usually based upon a formal task analysis (Xie & Salvendy, 2010), often incorporating elements of both Hierarchical Task Analysis (HTA) and Cognitive Task Analysis (CTA), with tasks described on a timeline. In many cases these task analyses are supplemented by an underlying model of human information processing (for example multiple resource theory - MRT: Wickens, 1984). The TLAP model (Time-Line Analysis Procedure: Parks & Boucek, 1989) is one such timeline-based model not dependent upon an underlying theory, but which simply imposes an additional workload penalty when two or more tasks have to be performed simultaneously, so may imply an approach requiring multiple cognitive resources.

However, most predictive models of workload are predicated upon a basic theory of human information processing which conceives that some tasks can be shared concurrently (e.g. a visual and an auditory task) but other combinations of tasks can only be shared with limited success (e.g. two verbal tasks, or two visual tasks: see Baddeley & Hitch, 1974; Wickens, 1984; Baddeley, 2000). AWAS (Aircrew Workload Assessment System) is such an analytical approach which utilises a description of piloting tasks arranged on a task timeline. Tasks are assessed for their individual demands on each of Wickens' MRT channels (Wickens, 1984) and conflicting demands (where two simultaneous tasks occupy the same information processing channel) are also identified (Davies *et al*., 1995). The results are in the form of a continuous output of predicted pilot workload. Like AWAS, W/INDEX (Workload Index) also draws upon MRT to identify task conflicts and assess their implications for operator workload (North & Riley, 1989). W/INDEX is scenario-based and is also predicted upon a timeline analysis. It is specifically aimed at establishing the likely workload benefits associated with the introduction of new equipment.

Hamilton & Bierbaum (1990) developed the Task Analysis Workload (TAWL) modelling approach. This was again based upon a multiple resource model of human information processing to estimate the cognitive, psychomotor and sensory components of workload. TAWL was used to estimate the workload imposed and assess the likely contributions to workload management offered by the development of new flight deck technologies. VACP (visual, auditory, cognitive, psychomotor) described by Mitchell (2000) is a further development of TAWL. It has been used to make design decisions early in the product development cycle to assess workload trade-offs (e.g. Mitchell & Samms, 2010). Wickens also developed a predictive model based upon the computational version of MRT (Wickens, 2002) which operated on a similar basis to AWAS. Any shared tasks identified from a cognitive task analysis were analysed with respect to the demands they placed on the operator in four categories: working memory (spatial); working memory (verbal); response (manual) and response (verbal). This analysis for the basis of an estimate of workload and identified conflicting task demands that could potentially degrade performance.

It is desirable to assess any effect on workload of new flight deck functions early in the design stages, before commencing detailed design work and building prototypes. There is a 1:10:100 "rule" for design costs (Pressman, 1992). For every £1 to fix a problem identified in the early design stages it will cost £10 to fix it during detailed design and development, and £100 to fix it once the system is operational. An early predictive analysis of task scheduling and sharing on the flight deck helps to determine potential pilot information processing bottlenecks and identify periods of unacceptably high workload. However, to be of any utility the estimates derived by the model should show a high degree of predictive validity. Many earlier predictive models have been validated against measures of performance (see Sarno & Wickens, 1992; Wickens *et al*., 2013: McMillan *et al*., 1991) rather than against measures of workload, which is less common (e.g. Rusnock, *et al*., 2015; Rusnock & Geiger, 2016; Borghetti, *et al*., 2017). Where there has been an attempt to validate against an extant workload measure, the correlations have generally been higher for task performance than for workload (Wickens *et al*., 2013). However, there is often a dissociation observed between workload and performance (Yeh & Wickens, 1988; Mansikka, *et al*., 2019). Because of the dissociation often observed between workload and performance it is important to compare any prediction of workload against an extant workload measure. In this study the predictive workload estimates from the model are compared against an existing workload scale in common usage (Bedford scale: Ellis & Roscoe, 1982) making this a direct assessment of the predictions from this modelling approach.

A key objective of this work was to develop a simple predictive model of workload that was not computationally complex, and which could be applied easily with minimal domain-specific expertise. The approach developed requires no bespoke software. It may be implemented using simple HTA software and a spreadsheet. Nevertheless, it remains theoretically robust, being grounded upon the principles of MRT.

# Model Development

## Premise underlying the predictive model

Workload is the cognitive cost of performing certain flight task(s), within a task environment using items of flight deck equipment. For simplicity, the effects of factors such as fatigue and stress on workload will be discounted in this instance.

For modelling purposes, cognitive input load is considered to be imposed primarily by the task and the environment in which it is performed; flight deck equipment should help to alleviate these imposed demands (rather than create additional demands on the crew). It should moderate the task demands imposed upon the pilot. This may be achieved in several manners, for example by the integration of information; simplification of procedures; reduction in required calculations, etc.

The predictive workload model developed was based upon a seven-stage process:

- Timeline-based analysis of required flight tasks
- Characterisation of the flight task components and derivation of task input cognitive load estimates
- Allocation of tasks to an appropriate cognitive category
- If two tasks are undertaken simultaneously, derivation of a task conflict score.
- Characterisation of the task environment using antagonistic noun pairs (used to characterise the task environment).
- Timeline-based calculation and presentation of the workload estimate.

## Timeline-based analysis of required flight tasks

Workload can vary dramatically from minute to minute during certain stages of flight. A major criticism of workload scales not used concurrently with the task is that it is unclear if measures relate to either peak or average workload, or workload at the point of questioning. However, when workload imposed exceeds a pilot's cognitive capacity to cope, no matter for how short a period, then from both a flight safety and performance perspective, this is unacceptable (workload redlines: see Rueb, *et al*., 1994).

The predictive model of workload is predicated upon a time-line based task analysis of a specific flight scenario. Timeline analyses are based upon the bottom-level tasks in an HTA (those tasks which are not subject to further de-composition and which constitute an operation – e.g. see Huddlestone & Stanton, 2016). Tasks undertaken simultaneously are also identified to allow the calculation of workload when there are potentially competing tasks. These bottom level tasks forming the basic building blocks of the modelling technique and are derived from a number of actual flight scenarios. However, it is anticipated that when being used in a design context, the task analyses would comprise prospective analyses of the way it is anticipated that the task would be undertaken. New equipment may change the basic nature of tasks (for example, automation of flight path control changes the nature of the task from one of manual control to one of supervisory control/monitoring). As a result, any new equipment will also impact the competition between tasks for cognitive resources which will serve to further change the workload experienced. This may require new HTAs to be undertaken.

## Characterising the features of the flying task

Each of the bottom level tasks depicted on the timeline are characterised using a set of verbs describing the task characteristics. In contrast to the VACP model (Mitchell, 2000) where task descriptors are based upon information processing codes, the present model adopts an approach where the descriptors are more directly task based. The verbs describe the cognitive activities undertaken by the pilot. Originally, nine verbs describing pilot-related activities were generated (see Harris, 2019). Initial trials showed some inconsistency in the application of the verb descriptors, in particular the consistency of the application of verbs associated with checking/verifying system status. As a result, the initial set of verbs was reduced to just five to increase coding reliability and reduce complexity. The final set of verbs to describe the pilot's tasks were:

- **Controlling** (flight path; systems): activities relating to the position and speed of the aircraft in three-dimensional space. Control of aircraft systems, including controls such as flaps, speed brakes, gear, etc.
- Information **Gathering** and **Monitoring**: state of systems; position of aircraft, etc. using aircraft instrumentation, specifically activities relating to actively obtaining data and information (including data about the external environment). Specifically, NOT controlling.
- **Communicating** and **Coordinating** (intention), including the coordination of people and equipment.
- Strategic **Planning** (of flight path; crew activities) and **Deciding** (choosing) between potential courses of action.
- Data **Verifying** and **Checking**: data; information; procedures; checklists and checking these data against plans and intentions.

Each task on the time is characterised using these verbs and the difficulty evaluated using a simple four-point scale (Hi - 5; Med - 3; Lo – 1; or Not Applicable - 0). Ratings are completed by a subject matter expert and are considered independently of either the task context or other concurrent tasks. As far as possible each task should be described by just a single verb, however some may involve two (or more) verbs, for example, many control tasks also require attention to the aircraft instrumentation. The individual difficulty scores are simply summed to provide a workload influence rating (see Table 1).

**Table 1** **Table for deriving workload input load estimates: Ratings made using a four-point scale (Hi - 5; Med - 3; Lo – 1; or Not Applicable - 0). Includes MRT Task category.**

| **Dimension** | **Rating** |
|---|---|
| Controlling | |
| Information Gathering and Monitoring | |
| Communicating and Coordinating | |
| Planning and Deciding | |
| Verifying and Checking | |
| **TOTAL** | |
| **MRT Task Category** | |

## Allocation of bottom level task to cognitive category of task

Each of bottom level task is also assigned to one category that best describes the cognitive nature of the task based upon Wickens' (2002) task characterisations: VS = Visual Spatial task; VV = Visual Verbal; AS = Auditory Spatial; AV Auditory Verbal; CS = Cognitive Spatial; CV = Cognitive Verbal; RS = Response Spatial; RV = Response Verbal: See Table 2. MRT makes certain predictions about performance. For example, performance of two simultaneous tasks will be better if one is task presented visually and the other is presented via the auditory channel, rather than both using the same modality. There is also a cognitive overhead over and above the individual workload contributions from each of the tasks resulting from two tasks being undertaken simultaneously.

In initial trials of the predictive workload modelling method, issues were identified in the consistency of the application of the definitions in the task tables and in particular the consistency of the categorisation of tasks associated with checking/verifying system status and the application of some aspects of Wickens' task categorisations (e.g. perception Vs cognition). Furthermore, there was also some difficulty assigning an appropriate information processing code. In the first revision of the model (reported in LPA-IADP-WP3.1.5—12), to aid in categorisation, any task that was not designated as 'verbal' was (by definition) codified as 'spatial', hence spatial tasks were designated as being 'non-verbal'. A similar approach was used by Ferris & Sarter (2010). In this instance codes were characterised as being either 'spatial' (data extracted from task relevant stimuli relating to spatial relationships) or 'non-spatial' (the data had symbolic meaning or had no spatial properties). However, in the latter instance 'symbolic meaning' or data that has 'no spatial properties' may themselves constitute two classifications of data. Not all 'non-verbal' tasks are necessarily 'spatial' tasks, for example the verification of the configuration of a system, checking a course deviation or verifying that thrust levers close under auto-thrust command. While these are certainly non-verbal, they are not truly spatial within Wickens' definitions. Furthermore, Kirby (1994) pointed out that simply because data are presented in a spatial format does not mean that the cognitive tasks involved in the processing of those data are solely of a spatial nature. Examples of this would include the incorporation of height/altitude data (digitally) on a 2-dimensional map containing spatial data. Memorisation of waypoints or route planning will also involve verbal aspects of what may initially be regarded as a spatial task. These shortcomings identified a requirement for further task categories to be defined relating to the nature of information processing undertaken during the perceptual and cognitive stages. Hence, a third processing code dimension was incorporated, entitled: 'symbolic'. Symbolic tasks may be visual in nature (particularly at the perceptual processing stage); auditory, for example non-verbal auditory warnings, or cognitive (reasoning and thinking): see Table 3.

## Task competition matrix

In addition to the complexity of tasks, scheduling and sharing also determines workload. The influence on workload of two tasks occurring simultaneously is incorporated into the predictive workload by specifically examining the cognitive nature of these tasks. The cognitive load on the pilot will partly be determined by competition for cognitive resources between tasks if two (or more) tasks need to be undertaken by the pilot simultaneously. Cognitive resource theories (e.g. Wickens, 1984; Wickens, 2002) hypothesise distinct pools of information processing resources. MRT (Wickens, 1984) determines that some simultaneous tasks can be shared more successfully than others. If both tasks share the same information processing resource (e.g. two verbal tasks, or two visual tasks) task performance is compromised to a greater extent than when they occupy different resources.

Wickens (1984; 2002) sub-divided information processing into three basic stages: perceptual encoding (what is it, where is it, what does it do, etc.); central processing (what shall I do about it) and responding (doing something about it, if required)? These processing stages are also categorised to incorporate input code and are described in Tables 2 and 3. Tasks undertaken simultaneously which place demands on the same underlying cognitive resource exhibit greatest task conflict, and hence result in higher workload. When two tasks are undertaken simultaneously there is also an information processing overhead attributable to the tasks competing for cognitive resources over and above the demands of the individual tasks. This is reflected in the modelling process.

**Table 2. Cognitive task categories (from Wickens, 2002).**

| Cognitive Category | Code | Processing Stage |
|---|---|---|
| **Visual Spatial** | VS | **Perceptual**: linked to the receiving of information from the aircraft or environment, for example reading; listening (spoken word); maps; visual information from the environment |
| **Visual Verbal** | VV | |
| **Auditory Spatial** | AS | |
| **Auditory Verbal** | AV | |
| **Cognitive Spatial** | CS | **Reasoning**: Either non-verbally or verbally/numerically |
| **Cognitive Verbal** | CV | |
| **Response Spatial** | RS | **Responding:** either via control systems or by voice (spoken word) |
| **Response Verbal** | RV | |

**Table 3. Additional cognitive task categories to incorporate symbolic data processing.**

| Cognitive Category | Code | Processing Stage |
|---|---|---|
| **Visual Symbolic** | VSym | **Perceptual**: linked to the receiving of information from the aircraft for example warning lights and sounds. |
| **Auditory Symbolic** | ASym | |
| **Cognitive Symbolic** | CSym | **Reasoning**: Either non-verbally or verbally/numerically |

These additional task categories required Wickens' original task conflict matrix to be extended to accommodate them. For the initial task conflict estimates, it was assumed that symbolic tasks would not conflict to the same degree as either the corresponding verbal or spatial tasks, however two symbolic tasks sharing the same modality and code would be equally as incompatible (see Table 4).

**Table 4. Original task conflict matrix was based upon the three primary dimensions of MRT: VS = Visual Spatial task; VV = Visual Verbal; AS = Auditory Spatial; AV Auditory Verbal; CS = Cognitive Spatial; CV = Cognitive Verbal; RS = Response Spatial; RV = Response Verbal. Numbers in the table indicate the degree of task conflict (from 1 to 2, where 2 = maximum conflict). Adapted from Wickens (2002). The revised matrix also incorporates the new categories of Visual Symbolic (VSym); Auditory Symbolic (ASym) and Cognitive Symbolic (CSym).**

**TASK 1 RESOURCES**

| **TASK 2 RESOURCES** | Perceptual | | | | | | Cognitive | | | Response | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *VS* | *VV* | *VSym* | *AS* | *AV* | *ASym* | *CS* | *CV* | *CSym* | *RS* | *RV* |
| VS | 1.8 | 1.6 | 1.4 | 1.6 | 1.4 | 1.2 | 1.7 | 1.5 | 1.3 | 1.4 | 1.2 |
| VV | | 1.8 | 1.6 | 1.4 | 1.6 | 1.4 | 1.5 | 1.7 | 1.5 | 1.2 | 1.4 |
| VSym | | | 1.8 | 1.4 | 1.2 | 1.6 | 1.3 | 1.5 | 1.7 | 1.1 | 1.3 |
| AS | | | | 1.8 | 1.4 | 1.2 | 1.7 | 1.5 | 1.3 | 1.4 | 1.2 |
| AV | | | | | 1.8 | 1.6 | 1.5 | 1.7 | 1.5 | 1.2 | 1.4 |
| Asym | | | | | | 1.8 | 1.3 | 1.5 | 1.5 | 1.1 | 1.3 |
| CS | | | | | | | 1.8 | 1.6 | 1.4 | 1.6 | 1.4 |
| CV | | | | | | | | 1.8 | 1.6 | 1.4 | 1.6 |
| CSym | | | | | | | | | 1.8 | 1.3 | 1.5 |
| RS | | | | | | | | | | 1.8 | 1.6 |
| RV | | | | | | | | | | | 2.0 |

## Characterisation of the task environment

All tasks take place within an environment (meteorological; airspace; air traffic, etc.). The demands of the environment partially determine the pace, difficulty and complexity of the pilot's tasks. Flight deck equipment can moderate some of the task demands imposed by the operational environment (but not all).

A list of generic environmental characteristics that may be applied to task environment as a whole which may impact on pilot workload was developed. These characteristics are described in the form of seven antagonistic noun pairs:

- **Simplicity/Complexity** – how difficult is the task; how many task components are there with how many potential interactions?
- **Predictability/Stability** – is the task environment stable and predictable, or could a number of unexpected things happen at short notice?
- **Visibility/Invisibility** – can the necessary task factors be viewed directly, or can they only be perceived using equipment on the aircraft (e.g. radar)?
- **Clarity/Ambiguity** – is it obvious what demands and/or risks are being imposed from the operational environment?
- **Velocity (Rapidity)** – how quickly are the environmental demands changing/evolving? Does this determine the required pace of the task?
- **Controllability/Uncontrollability** – are the environmental demands controllable from the flight deck in any way (e.g. requesting alternative ATC instructions)?
- **Accuracy/Inaccuracy –** what degree of accuracy of performance is required to complete the task successfully?

Furthermore, allowance in the modelling process is also provided to alter environmental demands during a scenario irrespective of task demands (e.g. weather may deteriorate as the aircraft descends; flight path accuracy demands may change with respect to distance from the runway, etc.).

Ratings based upon these noun pairs supplement the task analysis components to describe the workload requirements resulting from the operational task environment. This serves to further determine the amount and complexity of the task input load. The degree to which various environmental characteristics contribute to the input load of sub-tasks is again determined by expert judgement. Each environmental characteristic influencing workload demands in a particular scenario is rated using a simple three-point scale and entered into a table:

- 0 – no effect on pilot workload.
- +1 – potentially increases workload demands on the pilot slightly.
- +3– potentially increases workload demands on the pilot significantly.

See Table 5.

These ratings are then summed across all components to produce an environmental workload influence score. The environmental influence score may be calculated just once for a particular task/scenario or be revised as demands change.

**Table 5 Table for deriving environmental influence estimates on input workload estimates.**

| Dimension | Rating |
|---|---|
| Simplicity/Complexity | |
| Accuracy/Inaccuracy | |
| Predictability/Stability | |
| Visibility/Invisibility | |
| Clarity/Ambiguity | |
| Velocity (Rapidity) | |
| Controllability/Uncontrollability | |
| **TOTAL Environmental Influence** | |

## Derivation and presentation of task workload estimate

For each task (or task pair when two tasks occur simultaneously) the predicted workload is derived from the following formulae:

Equation 1: Single task predicted workload estimate

$$\left(\text{Workload input load estimates}\right) * 1.5 + \text{Environmental influences on workload} = \text{Predicted Workload}$$

Equation 2: Dual task predicted workload estimate

$$\left[\left(\text{Workload input load estimates task 1} + \text{Workload input load estimates task 2}\right) * \text{Task conflict modifier value}\right] + \text{Environmental influences on workload} = \text{Predicted Workload}$$

The estimates from pilot trials exhibited a tendency to underestimate workload predictions when just a single task was being undertaken. To reflect the additional attention that a pilot undertaking just a single task would be able to allocate to a single task, in these circumstances the workload estimate was increased by 50% (see equation 1).

The task-by-task workload predictions are presented graphically on a timeline, with time on the x-axis and overall workload on the y-axis.

## Workload Model Predictions

### Method

To assess the predictions from the modelling approach, workload estimates were derived for the pilot flying (PF) position during of an extended flight scenario encompassing approach, landing and roll-out. The approach and landing analyses were undertaken for a manual approach (with autothrust engaged) with a 0 (zero) feet decision height and 0 (zero) feet runway visual range performed in a Boeing 777.

The comparative assessment process proceeded in the following stages:

- Timeline task analyses of the flight scenario and development of task timeline(s)
    - Identification of the nature of the cognitive tasks
    - Task environment assessment
- Subjective assessment of workload experienced using the Bedford workload scale (Ellis & Roscoe, 1982) for comparison with the predictive workload estimates.
- Calculation of predictive workload estimates.

The task timeline analysis was constructed from data derived from Verbal Protocol Analyses (see Stanton *et al*., 2013) supplemented by further critical decision-making analyses (Crandall *et al*., 2006). Task time estimates were derived from height Vs speed calculations during the approach and validated from review of on-line approach videos.

## Subject Matter Experts

A fully qualified, senior airline Captain with Boeing 777 type-rating provided the principal data for the timeline task analyses. This was supplemented by data from two further commercial pilots holding ATPLs – Air Transport Pilot's Licence - (including dual qualified ATPL pilots with Test Pilot qualifications) who also provided workload estimates.

### Timeline Task Analysis Method

A timeline task analysis was derived from bottom level tasks (Kirwan & Ainsworth, 1992; Stanton *et al*., 2013). Initially a high-level outline of the flight task under consideration was elicited, in particular specifying the beginning and end points for tasks. This was subsequently de-composed into logical sub-tasks undertaken by the PF and if necessary, these sub-tasks themselves were then subject to further de-composition based upon a classic input-process-output model to help identify information sources, decisions, communications (on board and with Air Traffic Control, if necessary) and control outputs made. These were then arranged upon a timeline with tasks being undertaken in parallel identified. Typical timings of the average duration for all tasks and sub-tasks were estimated.

### Subjective assessment of workload

Subject matter experts were asked to make workload assessments using a simple, uni-dimensional Modified Copper-Harper scale (Bedford scale: Ellis & Roscoe, 1982). These measures were elicited for each of the major tasks during takeoff and landing and approach. To increase the sensitivity of the Bedford Scale, especially at the lower end, participants were allowed to make half ratings. Bedford ratings were tied to specific actions in the timeline task analysis to allow for the workload ratings to be synchronised with the pilot's activities.

### Identification of the nature of the cognitive tasks

The underlying cognitive activities in each task were identified using the set of descriptors and rated using the simple four-point scale described Table 1. The cognitive nature of each sub-task was classified using the categories described in Tables 2 and 3. Ratings for each bottom level task were completed independent of either the task context or any concurrent tasks.

### Task environment assessment

The overall task environment's contribution to workload was assessed with reference to the list of antagonistic noun pairs described in Table 5 and rated using the simple three-point scale. Ratings were summed to produce an overall environmental influence score.

Categorisation of sub-tasks; sub-task contribution to workload and the assessment of the contribution of the task environment to workload was performed by a Subject Matter Expert independent of those providing the data for the task analyses and Bedford scale workload ratings.

## Predicted Workload

A section of the Excel data file describing the initial part of the approach and landing flight segment is presented in Figure 1. Each bottom level sub-task is described, given an approximate duration; assigned to a cognitive task category and each sub-task evaluated for its contribution to workload (see exemplar in Table 6). Environmental contributions to workload are evaluated as required (these may change during the task) – see exemplar in Table 7. When two tasks are undertaken simultaneously the task conflict multiplier is assigned with reference to the matrix presented in Table 4.

| Task Description (1) | Task Description (2) | Time (seconds) | Cumulative time | Task 1 Workload Input | MRT Category Task 1 | Task 2 Workload Input | MRT Category Task 2 | Task Conflict Weight | Environmental Influence | Total Predicted Workload | Bedford Rating | Smoothed Total Workload (Environment Workload Additive) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 5 | 32 | | | 6 | RS | | 3 | 3 | 3 | 0 |
| Verifies a/c thrust closes, speed decreases to flap up speed | | 2 | 34 | 1 | Csym | 6 | RS | 1.3 | 3 | 12.1 | 3 | 0 |
| | | 5 | 39 | | | 6 | RS | | 3 | 3 | 3 | 0 |
| Verifies auto-throttle maintains 250kts | | 2 | 41 | 1 | Csym | 6 | RS | 1.3 | 3 | 12.1 | 3 | 0 |
| | | 15 | 56 | | | 6 | RS | | 3 | 3 | 3 | 0 |
| Actions appropriate heading and speed adjustments as required by ATC | | 5 | 61 | 2 | RS | 6 | RS | 1.8 | 3 | 17.4 | 3 | 0 |
| | | 15 | 76 | | | 6 | RS | | 3 | 3 | 3 | 9.12 |
| Abstracts and predicts expected height/distance to go/speed ATC requirements | | 5 | 81 | 3 | CS | 6 | RS | 1.6 | 3 | 17.4 | 3.5 | 8.96 |
| | | 10 | 91 | | | 6 | RS | | 3 | 3 | 3.5 | 8.96 |
| When appropriate, calls for FLAP UP SPEED | | 2 | 93 | 3 | CV | 6 | RS | 1.4 | 3 | 15.6 | 3.5 | 8.96 |
| | Manual Control flight path of aircraft in X and Y axes (to within 0.5 dots) | 10 | 103 | | | 6 | RS | | 3 | 3 | 3.5 | 8.96 |
| Verifies a/c thrust closes, speed decreases to flap up speed | | 2 | 105 | 1 | Vsym | 6 | RS | 1.1 | 3 | 10.7 | 3 | 8.82 |
| | | 10 | 115 | | | 6 | RS | | 3 | 3 | 3 | 8.82 |
| Verifies auto-throttle maintains flap up speed. | | 2 | 117 | 1 | VSym | 6 | RS | 1.1 | 3 | 10.7 | 3 | 8.68 |
| | | 10 | 127 | | | 6 | RS | | 3 | 3 | 3 | 8.68 |
| Verifies transition height, sets QNH, crosscheck the altimeters | | 10 | 137 | 3 | CSym | 6 | RS | 1.3 | 3 | 14.7 | 3 | 8.41 |
| | | 10 | 147 | | | 6 | RS | | 3 | 3 | 3 | 8.41 |

eet2 70%

**Figure 1** **Section of Excel™ predictive spreadsheet describing sub-tasks; duration; cognitive task category; sub-task contribution to workload; environmental contribution to workload; task conflict modifier; predicted workload value and Bedford workload rating.**

Output from the workload predictions plotted against corresponding Bedford scale workload scores are presented in Figures 2 and 3. For clarity, the approach and landing, and landing roll-out segments and depicted separately, however all analyses were conducted on the aggregated data.

**Table 6** **Exemplar of sub-task contribution to predicted workload: Manual Control flight path of aircraft in X and Y axes (to within 0.25 dots) during middle stages of approach.**

| **Dimension** | **Rating** |
| --- | --- |
| Controlling | 5 |
| Visual Information Gathering and Monitoring | 3 |
| Communicating and Coordinating | 0 |
| Strategic Planning | 0 |
| Data Verifying | 0 |
| **TOTAL Workload Influence** | **8** |
| **MRT Task Category** | **RS** |

**Table 7** **Exemplar of environmental influence estimates on input workload estimates: Environmental influences on workload for undertaking a manual ILS approach during daylight in very light turbulence (early stages of approach).**

| **Dimension** | **Rating** |
| --- | --- |
| Simplicity/Complexity | 0 |
| Accuracy/Inaccuracy | 0 |
| Predictability/Stability | +1 |
| Visibility/Invisibility | +1 |
| Clarity/Ambiguity | 0 |
| Velocity (Rapidity) | +1 |
| Controllability/Uncontrollability | 0 |
| **TOTAL Environmental Influence** | **+3** |

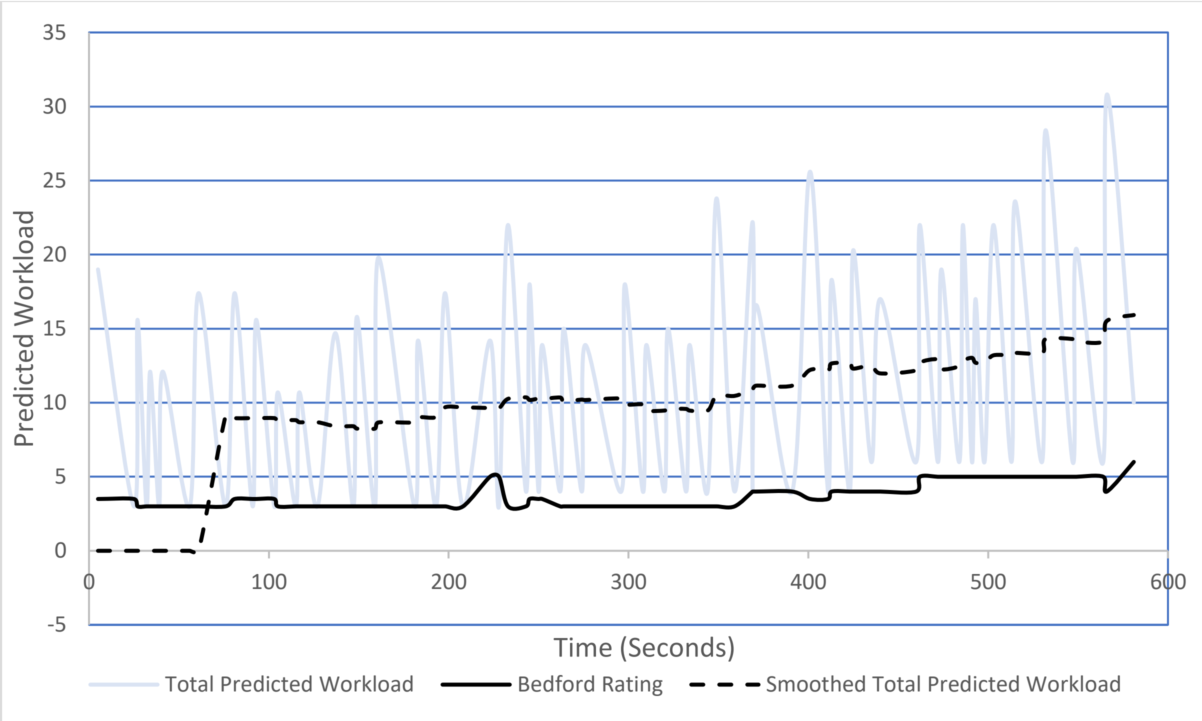


**Figure 2** **Manual approach segment (under autothrust) predictive workload model output (presented both as individual task predicted workload ratings and smoothed workload ratings) plotted against corresponding Bedford scale workload rating data.**

As workload predictions were made on a task-by-task basis, to produce a more meaningful workload profile (see Figures 2 and 3) data were smoothed and depicted as a rolling average over ten sub-tasks. In general, the predicted workload values were approximately four times higher than the corresponding Bedford scale values (see Figures 2 and 3). Smoothed predicted workload values approximately mirrored those of the corresponding Bedford scale scores.

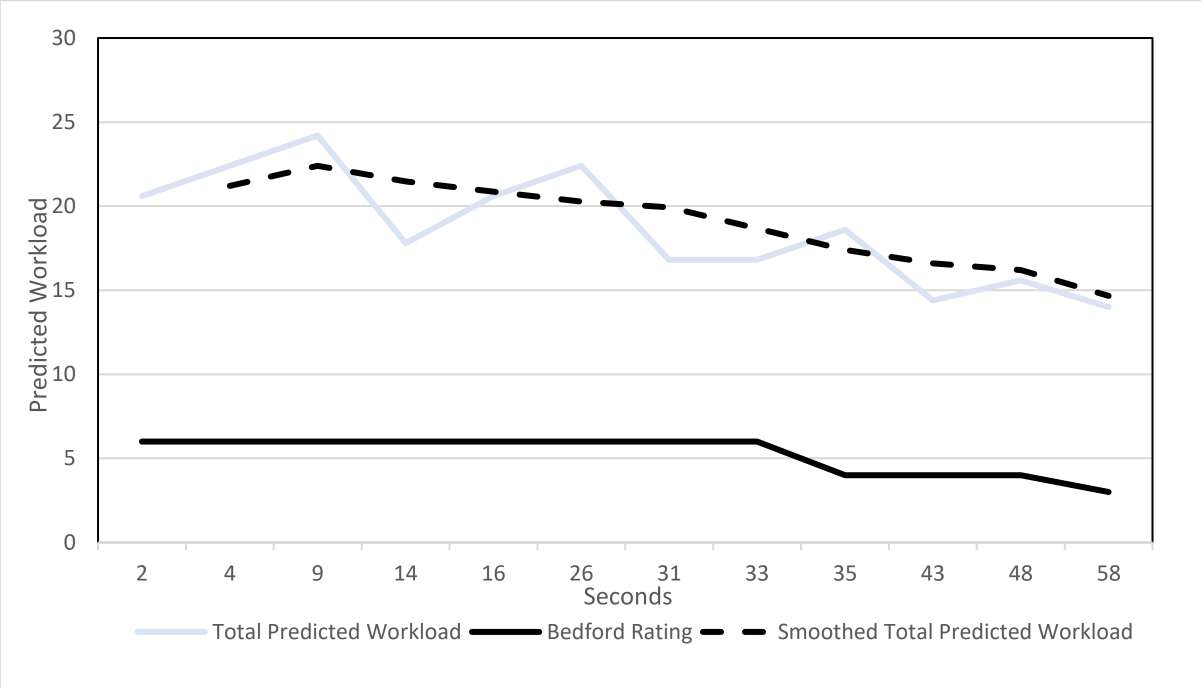


**Figure 3** **Roll out flight segment predictive workload model output (presented both as individual task predicted workload ratings and smoothed workload ratings) plotted against corresponding Bedford workload rating data.**

## Discussion

The work describes the preliminary stages of the development and proof of concept of a predictive workload. The emphasis in the development of the technique was on producing a method that was easy to use in the early stages of design and which could be implemented without recourse to complex software. The predictive modelling technique developed produced workload estimates that corresponded with workload estimates elicited using the Bedford scale. Furthermore, the workload predictions were elicited from straightforward task-analysis timeline data which should be available in the basic design documentation for most new systems.

Many approaches to the validation of predictive workload models have been based upon assessing the relationship with performance. Sarno & Wickens (1992) observed that across the predictive models of workload, between 61% and 77% of the variance in performance was accounted for. Of the models assessed, TLAP (Parks & Boucek, 1989) produced the best performance.

It has previously been noted, though, that the relationship between performance and mental workload is not a straightforward one, and is often mediated by other factors (Mansikka, *et al.*, 2019). Wickens *et al*. (2013) reported much lower correlations of predicted workload with measures of cognitive workload elicited. Average correlations were in the range of 0.21-0.61, accounting for between approximately 4.5 - 37% of the variance. The results from the initial proof of concept evaluations in this study show good correspondence between the predicted workload scores and Bedford workload scores. Comparing predictions against a known workload measure is essential for establishing the predictive utility of the approach.

The approach described in this paper has the advantage of producing a rolling prediction of workload which allows the identification of potential workload peaks at an early stage in the design process, thereby allowing for early re-design of equipment and/or re-scheduling of tasks (predictive workload redlines – see Rueb, *et al.,* 1994).

The predictive workload technique described is currently restricted in its potential generalisability as a result of only having been applied to a limited number of flight scenarios and aircraft types, and also with a small sample size of participants. At sustained periods of very low workload, other measurement techniques to validate predictions against may be required (Braby, *et al*., 1993). Work is underway to further develop the approach to establish stable regression coefficients for use in the next refinement of the predictive workload model calibrated against Bedford scale estimates (Ellis & Roscoe, 1982), to establish its generalisability in a wider range of flight phases and to further assess the reliability of the task categorisation and coding phases in the modelling process. It is proposed to undertake formal model validation (including statistical analyses) which will include a wider range of normal, non-normal and emergency scenarios which impose higher levels of task demand. Future scenarios will also incorporate the predicted effects of new flight deck equipment to estimate their contributions to alleviating crew workload. The present, early iteration of the model produces relative estimates of workload. However, the workload experienced in a given context will also vary from person to person dependent upon individual differences (such as experience) and contextual factors (such as fatigue). The next iteration of the model will specifically include factors to incorporate these issues. Future work could also include collecting Bedford workload data in real-time during simulated flight scenarios. The use of simulated trials will allow better replication and manipulation of task and environmental factors and avoid any potential safety issues associated with collecting data during high workload flight phases.

## Ethics

Work was conducted under Coventry University ethics certificate No. P93317.

## Note

Steve Scott is now at School of Engineering and Built Environment (Aviation), Griffith University, Queensland, Australia

## References

Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A.D. & Hitch, G.J. (1974). 'Working memory'. In, G.H. Bower (Ed) *The psychology of learning and motivation, volume 8*. New York NY: Academic Press, pp. 47–89. doi:10.1016/S0079-7421(08)60452-1

Braby, C.D., Harris, D. & Muir, H.C. (1993) A psychophysiological approach to the assessment of work underload. *Ergonomics, 36*(9), 1035-1042. doi: 10.1080/00140139308967975

Crandall, B., Klein, G.A., & Hoffman, R.R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: MIT Press.

Davies, A.K., Tomoszeck, A., Hicks, M.R. & White, J. (1995). AWAS (Aircrew Workload Assessment System): Issues of theory implementation and validation. In, R. Fuller, N. Johnston & N. McDonald (Eds) *Human factors in aviation operations*. Aldershot: Avebury, pp. 319-324.

Ellis, G.A. & Roscoe, A.H. (1982). The airline pilot's view of flightdeck workload: a preliminary study using a questionnaire. *Technical Memorandum FS(B) 465*. Bedford: Royal Aircraft Establishment.

Ferris, T.K. & Sarter, N. (2010). When content matters: the role of processing code in tactile display design. *IEEE Transactions on Haptics, 3*(3), 199-210. doi: 10.1109/TOH.2010.10

Hamilton, D. B., & Bierbaum, C. R. (1990). Task analysis workload (TAWL): A methodology for predicting operator workload. In, *Proceedings of the Human Factors Society 34th Annual Meeting*. Santa Monica, CA: Human Factors Society, pp. 1117–1121.

Harris, D. (2011). *Human performance on the flight deck*. Aldershot: Ashgate.

Harris, D. (2019). LPA-IADP-WP3.1.5-11 *Report developing a predictive model of workload*. Report to EU CleanSky2 Joint Undertaking.

Huddlestone, J.A., & Stanton, N.A. (2016). New graphical and text-based notations for representing task decomposition hierarchies: towards improving the usability of an Ergonomics method. *Theoretical Issues in Ergonomics Science, 17*(5-6), 588-606. doi: 10.1080/146

Johannsen, G. (1979). Workload and workload measurement. In, N. Moray (Ed) Mental Workload: Its Theory and Measurement (pp. 3-11). New York NY: Plenum.

Kirby. J.R. (1994). Comprehending and using maps: are there two modes of map processing? In, W. Schnotz & R. W. Kulhavy (Eds) *Advances in Psychology, Volume 108*. Amsterdam, NL: Elsevier Science B.V., pp. 63-76. doi: 10.1016/S0166-4115(09)60108-4

Kirwan, B. & Ainsworth, L. (1992). *A guide to task analysis.* London: Taylor & Francis.

Mansikka, H., Virtanen, K., & Harris, D. (2019). The dissociation between mental workload, performance and task awareness in fast jet pilots. *IEEE Transactions on Human-Machine Systems, 49*(1) 1-9. doi: 10.1109/THMS.2018.2874186

McMillan, G.R., Beevis, D., Stein, W., Strub, N.H., Salas, E., Sutton, R., & Reynolds, K.C. (1991). *A directory of human performance models for system design (AC/243(Panel 8) TR/1).* Brussels: NATO.

Mitchell, D.K. (2000). *Mental workload and ARL workload modeling tools. army research laboratory report (ARL-TN-161).* Aberdeen Proving Ground, MD: Army Research Laboratory.

Mitchell, D.K. & Samms, C. (2010). An analytical approach for predicting soldier workload and performance using human performance modelling. In, M. Barnes & F. Jentsch (Eds), *Human-robot interactions in future military operations*. Aldershot, UK: Ashgate Publishing, pp.125–141.

North R.A. & Riley V.A. (1989). W/INDEX: A predictive model of operator workload. In, G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton & L. van Breda (Eds) *Applications of human performance models to system design*. Boston, MA: Springer, pp. 81-89. doi: 10.1007/978-1-4757-9244-7_6

Parks, D.L. & Boucek, G.P. (1989). Workload prediction, diagnosis and continuing challenges. In, G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, & L.V. Breda (Eds). *Applications of human performance models to system design (Defense Research Series, Vol. 2).* New York City, NY: Plenum Press, pp. 47-63. doi: 10.1007/978-1-4757-9244-7_4

Pressman, R.S. (1992). *Software engineering: a practitioner's approach*. New York: McGraw-Hill.

Reuzeau, F. & Nibbelke, R. (2004). Flight deck design process. In D. Harris (Ed). *Human factors for civil flight deck design.* Aldershot: Ashgate, pp. 33-54.

Rueb, J.D., Vidulich, M.A. & Hassoun, J.A. (1994). Use of workload redlines: a KC-135 crew-reduction application. *International Journal of Aviation Psychology, 4*(1), 47-64. doi: 10.1207/s15327108ijap0401_3

Rusnock, C.F., Borghetti, B.J., & McQuaid, I. (2015). Objective-analytical measures of workload: the third pillar of workload triangulation? In, D.D. Schmorrow & C.M. Fidopiastis

(Eds.) *Foundations of augmented cognition*. Berlin: Springer-Verlag, pp. 124-135. doi: 10.1007/978-3-319-20816-9_13

Rusnock, C.F. & Geiger, C.D. (2016). Designing adaptive systems: selecting an invoking threshold to improve human performance. *International Journal of Human Factors and Ergonomics*, *4*(3/4), 292–315. doi: 10.1504/IJHFE.2016.083503

Sarno, K.J. & Wickens, C.D. (1992). Predictive workload models and multiple-task performance. In, *Proceedings of the Human Factors Society Annual Meeting, Volume: 36(1),* 12-16. doi: 10.1177/154193129203600105

Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., Baber, C. & Jenkins, D.P. (2013). *Human factors methods: a practical guide for engineering and design (2nd Edition)*. London: Taylor & Francis.

Wickens, C.D. (1984). Processing resources in attention. In, R. Parasuraman & R. Davies (Eds) *Varieties of attention*. New York NY: Academic Press, pp. 63–101.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 2*(2)*,* 150-177. doi: 10.1080/14639220210123806

Wickens, C.D., Sebok, A., Keller, J., Peters, S., Small, R., Hutchins, S., Algarín, L., Gore, B.F., Hooey, B.L. & Foyle, D.C. (2013). *Modeling and evaluating pilot performance in nextgen: review of and recommendations regarding pilot modeling efforts, architectures, and validation studies. NASA/TM—2013–216504.* Ames Research Center, Moffett Field, CA: National Aeronautics and Space Administration.

Xie, B. & Salvendy, G. (2000) Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work and Stress, 14*(1), 74-99. doi: 10.1080/026783700417249

Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of Performance and Subjective Measures of Workload. *Human Factors*, *30*(1), 111-120. doi: 10.1177/001872088803000110