

# Differentiating Cognitive Complexity and Cognitive Load in High and Low Demand Flight Simulation Tasks

Jemma Harris<sup>1</sup>(✉), Mark Wiggins<sup>2</sup>, Ben Morrison<sup>1</sup>,  
and Natalie Morrison<sup>1</sup>

<sup>1</sup> Australian College of Applied Psychology, Sydney, Australia  
jemma.harris@acap.edu.au

<sup>2</sup> Macquarie University, Sydney, Australia

**Abstract.** In the contemporary workplace, the design of interfaces has a significant impact on the cognitive demands experienced by operators. Previous approaches to the assessment of these designs have relied on measures of cognitive load to infer the level of cognitive demand imposed. Assessments of cognitive complexity may offer a complimentary measure of the demands of the task as they take into account the inherent nature of the task, rather than idiosyncrasies of the operator. Two studies are reported that examined the information acquisition behavior of pilots in response to a series of simulated flight sequences involving different levels of cognitive complexity. Information acquisition was recorded using an eye tracker. Taken together, the results suggest that assessments of the complexity of a task should be employed as a benchmark in task assessment.

**Keywords:** Cognitive complexity · Cognitive load · Eye movement · System design · Aviation

## 1 Introduction

Contemporary approaches to task assessment generally adopt a user-referent perspective in which a range of different users interact with a task and their perceptions and performance are recorded and distributed [1, 2]. The intention is to establish the limits to performance amongst prospective users. In doing so, strategies can be developed that either alter the nature of the task and/or restrict the task to users with particular capabilities or experience. This will pose implication for task design.

The design of tasks around the limitations of users reflects an underlying assumption that the performance of a task imposes a mental or physical load [3, 4]. Depending upon the nature of the task, the load will accumulate to a point where performance begins to deteriorate. This is the point at which the demands of the task exceed the capabilities of the user. At a cognitive level, the demands imposed by a task are presumed to be reflected in perceptions of mental or cognitive load which, in combination with the motivation to maintain a specified level of performance, is associated with the effort invested in the performance of task [5]. Perceptions of cognitive load will be further

moderated by the experience of the operator to a point where, arguably, it is possible to achieve relatively high levels of performance, with relatively little effort and a resultant lower perception of cognitive load [5].

From the system designer's perspective, the difficulty with an assessment based on cognitive load lies in anticipating the minimum requirements necessary to ensure that the least able operator is capable of undertaking the task successfully. This requires some understanding of the complexity of the task, since this represents the minimum level of processing necessary to undertake the task, irrespective of experience or performance shaping factors [6]. While experts might be more rapid and more accurate in their capacity to manage the complexity of the task, the complexity associated with the task remains. Therefore, while perceptions of cognitive load may be variable, depending upon the skills and the motivation of the operator, cognitive complexity is characterized by the nature task itself and therefore, can be established a priori.

This chapter explains how the cognitive complexity of a task can be operationalized as a means of standardizing the cognitive aspects of product development in the future. The chapter will start with a discussion on the notion of cognitive complexity and the relationship between the cognitive and behavioural elements of task performance. Sections 3 and 4 will then present empirical data from two studies where cognitive complexity and information acquisition were examined in an aviation context. The aviation focus represents a domain where the utility of usability testing requires a cognitive focus. Aviation further represents a context where the user experience might be generalized to other domains that evoke demands on information processing resources. A discussion on the outcomes of the empirical investigations will follow. Future research directions will then conclude this chapter.

## 2 The Notion of Cognitive Complexity

Wood [7] explains complexity as a combination of component, coordinative, and dynamic complexity. Component complexity refers to the number of distinct acts that need to be completed and the number of information cues that are required to be interpreted. Coordinative complexity comprises the form and strength of the association between task inputs and between inputs and task outcomes. Finally, the dynamic aspect of complexity includes the extent to which there are variations in the relationship, whether there are sudden changes in the relationships, and whether these changes are predictable.

The distinction between individual and task attributes in the context of task performance is further conceptualized by Woods [6]. Consistent with Wood [7], Woods [6] conceptualizes cognitive complexity as a combination of four indicators: (1) the time-constraint imposed, (2) the uncertainty associated with performing the task, (3) the inter-relationships between components, and (4) the risk associated with the activity.

In the context of cognitive complexity, the level of dynamism refers to the extent to which the information used to perform the task changes over time. The level of uncertainty is the extent to which an operator can rely on the information, while the interrelationship between components relates to the prospective impact of one task on the performance of other tasks [6, 8, 9]. Finally, the risk associated with a task refers to

the combination of the likelihood of performing the task correctly and the implications of performing the task incorrectly [6, 9, 10].

Both Woods [6] and Wood [7] argue that, independent of motivation and task experience, there is an association between the cognitive characteristics of a task and the demands on information processing resources. For example, increases in the rate at which task-related information changes (referred to as dynamism) will require an increased capacity to monitor, extract, and integrate information over shorter periods of time [1, 6]. Similarly, an increase in the number of components with which an element is associated will require an increased capacity to anticipate the consequences of events [6].

Where the indicators of system functioning become inaccurate, a level of uncertainty is generated that requires an assessment of the accuracy of the information, based on either previous experience or on other information available at the time [6, 8]. Finally, increases in the risk associated with an incorrect response to a task are associated with an increase in the demands for the management of affective responses that might reduce the capacity of working memory [1].

## 2.1 Cognitive Complexity and Task Performance

In establishing the relationship between cognitive complexity and task performance, Maynard and Hakel [4] distinguish objective and subjective conceptualizations of the construct. Their conceptualization of objective complexity broadly reflects the theoretical positions of Wood [7] and Woods [6] insofar as it is defined as the amount of information that is required to be integrated to optimize task performance. By contrast, their conceptualization of subjective complexity is more consistent with notions of cognitive load, since it is defined as the perception of the level of complexity associated with the performance of a task.

Like cognitive load, Maynard and Hakel [4] contend that the level of subjective task complexity reflects the impact of performance shaping factors such as motivation and experience, and moderates the effort ascribed to the performance of the task. Importantly, they found that levels of subjective and objective complexity, although related, contributed differently to performance on a managerial scheduling task. This suggests that it is both the subjective and objective demands of a task that contribute to task success.

Although subjective and objective complexity appear to contribute independently to task performance, the difficulty for system designers lies establishing systematically, the differences in the demands of the task. Where subjective complexity can be established through the administration of questionnaires, objective complexity needs to be established as an outcome of a task analysis. This process can be time-consuming and may, in and of itself, represent a subjective conceptualization of the demands of the task.

An alternative approach to a detailed task analysis involves an assessment of the process of information acquisition as indicative of the objective complexity of a task. Information acquisition and the behavioral strategies with which it is associated are well-established measures of underlying cognitive performance. First-order measures include the data arising from process tracing [11] and visual search [12, 13], while second-order measures include data from retrospective verbal protocols, self-reports [14, 15], and cognitive interviews [16, 17].

As a measure of visual search, fixation duration has been used to draw inferences to cognitive processing, albeit with inconsistent outcomes. For instance, in Tole, Harris, Stephens, and Ephrath [18] average fixation time increased when mental workload was raised, yet Rivercourt, Kurperus, Post and Mulder [19] observed a decline in fixation duration with increasing levels of task complexity. These discrepant outcomes might be explained by the inherent value of average values, and the time frame over which eye movements are analyzed.

The value of assessing the dispersion of fixation durations over time has been raised by Velichkovsky, Dornhoefer, Pannasch, and Unema [20]. In their study, the proportion of fixation durations occurring in five categories was calculated. In the first second following the release of a critical event, fixation durations increased significantly in the >601 ms category. This outcome could be interpreted as providing support for the notion that increases in task complexity are related to longer fixation durations. However, during the fifth second of the critical event, fixation durations were predominately occurring in the <151 ms category [20]. Therefore, it appears that the period in which fixation durations are analyzed is important for making inferences to cognitive processing.

### 3 Study 1: Task Complexity and Information Acquisition

Wiggins [21] proposes that, under conditions of uncertainty, the pattern of information acquisition during the diagnostic or situation assessment phase of a task reflects the underlying complexity of that task. Tasks that are objectively more complex will require the acquisition of greater amounts of information, with greater frequency, and within less time. By contrast, tasks that are less complex will require the acquisition of lesser amounts of information, with less frequency, and within a greater period of time.

Given that objective complexity is based on the intrinsic demands of the task, it should be possible, having completed the task, for users to rate the various dimensions of complexity, provided that questions relate to the underlying characteristics of the task, and not perceptions of the demands of those tasks. In doing so, designers could be offered a relatively inexpensive solution to the assessment of both subjective and objective complexity in the context of system design.

#### 3.1 Aims and Hypotheses

The present studies were designed to establish whether differences in the objective complexity of tasks are associated with differences in information acquisition during flight simulation; whether these differences are consistent with the effects implicit in Wiggins [21]; whether pilots' ratings of the elements of objective complexity relate to the intrinsic complexity of the tasks; and whether ratings of objective and subjective complexity discriminate between tasks of greater and lesser complexity.

Aviation was employed as a context for the research since the use of flight simulation provides a combination of ecological validity and experimental control. It also constrains the number of features available for problem resolution, and enables the

application of eye-tracking technology. Finally, pilots are required to record their experience in operating aircraft, thereby providing a relatively accurate indication of both the quality and the quantity of operational experience.

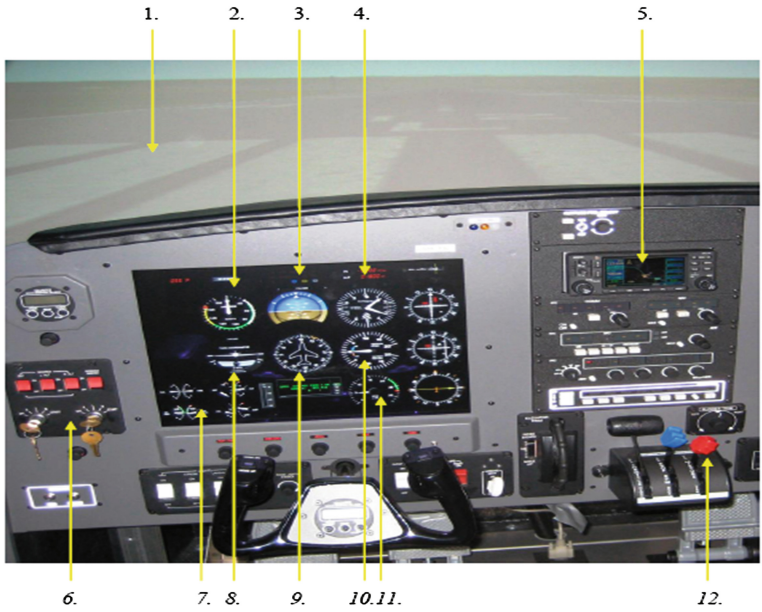
Study 1 involved the development of a series of flight simulated problem-solving scenarios that differed objectively in their complexity. On the basis of Wiggins [21], it was hypothesized that, in comparison to objectively complex scenarios, pilots' would access less information, with less frequency, and within more time during the less complex scenarios.

## 3.2 Method

**Subjects.** This study was approved by the UWS Human Research Ethics Committee and subject's gave their written informed consent to participate. The subjects comprised 41 general aviation pilots, of whom 38 were male and three were female. They ranged in age from 19 to 62 years, with a mean age of 34 years ( $SD = 13.19$ ). The subjects had accumulated between 70 and 8000 total flying hours ( $M = 937.80$  h,  $SD = 1587.10$ ), of which between 10 and 6300 were undertaken as pilot-in-command ( $M = 683.78$  h,  $SD = 1284.10$ ). The majority of subjects held a commercial or private pilots license, and were neither a flight instructor nor instrument-rated. The subjects represented a convenience sample recruited through a University-based aviation research register and through a number of flight training organizations. They were compensated \$40.00 for travel expenses.

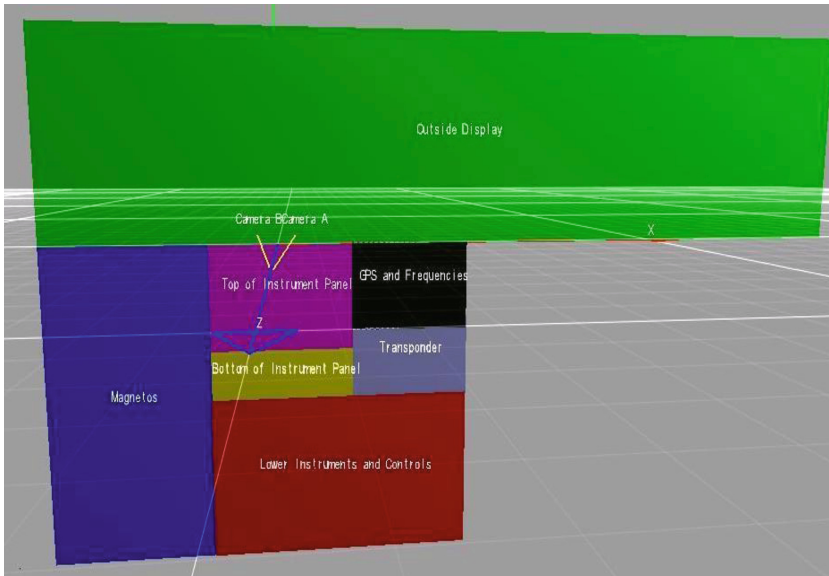
**Equipment and Materials.** The study was conducted using a simulated Cessna 172 aircraft operated on a Precision Flight Controls flight simulator, using the X-Plane 6.21™ program developed by Laminar Research Corporation. Figure 1 displays the layout of the flight simulator. The pilot was seated in a fibreglass cockpit and communicated with the experimenter through a headset. Through the instructor PC, the X-Plane program allowed for the airport, aircraft, weather, time of day, and instrument serviceability to be varied manually. It also contained a dynamic display of the aircraft's altitude, heading and airspeed, and allowed the experimenter to view the location of the aircraft with reference to nearby airfields.

Developed by Seeing Machines, FaceLAB™ Version 3.2 was used to record eye-gaze data. The system uses the input from two small video cameras (located on the flight simulator console) to determine gaze direction and fixation duration. A graphical model containing information on the size, orientation and spatial separation of real-world objects was created to represent the flight simulator environment. Within this graphical model, the flight simulator was divided into seven Areas of Interest (AOI, see Fig. 2). The individual instruments located on the instrument panel (e.g., altimeter) were not mapped onto the model, as a faceLAB AOI can be no smaller than  $32 \times 24$  cm in actual size and most of the individual flight simulator instruments were  $5 \times 10$  cm. Nevertheless, the model created for the present study could generate reliable information on the major AOIs that were accessed whilst managing the aircraft. The subjects' gaze interaction with the graphical model (e.g., point of fixation and fixation duration) was logged every 16.66 ms (sampling rate of 60 Hz). The 'xIFAT' add-in software, also developed by Seeing Machines, was used to analyze the gaze data.



Note. 1. outside, 2. airspeed indicator, 3. artificial horizon, 4. altimeter, 5. GPS, 6. magnetos, 7. engine instruments, 8. turn and balance coordinator, 9. directional indicator, 10. vertical speed indicator, 11. RPM, and 12. mixture.

**Fig. 1.** The geographic location of aircraft instruments in a simulated Cessna 172 aircraft operated on a Precision Flight Controls flight simulator.



**Fig. 2.** The seven areas of interest (AOI) created for the present study.

The flight sequences were designed in consultation with five subject-matter experts who were cognizant of Woods' [6] notion of cognitive complexity. The five subject-matter experts (SMEs) were asked to develop three flight sequences that differed systematically from straight and level flight on the basis of four elements of objective complexity, including: (1) the requirement for an accurate diagnosis, (2) the uncertainty associated with task, (3) the number of components required to reach a diagnosis, and (4) the risk associated with a misdiagnosis. The three flight sequences in successively increasing levels of objective complexity included an airspeed indicator failure, a right magneto failure (partial engine failure), and a low oil pressure indicator (the precursor to a total engine failure). The straight and level flight sequence acted as a baseline against which the assessments of performance in response to other flight sequences could be compared.

A questionnaire was administered to the subjects following the completion of each of the four flight events. The questionnaire comprised five questions relating to the objective complexity of the task and was employed as a manipulation check for the purposes of Study 1. Subjects responded to each question on a three-point scale. The summed scores across the five questions ranged from 5 to 15, with a higher score corresponding to a relatively greater level of objective complexity (see Table 1 for the post-flight sequence questionnaire).

**Table 1.** Post flight sequence questionnaire used in Study 1

Cognitive complexity dimension	Question/s
Dynamism	To what extent did the information that you used to perform the scenario change over time?
Uncertainty	To what extent could you rely on the information that you used to perform the scenario?
Interconnectedness among sub systems	How difficult was the scenario?
Risk	(a) How often did you make mistakes during the scenario? (b) How important was it to avoid errors during the scenario?

**Procedure.** Having provided written informed consent, the gaze tracker was calibrated and the subjects were instructed to complete a two-minute practice flight to become accustomed to the simulator. Following the practice flight, the experimenter explained the details of the study to the participant. They were instructed that they would be the pilot-in-command of a Cessna 172S aircraft and navigating under visual flight rules. They would complete a take-off, climb and cruise around Tamworth (located in New South Wales, Australia) airport. This airport was selected for its relative novelty to the subjects.

As a context for the scenario, subjects were asked to follow air traffic control instructions in undertaking a visual search for possible bushfires (forest fires). The subjects were naïve as to the exact nature of the study, and particularly the failure of



cockpit instruments. However, they were advised that an ‘intelligent aviation assistant’ would inform them of any abnormalities during the flight by announcing the event that had occurred. The use of an ‘intelligent aviation assistant’ was explicitly designed to ensure a consistent trigger for the onset of events. The experimenter acted as both air traffic controller and ‘intelligent aviation assistant’.

Four flight sequences were completed by the subjects over a period of 20.5 min. Following a period of time during which the participant responded to the failure, the ‘intelligent aviation assistant’ informed the subjects that the failure had been rectified and the flight had resumed. The straight and level flight sequence was always completed on commencement both to orientate the pilot to the task and to avoid priming. The presentation of the remaining flight sequences was counterbalanced.

### 3.3 Results

Prior to analysis, the data were screened to ensure that assumptions of normality were satisfactory. Five univariate outliers were identified and the data corrected using a square root transformation. The assumptions of normality were subsequently achieved for univariate and multivariate analyses. Data screening revealed no significant correlation between total hours of flight experience and each of the dependent variables, indicating that total hours of flight experience (as a measure of expertise) did not need to be included as a covariate in subsequent analyses. As a manipulation check, differences in the self-rated objective complexity scores were assessed for the four flight sequences (take off/climb/cruise, airspeed indicator failure, low oil pressure, and right magneto failure) using a one-way repeated measures ANOVA. With alpha set at .05, a statistically significant outcome was observed,  $F(3, 120) = 17.58$ ,  $p = .00$ , partial  $\eta^2 = .305$ . Table 2 lists the average self-rated cognitive complexity score (+SD) and outcomes of Bonferroni pairwise comparisons. Consistent with expectations, the right magneto failure and low oil pressure events achieved objective complexity scores that were significantly greater than the scores for the take off/climb/cruise and airspeed indicator failure flight sequences. Finally, unless otherwise stated, for all statistical tests alpha was set at .05.

The number of AOI accessed during the two time-periods (30–0 s prior to the onset of the instrument abnormality, and 0–30 s following the release of the instrument failure) associated with the three instrument failure events were analyzed using a  $3 \times 2$  repeated measures ANOVA. A statistically significant interaction was observed,  $F(2, 80) = 8.27$ ,  $p = .003$ , partial  $\eta^2 = .171$ . Table 3 shows the means, standard deviations and significance level for each variable. The pilots acquired a significantly greater number of AOI during the higher objective complexity events in comparison to the lower cognitive complexity events.

To assess the time spent examining the information acquired, fixation durations were divided into five categories: (1) <151 ms, (2) 151–301 ms, (3) 301–450 ms, (4) 451–600, and (5) >601 ms; and the proportion of fixation durations occurring in each category was calculated. As employed by Velichkovsky et al. (2001), this method of assessing fixation duration data is favored over the interpretation of average fixation durations, as measures of central tendency tend to ignore the dispersion of data. Across



**Table 2.** Descriptive statistics and analysis of variance for self-rated cognitive complexity scores across the four flight sequences (N = 41).

	Flight sequences				Significance (2-tailed, post-hoc)
	Take/off, climb, cruise	Airspeed indicator failure	Right magneto failure	Low oil pressure	
Self-rated cognitive complexity		7.88 (1.34)	9.04 (1.57)		<i>p</i> = .000*
	8.28 (1.18)		9.04 (1.57)		<i>p</i> = .006*
		7.88 (1.34)		9.50 (1.43)	<i>p</i> = .000*
			9.04 (1.57)	9.50 (1.43)	<i>p</i> = .089
	8.28 (1.18)			9.50 (1.43)	<i>p</i> = .000*
	8.28 (1.18)	7.88 (1.34)			<i>p</i> = .069

\* Difference between group means is significant at the .0083 level.

**Table 3.** Descriptive statistics and analysis of variance for the number of AOI accessed: Prior and post abnormality (N = 41)

Event	Time period		Significance (2-tailed, post hoc)
	Pre failure	Post failure	
Airspeed indicator failure ( <i>Lower CC</i> )	5.68 (1.01)	5.51 (1.40)	<i>p</i> = .489
Right magneto failure ( <i>Higher CC</i> )	5.39 (1.02)	6.17 (.80)	<i>p</i> = .000*
Low oil pressure ( <i>Higher CC</i> )	5.14 (1.28)	6.07 (1.82)	<i>p</i> = .000*

\* Difference between group means is significant at the .017 level

the subjects, the proportion of the total number of fixation durations occurring in each fixation duration category was calculated for the 30–0 s prior to the onset of each failure event and the 0–30 s following the activation of these failure events.

With alpha set at .017 (.05/3), a two way chi-square revealed differences in the proportion of fixation durations prior to, and post the right magneto failure,  $\chi^2$  (4, N = 8145) = 160.67, *p* = .000 and low oil pressure events,  $\chi^2$  (4, N = 8145) = 365.12,

$p = .000$ . In the 0–30 s following the onset of the right magneto failure the proportion of fixation durations in the <151 ms category increased significantly from 39.9 % to 60.1 %, while the proportion of fixation durations in the >601 ms category moved from 57.5 % to 42.5 %. Similarly, in the 0–30 s following the release of the low oil pressure failure the proportion of fixation durations in the <151 ms category increased significantly from 37.3 % to 62.7 %, while the proportion of fixation durations in the >601 ms category decreased from 63.45 % to 36.6 %. No significant differences were observed for the 151–301 ms, 301–450 ms, and 451–600 ms categories. Finally, no significant difference was evident for the airspeed indicator failure event, indicating relatively comparable fixation durations prior to, and following this instrument abnormality.

### 3.4 Outcomes

Study 1 sought to establish whether differences in objective complexity were associated with differences in patterns of information acquisition across a range of simulated failure events. Specifically, it was hypothesized that, in comparison to objectively complex scenarios, pilots' would access less information, with less frequency, and within more time during the less complex scenarios.

As a manipulation check, subjective assessments of objective complexity confirmed the a priori classification of the flight sequences as either more or less objectively complex. Significant differences were observed for the frequency of AOI accessed, and the frequency of longer and shorter fixation durations across the scenarios. Consistent with the hypothesis, a greater frequency of shorter fixations, fewer longer fixations, and a greater number of AOI were accessed during those failure events that embodied greater levels of objective task complexity.

In addition to establishing the relationship between objective task complexity and patterns of information acquisition, the outcomes of Study 1 also suggest that operators, having completed a task, possess the capacity to differentiate flight sequences of greater or lesser objective complexity. However, there was no comparison against measures of subjective complexity. Therefore, it is not clear whether measures of objective complexity account for the perceived cognitive load associated with the tasks or whether they account for the intrinsic characteristics of the task, independent of cognitive load. Study 2 was designed to both replicate and extend the outcomes of Study 1 by incorporating a comparative analysis of measures of subjective and objective complexity.

## 4 Study 2: Explanatory Power of Task Complexity

From the perspective of cognitive load theory, the resources expended by an operator and therefore, the load imposed, can be established through subjective assessments of task difficulty, while the resources allocated to achieve successful task performance can be established through perceptions of the effort invested [5]. Collectively, subjective assessments of difficulty and effort correspond to elements of Maynard and Hakel's [4] notion of subjective complexity. The aim of Study 2 was to establish which of the operator-rated measures of subjective or objective complexity better discriminates flight simulation tasks representing different levels of objective complexity.

#### 4.1 Method

**Subjects.** This study was approved by the UWS Human Research Ethics Committee and subject's gave their written informed consent to participate. The subjects comprised 38 general aviation pilots, of whom 34 were male and four were female. They ranged in age from 20 to 63 years, with a mean age of 33 years ( $SD = 13.19$ ). The subjects had accumulated between 50 and 7100 total flying hours ( $M = 583.55$  h,  $SD = 1138.97$ ), of which between seven and 6850 were accumulated as pilot-in-command ( $M = 436.55$  h,  $SD = 1102.41$ ). The majority of subjects held a commercial or private pilots license, and were neither a flight instructor nor instrument-rated. The subjects represented a convenience sample recruited through a University-based aviation research register and through a number of flight training organizations. They were compensated \$20.00 for travel expenses.

**Equipment and Materials.** The study was conducted using a simulated Cessna 172 aircraft operated on the Precision Flight Controls flight simulator employed in Study 1. The gaze tracker employed in the present study was identical to that employed in Study 1. The study incorporated two flight sequences that corresponded to two levels of objective complexity (higher and lower) based on the outcomes of Study 1. The low objective complexity task comprised a straight and level flight sequence while the high complexity event comprised the failure of the right magneto (partial engine failure). Immediately following the completion of each flight sequence, subjects were asked to complete the post-event questionnaire. The questionnaire comprised six questions relating to elements of objective complexity (see Table 4), and subjects responded to each question on a seven-point scale. Objective complexity rating scores ranged from 6 to 42, with a higher score corresponding to a relatively greater rating of objective complexity. Subjects were also asked to rate from 0 to 100 %, the effort invested during the performance of the two scenarios. Finally, subjects were asked to recall as much of the information that they had acquired during each scenario.

**Table 4.** Post flight sequence questionnaire used in Study 2

Cognitive complexity dimension	Question/s
Dynamism	To what extent did the information that you used to perform the scenario change over time?  How many different pieces of information did you need to be aware of to complete the scenario?
Uncertainty	To what extent could you rely on the information that you used to perform the scenario?
Interconnectedness among sub systems	How difficult was the scenario?
Risk	(a) How often did you make mistakes during the scenario? (b) How important was it to avoid errors during the scenario?

**Procedure.** Following the provision of informed consent, the gaze tracker was calibrated and subjects were asked to complete a two-minute practice flight to become accustomed to the simulator. On completion of the practice flight, they were advised that they would assume the role of pilot-in-command of a Cessna 172S aircraft and navigate under Visual Flight Rules (VFR). The scenario involved an approach and landing at Cairns airport on the North-East Coast of Australia.

Subjects were advised that Cairns airport operates under Class C airspace and therefore, they would need to maintain contact with air traffic control using the headset provided. They were also advised that an ‘intelligent aviation assistant’ would inform them of any abnormalities during the flight. As in Study 1, the use of the ‘intelligent aviation assistant’ was explicitly designed to ensure a constant trigger for the onset of the failure event. The experimenter acted as both the air traffic controller and the ‘intelligent aviation assistant’.

Using the Instructor PC, the experimenter positioned the aircraft 15 miles from Cairns Airport. This ensured that all subjects commenced the flight at the same distance from the runway. The first flight sequence required the management of the aircraft from cruise into a descent profile. The simulation was paused at 10 miles from the aerodrome, and the subjects were asked to complete the relevant post-flight event questions.

In the second flight sequence, the subjects were asked to continue the approach and landing. The right magneto failure was activated when the aircraft was 4.7 miles from Cairns airport and the ‘intelligent aviation assistant’ informed the subjects of the failure. The second flight sequence was completed once the subjects had landed the aircraft. On landing at Cairns airport, the subjects were asked to complete the final post-flight event questions.

## 4.2 Results

Prior to commencing the analyses, data were screened to ensure that the assumptions of normality were satisfactory. Across the analyses, seven univariate outliers were observed and corrected for, and assumptions of normality were achieved for univariate and multivariate analyses. Data screening revealed no significant correlation between total hours flight experience and each of the dependent variables, indicating that total hours of flight experience (as a measure of expertise) did not need to be included as a covariate in the analysis. As manipulation checks, differences in self-rated objective complexity, difficulty, and effort associated with the scenarios (cruise/descent, and right magneto failure) were examined using three dependent samples *t* tests. With alpha set at .017, the right magneto failure scenario was associated with a significantly greater level of objective complexity,  $t(37) = 9.10$ ,  $p = .000$ , partial  $\eta^2 = .691$ , perceived difficulty, and perceived effort invested,  $t(37) = -8.09$ ,  $p = .000$ , partial  $\eta^2 = .639$ , than the cruise/descent scenario (see Table 5 for descriptive statistics). Finally, unless otherwise stated, for all statistical tests alpha was set at .05.

A dependent samples *t* test was conducted for the mean number of AOI accessed during the two time-periods (30–0 s prior to the completion of the cruise/descent flight sequence, and 0–30 s following the release of the right magneto failure). No statistically significant difference was evident,  $t(37) = -1.17$ ,  $p = .426$ , partial  $\eta^2 = .036$ .

**Table 5.** Descriptive statistics for self-rated cognitive complexity, perceived difficulty, and perceived effort across the two scenarios (N = 38).

Variable	Scenario	
	Cruise/descent	Right magneto failure
Self-rated cognitive complexity	20.68 (4.90)	29.08 (3.07)
Perceived difficulty	3.16 (1.53)	5.47 (1.18)
Perceived effort	25.27 (12.45)	52.25 (82.24)

A dependent samples t test established whether differences existed between the frequency of different pieces of information that the subjects reported as having acquired during the two flight sequences. The results indicated a statistically significant difference between the number of different pieces of information that were reportedly acquired during the two tasks,  $t(37) = -2.87$ ,  $p = .003$ , partial  $\eta^2 = .182$ . Specifically, the mean number of features reported for the right magneto failure event ( $\bar{X} = 7.00$ ,  $SD = 1.59$ ) was greater than the mean number of features reported for the cruise/descent event ( $\bar{X} = 6.21$ ,  $SD = 1.73$ ).

The method of calculating fixation durations was consistent with Study 1. A two way chi-square revealed that the proportion of fixation durations differed significantly prior to, and post the right magneto failure,  $\chi^2(4, N = 7600) = 153.30$ ,  $p = .000$ . In the 0–30 s following the release of the right magneto failure, the proportion of fixation durations in the <151 ms category increased from 40.2 % to 59.8 %, while the proportion of fixation durations in the >601 ms category decreased from 63.4 % to 42.5 %. No significant differences were observed for the 151–301 ms, 301–450 ms, and 451–600 ms categories.

A direct, discriminate function analysis was used to establish the precision with which the subjective and objective measures of complexity discriminated the cruise/descent flight sequence from the right magneto failure flight sequence. The predictors were: (1) rated objective complexity scores, (2) perceived difficulty scores, (3) perceived effort scores, (4) the number of different pieces of information that were reportedly accessed, (5) the number of AOI accessed, (6) the proportion of fixation durations in the <151 ms category, and (7) the proportion of fixation durations in the >601 ms category. Predictors 5–7 occurred either during in the 30–0 s prior to the completion of the cruise/descent event or during the 0–30 s following the release of the release of the right magneto failure.

Using a conservative alpha of .01, one discriminant function was calculated, with a strong association between groups and predictors,  $\chi^2(7) = 68.63$ ,  $p = <.01$ , that accounted for 100 % of between-group variability. The loading matrix of correlations

between predictors and discriminant functions suggests that the best three predictors for discriminating the first function are, respectively: (1) rated objective complexity, (2) perceived difficulty scores, and (3) perceived effort scores. As evident in Table 6, the standardized canonical coefficients indicate the stability of rated objective complexity, while perceived difficulty may be an unstable outcome.

With the use of a jack-knifed classification procedure for the total sample of 38 pilots, across the two events, 88.2 % of subjects were classified correctly, compared to 50.0 % who would be correctly classified by chance alone. The stability of the classification procedure was confirmed by a cross-validation run, where there was an 86.8 % correct classification rate, indicating a high degree of consistency in the classification scheme.

**Table 6.** Results of discriminant function analysis variables related to the cruise/descent (*lower CC*) and right magneto failure (*higher CC*) events

Predictor variable	Correlations of predictor variables with discriminant functions (Standardized canonical coefficients)	Univariate $F(1, 74)$	Wilks' Lambda
	1		
Self-rated cognitive complexity	.811 (.735)	80.10*	.480
Perceived difficulty	.668 (-.099)	54.43*	.576
Perceived effort	.595 (.361)	43.08*	.632
< 151 ms fixation duration category	.582 (.375)	41.33*	.642
> 601 ms fixation duration category	-.380 (-.002)	17.62*	.808
Number of AOI accessed	.188 (-.193)	1.09	.985
Self reports of pieces of information accessed	.095 (.281)	4.29	.945
Canonical R	.789		
Eigenvalue	1.647		

Note. \*  $p < .001$

### 4.3 Outcomes

The aims of Study 2 were to replicate the outcomes of Study 1 and to examine the relative utility of measures of subjective and objective complexity in discriminating flight sequences that differed in complexity. Apart from the results pertaining to the frequency of fixations across the various AOI, the results confirmed the outcomes of Study 1 with the more complex scenario eliciting a relatively greater proportion of fixation durations in the <151 ms category, and a relatively lesser proportion of fixation durations in the >601 ms category. Differences were also observed in the ratings of objective complexity.

The outcomes of Study 2 also indicated that amongst the variables, ratings of objective complexity best discriminated the two flight sequences and provided a measure of performance distinct from ratings of difficulty and effort. This suggests that objective and subjective measures of complexity contribute differently to the performance of a task and that both measures can be self-rated at a level that discriminated more complex from less complex tasks.

## 5 General Discussion

The inherent complexity associated with a product or task is a feature that potentially has implications for human performance. Overly complex systems are likely to be error-prone and/or discarded in favor of systems that are relatively less complex. Maynard and Hakel [4], amongst others (e.g., [7]), conceptualize the complexity of a task as a combination of both users' subjective perception of the complexity of the task and the complexity of the task inherent in its execution.

The subjective perception of complexity corresponds to notions of cognitive load and the difficulty that is expected to be encountered in performing a task successfully. The nature of the situation, and the experience and inherent capability of the user determine both perceptions and difficulty, and the subsequent effort that is likely to be invested in the task. Therefore, the difficulty in relying solely on subjective perceptions of complexity lies in the individual differences between users.

The aim of this study was first to establish whether it is possible to differentiate tasks on the basis of objective complexity. Objective complexity relates to the inherent features of the task and the various cognitive and perceptual activities that need to take place to ensure successful performance. In establishing differences in objective complexity, it becomes possible to assess the utility of alternative designs on the basis of the underlying information processing demands, irrespective of factors such as experience or motivation. A series of four simulated flight tasks were developed by subject-matter experts that differed on the four aspects of complexity proposed by Woods [6]. Controlling for task experience, qualified pilots 'flew' the simulated flights and their process of information acquisition was compared during a diagnostic event.

The results indicated that the process of information acquisition differed on the basis of the objective complexity of the task, with more complex tasks associated with a greater frequency of Areas of Interest (AOI) in the 30 s following the onset of the event. More complex tasks were also associated with a lower proportion of fixations at less than 151 ms. The latter effect occurred in both Studies 1 and 2 and, together with a reduction in the proportion of fixations in the >601 ms category of fixations, highlights the behavioral changes that occur with changes in the complexity of a task. These changes corresponded to user-rated assessments of the objective complexity of the task.

A secondary aim of this study was to determine whether subjective and objective complexity contribute differently to the performance of a task and whether self-rated assessments of objective complexity discriminated tasks of greater and lesser complexity. Consistent with expectations, the results indicated that, amongst the outcome variables, self-rated assessments of objective complexity was the variable that best discriminated the tasks, while self-rated assessments of subjective complexity discriminated the tasks to



a slightly lesser extent. This suggests that subjective and objective measures of task complexity comprise distinct, but related constructs that contribute to an overall assessment of complexity.

The joint contribution of subjective and objective complexity corresponds to the integrated approach to the assessment of complexity proposed by Maynard and Hakel [4]. However, the outcomes of the present study also suggest that assessments of objective complexity can be self-rated, thereby reducing the potential costs in establishing the objective complexity of the task through the use of subject matter experts or task analyses.

## 5.1 Limitations and Implications

By establishing the utility of measures of subjective and objective complexity, the flight scenarios developed in the present study inevitably represented extremes as a means of establishing differences between more and less complex tasks. However, there remains a need to determine whether the self-rated measure of objective complexity retains a degree of sensitivity for graduated levels of complexity. This would enable comparisons whereby relatively small changes in the design of a system can be assessed in terms of their impact on the level of task complexity.

In addition to assessments of graduations in complexity, there is a need to establish whether systematic changes in subjective complexity impact either the process of information acquisition or the self-rated assessment of objective complexity. Since objective complexity relates to the nature of the task, variations in the perceived difficulty or effort associated with the performance of a task should occur independently of the intrinsic characteristics of the task.

From the perspective of system design, the outcomes of the present studies offer an opportunity to ensure that the subjective perceptions of users represent the intrinsic complexity of the task. For example, in assessing a new product, a designer may seek the responses of users ranging in experience from the expert to the novice and may find a breadth of subjective perceptions that are not necessarily based on the inherent complexity of the task specified. Experts may rate the tasks as relatively simple while novices may rate the same tasks as particularly difficult. By comparing these data with the data pertaining to objective complexity, it becomes possible to establish a standard against which the complexity of a task can be compared.

## 5.2 Conclusions and Future Outlook

This research sought to differentiate subjective and objective complexity as distinctive features associated with the overall complexity of a task. Using simulated in-flight events that required a diagnostic response, differences between more complex and less complex tasks were evident which suggested that information acquisition behavior changes as a function of the objective complexity of a task. However, the results also suggested that self-ratings of objective complexity can be employed to establish the intrinsic complexity associated with the task. The outcomes have implications for system design and development in the future. The authors will continue to work with

projects examining the relationship between self and objective task complexity and the behavioral elements of task performance.

The current research would be further strengthened by assessing ‘primary-task performance’ at the level of the individual operator. In particular, it is likely that, where the primary-task has been performed at a less than satisfactory level, deficiencies in operator behavior might be traced to problems with the perception of the cognitive complexity of the task and the pattern of feature acquisition. Therefore, assessing the alignment, or misalignment, between primary-task performance, the perception of cognitive complexity, and the pattern of feature acquisition, would provide further validation of the outcomes arising from the current project.

## References

1. Bedney, C., Karwowski, W.: *A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design*. Taylor & Francis, Boca Raton (2007)
2. Woods, D.D., Hollnagel, E.: Mapping cognitive demands in complex problem-solving worlds. *Int. J. Man-Mach. Stud.* **26**, 257–275 (1987)
3. Bach, C., Scapin, D.L.: Comparing inspections and user testing for the evaluation of virtual environments. *Int. J. Hum.-Comput. Interact.* **26**(8), 786–824 (2010)
4. Maynard, D.C., Hakel, M.D.: Effects of objective and subjective task complexity on performance. *Hum. Perform.* **10**, 303–330 (1997)
5. Sweller, J., van Merriënboer, J.J.G., Paas, F.: Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **10**, 251–296 (1998)
6. Woods, D.: Coping with complexity: the psychology of human behaviour in complex systems. In: Goodstein, J., Andersen, H., Olsen, B. (eds.) *Tasks, Errors and Mental Models*, pp. 128–148. Taylor & Francis, London (1988)
7. Wood, R.E.: Task complexity: definition of the construct. *Organ. Behav. Hum. Decis. Process.* **37**, 60–82 (1986)
8. Campbell, D.J.: Task complexity: a review and analysis. *Acad. Manag. Rev.* **13**(1), 40–52 (1988)
9. Reason, J.: *Human Error*. Cambridge University Press, Cambridge (1990)
10. Williams, M.D., Hollan, J.D., Stevens, A.L.: Human reasoning about a simple physical system: a first pass. In: Gentner, D., Stevens, A.S. (eds.) *Mental Models*, pp. 131–153. Lawrence Erlbaum, Hillsdale (1983)
11. Wiggins, M.W., O’Hare, D.: Expertise in aeronautical weather-related decision-making: a cross-sectional analysis of general aviation pilots. *J. Exp. Psychol. Appl.* **1**, 305–320 (1995)
12. Sarter, N.B., Mumaw, R.J., Wickens, C.D.: Pilots’ monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye tracking data. *Hum. Factors* **49**, 347–357 (2007)
13. Schriver, A.T., Morrow, D.G., Wickens, C.D., Talleur, D.A.: Expertise differences in attentional strategies related to pilot decision making. *Hum. Factors* **50**(6), 864–878 (2008)
14. Patrick, J., James, N.: Process tracing of complex cognitive work tasks. *J. Occup. Organ. Psychol.* **77**, 259–280 (2004)
15. Renshaw, P.F., Wiggins, M.W.: A self-report critical incident assessment tool for army night vision goggle helicopter operations. *Hum. Factors* **49**, 200–213 (2007)

16. Salmon, P.M., Stanton, N.A., Walker, G.H., Barber, C., Jenkins, D.P., McMaster, R., et al.: What's really going on? Review of situation awareness models for individuals and teams. *Theor. Issues Ergon.* **9**(4), 297–323 (2008)
17. Wiggins, M.W.: The development of computer-assisted learning systems for general aviation. In: O'Hare, D. (ed.) *Human Performance in General Aviation*, pp. 153–172. Ashgate, Aldershot (1999)
18. Tole, J.R., Harris, R.L., Stephens, A.T., Ephrath, A.R.: Visual scanning behavior and mental workload in aircraft pilots. *Aviat. Space Environ. Med.* **53**, 54–61 (1982)
19. Rivercourt, M., Kuperus, M.N., Post, W.J., Mulder, L.J.M.: Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics* **51**(9), 1295–1319 (2008)
20. Velichkovsky, B.M., Dornhoefer, S.M., Pannasch, S., Unema, P.J.A.: Visual fixations and level of attentional processing. In: *Proceedings of the 2000 Eye Tracking Research and Applications Symposium*, pp. 79–85. ACM Press, New York (2000)
21. Wiggins, M.W.: Cue-based processing and human performance. In: Karwowski, W. (ed.) *Encyclopedia of Ergonomics and Human Factors*, pp. 641–645. Taylor and Francis, London (2006)