Computer-assisted cognitive function assessment of pilots

Roderick Westerman, PhD, MD, FRACGP; David G Darby, PhD, FRACP; Paul Maruff, PhD; and Alexander Collie, BAppSc(Hons)

Pilots are one of several special groups who operate in an environment unforgiving of human error, where cognitive failure can lead to catastrophic consequences. Cognitive function testing is one means of selecting mentally capable pilots. It is also a means of screening for covert disease, and it can be used to establish baseline performance data and provide ongoing monitoring of health. (For these reasons, cognitive function testing of all active service personnel might be considered appropriate.)

Pilot selection procedures used by the modern armed forces of most nations have come a long way from the ad hoc processes used at the time of World War One, when cavalry officers with “a good pair of hands” were the preferred recruits into the fledgling Royal Flying Corps. Before the war’s end it was clear that eyesight, cardiorespiratory fitness and even personality were of vital importance to success and survival in the air. Better selection procedures and medical standards for aviators were introduced at the start of World War Two, and by its end the increasingly high performance of aircraft and equipment stood in stark contrast to the physical performance limitations of the pilots.

Since World War Two, the physical and cognitive requirements for aircrew in civil airline aviation has moved towards the elite standard required in military aviation. However, despite the high standard required by aircrew selection procedures, conditions impairing cognition can still be present at recruitment or develop during service life, suggesting major roles for baseline and ongoing cognitive assessments. Closed head injuries from motor vehicle accidents are increasingly common in young adults and may be covert. Alcohol or illicit drug use, increasingly stressful lifestyles, job demands and economic pressures can all lead to psychiatric illness and secondary cognitive impairment. Occasionally, neurological illness, including neurodegenerative disease, can affect pilots. Therefore, the ability to quickly and accurately measure critical aspects of cognition is of vital importance.

When combined with other screening procedures, the use of good cognitive function assessment tools can help to improve personnel selection. The benefits of improved selection include fewer training failures and more adaptable and successful personnel “on the job”. Moreover, by archiving the entrance cognitive test results, Defence Force medical sections...
have overcome many of the delays and difficulties in assessing fitness to resume full duties in the recovery phase after head injury, brain trauma, CNS infection (including HIV infection), and alcohol or drug problems. In the civil aviation setting, cognitive function screening is becoming more relevant with the increased number of older pilots.

I Examples of cognitive testing tasks from Cogscreen-AE

Symbol Digit Coding task
A screen well into the explanatory instructions for this subtest. As with all the Cogscreen-AE subtests, a practice session precedes the test itself.

Visual Sequence Comparison task
The screen with the first page of instructions for this subtest. The several pages of instructions are followed by a short practice session and then the actual test.

Cogscreen-AE

Background and history

The first systematic attempt at computerised assessment in military aviation was Cogscreen Aeromedical Edition (Cogscreen-AE). Cogscreen-AE stems from an eight-year international effort aimed at developing a sensitive, specific tool for detecting cognitive changes resulting from mild brain dysfunction. In 1987 the US Federal Aviation Administration (FAA) issued a request for proposals from investigators to evaluate existing cognitive testing approaches with respect to their ability to detect subtle brain dysfunction and the risks these posed to aviation safety. Based upon these initial studies, the FAA then sought proposals for developing an automated or computerised instrument that would be suitable for neuropsychological screening in medical certification.

This resulted in a three-phase research and development project which included an extensive review of the literature surveying mental status examinations, neuropsychological screening tests, neuropsychological test batteries, and computer-based performance assessment batteries.

An empirical study (Phase A) then followed, comparing the performance of 60 aviators and 60 mildly brain-impaired patients who were matched for age and education on formal mental status tests, conventional neuropsychological measures, and computerised performance tests of aviation-related abilities. As specified by the FAA, the primary mental status test used was the Mini-Mental State Examination (MMSE).
which was compared with a battery of tests with known sensitivity and specificity for brain dysfunction, and performance tests with known validity for prediction of aviation-related performance. The study results were interesting in that patients with brain dysfunction could not be discriminated from pilots by MMSE, but a computerised subtest from the Naval Medical Research Institute assessment battery (Matching-to-Sample) revealed significant group differences.

Overall, the computerised measures demonstrated excellent sensitivity to mild brain dysfunction with acceptable levels of specificity, and were at least as sensitive as conventional neuropsychological instruments in detecting mild brain dysfunction.

The advantages of computer-based cognitive testing include standardised administration, self-paced instruction, multiple randomised forms, accurate measurement of reaction times and automatic standardised scoring and reporting. These advantages led the FAA to award Phase B contracts to develop a fully computerised, 30 minute cognitive screening examination. A design review advisory committee of aviation and medical specialists used the Phase A results to specify the subtests, hardware and software to be included, and the requirements of automated logging and scoring of performance data. Validation of the Phase B version of Cogscreen-AE used 40 fit pilots balanced across four age-groups and 40 patients with mild brain dysfunction secondary to head injury, stroke, substance abuse and mild dementia who were matched to the pilots for age and IQ score. Cogscreen-AE was able to correctly classify 32 of the 40 patients with brain dysfunction and incorrectly classified only three pilots. By contrast, the conventional neuropsychological test battery correctly classified only 20 of the 40 patients and incorrectly classified two pilots.

The findings from the Phase B project demonstrated Cogscreen-AE’s capacity to serve as a relatively inexpensive and efficient adjunct to conventional baseline neuropsychological assessment in the evaluation and medical certification of airmen. To establish the validity of Cogscreen-AE as an occupationally relevant assessment instrument, it was also essential to demonstrate a relationship between Cogscreen-AE scores and flight performance. This aim has been achieved by the studies of Yakimovich et al., Hyland et al., Hoffmann et al., and extended in the study by Taylor et al.

Structure, sequence and subtest content of Cogscreen-AE

Cogscreen-AE consists of a series of computerised tasks. Although subjects are supervised, the computer administers each test item, records the responses and scores the tests. Each subtest is self-contained and presented with instructions and a practice segment (which allows repeats) before the test item begins. The total testing takes about 50 minutes to complete and uses a light-pen as the input device. The light-pen, unlike the keyboard or mouse, is a device which most pilots are equally capable of handling (few have the advantage of prior experience).
experience, at least for the first administration). The Cogscreen-AE subtests include the following items:

- backward digit span
- simple maths problems
- visual sequence comparison
- symbol digit coding (immediate and delayed recall)
- matching to sample
- manikin
- divided attention task
- auditory sequence comparisons
- pathfinder (numbers and letters)
- shifting attention task
- dual task.

These tests cover a wide cross-section of cognitive abilities. Subjects are requested to read the specific instructions for each subtest and perform the test as quickly and accurately as possible. Examples are shown in Box 1.

**Competencies tested by Cogscreen-AE**

The cognitive functions sampled by Cogscreen-AE include visual scanning and sequencing, attribute identification, visual perception and spatial processing, motor coordination, choice visual reaction time, tracking, working memory, and numerical operations.

The computer measures and scores

1. speed (responses per minute)
2. accuracy (percentage of correct responses)
3. throughput (number of correct responses per minute)
4. process (success with subtests emphasising new concept formation, logical reasoning, and adaptability).

The program also estimates the probability of brain dysfunction from a logistic regression algorithm and this is expressed as the logistic regression probability value (LPRV) coefficient. A higher LPRV score represents an increasing probability of brain dysfunction. For healthy younger pilots (aged below 45) only 10% will have a LPRV score above 0.6, yet 29.4% of pilots older than 45 in the US Aviator Normative sample group had a score above 0.6. Users of Cogscreen-AE must be sensitive to this age relationship, yet LPRV scores above 0.6 should be taken seriously and indicate a need for fuller neuropsychological evaluation.

**Relationship of Cogscreen-AE to flight simulator performance and age**

Taylor et al reported the relationship between Cogscreen-AE scores and flight simulator performance in 100 pilots aged between 50 and 69 years. They concluded that Cogscreen-AE taps skills relevant to piloting, and that their study, with other existing data, justifies further validation studies of Cogscreen-AE as a clinical instrument for assessing aviators.

An important goal is to be able to predict individual aviator performance. In an initial attempt, Hyland et al and Hardy and Parasuraman have proposed a scheme for organising the various predictors of flight performance. These predictors include:

2 **Examples of cases tested in Australia with Cogscreen-AE**

**Case 1.** A 33-year-old Airline First Officer suffered sufficient clinical depression in February 2000 to require treatment with antidepressant (selective serotonin reuptake inhibitor) medication. He ceased flying at that time. Within three months of commencing treatment he was so much improved that both he and his psychiatrist were asking about the possibility of his resuming work. Approached about this by the psychiatrist, one of us (R.A.W) pointed out that fortuitously during 1999 the same pilot had been tested with Cogscreen-AE as a volunteer in a trial of Cogscreen as a pilot selection tool. With the pilot’s written consent, the results of both tests were evaluated and compared and the detailed results and interpretation were presented to the Civil Aviation Safety Authority. These data, together with the psychiatrist’s report, were used to allow resumption of flying with the stipulated endorsement “As or with co-pilot”, regular psychiatric reviews and a Cogscreen retest after six months.

**Case 2.** A 32-year-old rotary wing pilot (First Officer) in general aviation complained of “headache and feeling unwell” during the return flight from an oil-rig. After the Captain took over and landed, the First Officer was driven home by his girlfriend. Later in the evening, after making love, his condition worsened and she drove him to the hospital emergency department. He was investigated for severe headache and some visual field disturbance and was found to have sustained a spontaneous dissection of the internal carotid artery involving some ischaemia of the middle cerebral artery territory. Neurosurgical treatment reduced the immediate risk of further injury. The vascular accident left him with a small visual field defect. Magnetic resonance imaging showed evidence of some right frontal lobe damage. Seven months later he had improved considerably, with no seizures, but his cognitive performance on Cogscreen-AE put him in the bottom fifth percentile on 11 of 16 subtests and the LRPV probability of brain dysfunction was 1.0. His rehabilitation has included returning to his original trade of cabinet-making. Whether he ever could be fit to resume flying again (multicrew only) would depend upon all the results of neurological follow-up over several years. This should include assessment of any residual neurological deficit, risk of seizures or reoccurring vascular accident, and reassessment of cognitive function. The measurement of cognitive function recovery in such a case could be more accurately followed by serial performance of Cogscreen-AE or similar test.
domain-independent cognitive and motor skills (regarded as assessable with Cogscreen-AE)

- domain-dependent aviation knowledge
- pilot characteristics (eg, age, cardiovascular status, drug dependency, agreeableness)
- stressors (eg, difficult flight conditions, fatigue and interpersonal conflicts). A study by Hoffmann et al shows clearly how the predictive power of different Cogscreen-AE variables can depend on the criterion or outcome of interest. For instance, domain-dependent knowledge predicted training success ($P<0.05$) but not line-check ratings of aircraft control. Conversely, Cogscreen-AE Manikin and Symbol Digit Scores predicted aircraft control but not training success. Furthermore, Cogscreen-AE Dual Task and Divided Attention Scores predicted training success, compliance with procedures, and crew resource management ratings, but not aircraft control.

These studies attest to the difficulty of predicting aviator performance using single measures. Assessment of a combination of skills is much more likely to be required. It is also unlikely that any single criterion of occupational fitness will suffice. Hence, approaches based on profiles or patterns of combined test performance are more effective. These mirror the clinical neuropsychological method and allow the importance of strengths and weaknesses to be evaluated in context.

### Further directions of cognitive test development

The application of computer methods to the cognitive assessment of pilots has been an important step in the development of tools to ensure competence and fitness to operate in service personnel. However, there is still room for significant improvement in computer assessment protocols.

### Sensitivity and specificity

Although it is crucial for a cognitive test to be relevant to the target group, restricting test development to that group raises the potential for the test to lack sensitivity to all forms of cognitive impairment. New tests should be developed and then tested in groups of patients with well-defined clinical conditions that are known to interfere with cognitive function. For

### 3 Congruency decision task from CogState

1: The keyboard representation shows highlighted keys to be used in this task with preferred hand placement for two-handed subjects. A green background underlies all CogState tests. Cards above the keyboard are both red (ie, congruent in colour) requiring a “k” key response. This is the appearance immediately the cards turn face up.

2: Cards shown in another trial are non-congruent in colour, requiring a “d” key response. The subject has not responded during the required time, an error warning has sounded and visual feedback assistance is being presented (the required key on the keyboard inverts and an arrow emphasises its relationship to the pair of cards).

When the subject correctly responds several times, the keyboard disappears and feedback is given visually and aurally using the cards only. This task is preceded by a similar reaction time task of lesser complexity, which is repeated during the test to monitor consistency.
example, the performance of patients with anxiety disorders or depression can define cognitive profiles or patterns of performance that could indicate the reason for poor performance in pilots. An understanding of the nature and severity of cognitive impairments in these serious conditions can provide important clues to the presentation in pilots of the cognitive consequences of more mild conditions such as chronic stress, acute stress or dysphoria. Similarly, the effect of common drugs on cognitive function (eg, benzodiazepines and alcohol) can also be determined by such computerised studies. These results can then be used to infer differential diagnostic causes in pilots where similar patterns of deficiency are found. Such screening tests should aim to do more than just detect deviation from normal. They should limit the diagnostic possibilities by defining recognisable patterns of abnormal performance.

Repeatability of testing

One important rule in measuring any type of mental function is that there are often considerable differences between individuals in their performance on different tests. Also, the significance of an observed cognitive deficit is better determined by comparing the individual’s performance with that from an earlier time rather than with an average estimated from a large normative database. It is therefore crucial that cognitive tests are able to be used repeatedly, with each test being as difficult as the tests preceding it. If not, performance improvements related to practice (practice effects) could in theory mask any impairments. While providing alternative forms is one solution (as used by Cogscreen-AE), most computerised tasks are limited to a finite number of alternative forms at present. Clearly, when operational fitness must be determined regularly, it would be preferable to have tests with an infinite number of possible forms.

4 Monitoring task from CogState

1: The space key is highlighted on the keyboard, indicating that it is the key to be pressed. There are 5 cards above the keyboard which move independently up or down towards the white horizontal lines. The direction of movement (up or down) and distance moved are randomised many times a second, causing the cards to move in an unpredictable jittering motion. The task is to press the space key when any card reaches either white horizontal line. No cards have yet reached a line, and no response is required, with key presses eliciting an error buzzer.

2: Some time later, the king of diamonds has reached the lower white line, so a space key response is required. However, the subject has not responded within the allowed maximum response time. The space key has inverted and further visual feedback has been provided by a yellow line drawn from the card to the space key.

A subsequent task combines this monitoring task with a binary choice task like the congruency decision task, creating a very difficult continuous performance task.
5 Matching task from CogState

In this task an array of six card pairs is dealt above two card packs. These are shown face up. Two cards appear face-up on the card decks and a response is required. Two screens from this task are shown.

1: The cards are not part of the six-pair array and a “d” key response is required. Here the subject has delayed in responding and the “d” key has inverted to provide feedback.

2: The face-up cards now match one of the pairs above and a “k” key response is required. Once again, the subject has delayed in responding and the “k” key has inverted to provide feedback. Visual feedback is given differently depending upon whether the cards are part of the six-pair array. After all pairs have been presented several times (with additional non-legend pairs randomly intermingled) the six-pair array turns face down and the subject must respond by memory. A subsequent test presents cards in a similar arrangement, but turns them face-down during learning trials, allowing the efficiency of new learning and retention to be measured.
sonnel could reduce the requirements for remote trained staff and improve efficiency.

**Availability of test results**

Using telecommunications transmission, the results of computerised testing could be made available to the individual, a commander, and medical or psychological sections simultaneously, thereby ensuring that proper objective operational safety requirements are enforced. Importantly, the results should be presented in forms that are understandable and meaningful to these different groups. Remote testing could occur in real time and even be incorporated into normal flight operations. Instant central comparison with prior performance data could allow rapid and accurate deployment decisions based on real fitness to operate or fly.

**CogState: An Australian computerised cognitive assessment tool**

In an attempt to satisfy the higher requirements of computerised cognitive function testing, we have developed and are testing “CogState”, a computer-based test to be delivered and scored via the Internet. CogState probes a number of cognitive domains, defined by a clinical neuropsychological model, including alertness, attention, working memory, spatial awareness, memory and executive functions. It can assess motivation, perseverance, the ability to sustain efficient performance, consistency and adaptability of learning, acquisition and retention of material and abstraction. It is self-administered, automatically scored and requires about 15–20 minutes to complete. Subtests merge into one another and utilise familiar visual forms (playing cards) which instruct the subject of the rules of each test by demonstration and feedback only (Boxes 3 to 5). An almost infinite number of forms are available due to randomisation and variably timed binary choices. Speed and accuracy are measured and integrated over subtests.

CogState is currently undergoing clinical trials in patients with neurological and psychiatric illness, and in healthy younger and older subjects experiencing pharmacological and environmental challenges. These data will allow differential profiles of disease-based or drug-induced cognitive impairment to be constructed for comparison with subject performance. It is being evaluated in aerospace applications and is part of a selection process for at least one domestic airline.

In healthy individuals, CogState performance varies with induced fatigue and stress levels. Reports of performance based on repeated tests or comparison with other individuals are generated. CogState can be set up to email test results to any predetermined individual. The test will be available for download from the Internet and therefore can be given in any location. All major computing platforms are supported and shortened forms for PDA administration are being tested.

Although in its early stages of development, CogState promises to be an extremely useful adjunct to both recruitment and fitness-to-fly assessments in the future.

**References**