A novel method of measuring the mental workload of anaesthetists during simulated practice

D. H. J. Davis¹, M. Oliver¹ and A. J. Byrne²*

¹Abertawe Bro Morgannwg University NHS Trust, Morriston Hospital, Swansea SA6 6NL, UK. ²School of Medicine, Swansea University, Swansea SA2 8PP, UK

*Corresponding author. E-mail: a.byrne@swansea.ac.uk

Background. Cognitive overload has been recognized as a significant cause of error in industries such as aviation, and measuring mental workload has become a key method of improving safety. The aim of this study was to pilot the use of a new method of measuring mental workload using a previously published study design.

Methods. Ten trainee anaesthetists were exposed to a simulated crisis, similar to that used in a previous study. The mental workload of the trainees was assessed by measuring their response times to a wireless vibrotactile device.

Results. Although all subjects treated the ‘patient’ adequately, response times increased significantly during the crisis (P=0.005). These findings are consistent with increased mental workload and with the findings of other studies using similar techniques.

Conclusions. These findings confirm the importance of mental workload to the performance of anaesthetists, and suggest that raised mental workload is likely to be a common problem. Although further studies are required, the method described may provide a useful method for the measurement of the mental workload of anaesthetists.

Br J Anaesth 2009; 103: 665–9

Keywords: education; human error; risk; safety, patient safety; simulation; training

Accepted for publication: August 5, 2009

The aim of this study was to pilot the use of a novel secondary task to measure the mental workload of anaesthetists using a simulated critical incident described in a previous study.¹

Mental workload is used to describe the amount of mental effort involved in performing any given task.¹ Given that there is a limit to the mental workload of any individual (mental capacity), mental workload is the proportion of that capacity in use at any point in time and will vary depending on the task being performed by the individual.²

The measurement of workload is considered important in many high-risk environments because of its implications for safety, crew size, and the effects of automation.³ Raised mental workload has been associated with greater likelihood of an error and poor performance.⁴

Although less well investigated, mental workload in clinicians is no less important to the occurrence and recovery from human error, and to the performance of individuals during critical incidents.⁵ Effective measures of mental workload can provide us with the means to evaluate strategies for reducing mental workload and thereby preventing error and poor performance in routine practice.⁶–¹⁰

Methods

Ethical approval was granted from the Swansea Medical School Ethics Committee and after obtaining informed consent, 10 trainee anaesthetists volunteered to take part in the study. Each subject had between 2 and 6 yr experience and was on a recognized training programme and responsible for delivering anaesthesia unaided, but with immediate supervision available.

Before commencing the study, a vibrotactile wireless device was attached to each subject’s arm using a soft material strap and holster. A signal was sent to this device at random intervals (10–30 s) causing it to vibrate. The
vibration was terminated by the subject pressing a button on the device. Subjects were asked to press the button whenever the device vibrated during the session, but were not given any other specific instructions. The computer automatically recorded the time the signal was generated and the time taken for the subject to respond in milliseconds. All data were recorded as numbered text files. If the subject did not press the button immediately, the device continued to vibrate until the button had been pressed.

A mock operating theatre was set up using a SimMan® (Laerdal) Advanced Patient Simulator®, operating table, anaesthetic machine (Draeger, Cato®), an anaesthetic assistant, alternative oxygen supply, drugs, needles, syringes, i.v. fluids, a stethoscope, and the usual airway devices and was based in a training area as used in the previous study using the same method.1

Each anaesthetist was introduced to the simulation as if taking over a routine case from a colleague, with the patient described as a 25-yr-old male undergoing a knee arthroscopy under general anaesthesia. He had been intubated due to a past medical history of oesophageal reflux.

Initially, all physiological parameters were normal and remained so for a 5 min baseline period. This allowed the performance of each subject to be measured during conditions of low mental workload.

The experimental period began after 5 min. There was a rapid decrease in heart rate to 30 beats min⁻¹, a decrease in oxygen saturation to 69%, a decrease in arterial pressure to 75/20 mm Hg, and a decrease in the end-tidal carbon dioxide partial pressure to 2.3 kPa. These abnormalities continued for the following 5 min, irrespective of the actions of the anaesthetist. After this period, the patient’s parameters gradually returned to pre-crisis levels in the final 5 min washout period.

As in the previous study, the patient’s physiological changes were set to continue through a series of pre-programmed abnormalities irrespective of the anaesthetist’s actions. This ensured that all anaesthetists were exposed to exactly the same information from the monitors and simulator over the time period, and that task demands were held constant from simulation to simulation.

After conclusion of the simulation, all volunteers were debriefed and were informed of the design of the study.

The data produced three results for each individual: the mean response time during the initial baseline period, the mean response time during the experimental period, and the mean response time during the washout period.

The difference between the baseline period and the experimental period was assessed using Wilcoxon’s rank signed test with a P-value of 0.05 being taken as significant.

Results

All anaesthetists responded to the scenario in an appropriate fashion. Each followed an ABC approach, including checking the airway, chest and circuits, administering oxygen 100%, i.v. fluid, vasoconstrictor drugs, atropine, and calling for help. All volunteers completed the scenario to full patient recovery.

In all cases, the device maintained its wireless connection to the computer, and data were captured for all subjects as intended.

Results from the baseline period show consistent, rapid responses from all subjects with most subjects showing little within-subject variability. Data approximated to a normal distribution. Mean response time for each individual was 683 ms with a range of (542–840). In eight of these 10 subjects, the standard deviation of their baseline responses was between 59 and 235 ms. In two subjects, one and two longer response times resulted in standard deviations of 700 and 513 ms. Median response times for this period were between 528 and 770 ms.

Results from the experimental period showed that the median response time for each subject increased from baseline as shown in Figure 1. The mean increase in response time was 83% with a range of 2–253%. The difference between the baseline period and the experimental period was statistically significant (n=10, T=0, P=0.005).

During the experimental period, variability increased with most subjects showing some values similar to baseline interspersed with prolonged times, with the results of all 10 subjects shown in Figure 2. To clarify the

![Fig 1](image-url) Median response time in milliseconds for each subject in the baseline (1), experimental (2), and washout (3) periods.
distribution of data, the responses from two subjects are shown in Figure 3. One of these shows a single long response time at the onset of the problem followed by a rapid response towards baseline. In contrast, the other subject shows the initial increase followed by a series of prolonged responses.

The apparent increase in response times just before 5 min is likely to be due to the asynchrony between the clocks in the two computers used in the study, as the crisis started without any prior warning.

Response times during the washout period returned towards baseline, with each subject’s response times 7% above baseline with a range of −9% to +40%. Four subjects’ response times were faster in the washout period. These results were not statistically different from baseline ($n=10$, $T=17$, $P=0.285$).

**Discussion**

Mental workload can be assessed concurrently in real-time using several different methods. These include psychological (rating scales), procedural (response to task-related demands), or physiological measures (heart/ventilatory frequency, skin resistance), and the use of secondary tasks (warning lights, mental arithmetic).

The secondary (subsidiary) task paradigm adds a minimally intrusive second task, whose performance is easily measured, to the primary task under study (i.e. administering anaesthesia or flying a plane).

The assumption in this method is that if the workload associated with the primary task becomes excessive, there will be limited spare capacity. At this point, the performance of the secondary task will be impaired.

Thus, if performance of the secondary task is monitored continuously, any decrease in its performance indicates that the mental workload of the operator is close to their limit and that their performance is likely to be impaired. This condition has been identified in anaesthetists through errors in simple mathematical problems, increased latency in identifying a warning light, and increased record keeping inaccuracies.

However, the key to this method is identifying a secondary task that is not intrusive and that would be acceptable in a clinical area. Previously described methods involving lights, sounds, and mental arithmetic are too intrusive in clinical areas or have stimuli that are masked by workplace noise or the mobility of staff.

This method of assessment is important as it could identify practitioners operating close to the limit of their abilities even in the absence of any visible deterioration in their performance. In addition, it could be used during training to ensure that clinicians were not just trained to the point of competence, but to the point where their responses became routine and effortless.
The results were consistent with those of previous studies and the thesis that the simulated crisis produced high levels of mental workload and this was measurable by deterioration in the secondary task. This supports the suggestion that cognitive overload may be a common problem in anaesthetists dealing with anaesthetic crises. However, we recognize that the lack of a validated measure of primary task workload means that we were unable to make a direct correlation between primary task workload and each subject’s mental workload.

We believe that this method has several advantages over previously described techniques. It is objective and avoids the inter-observer variability caused by a rating scale. The placement on the arm of the vibrotactile device does not impair mobility, and unlike a latency light, it does not have to be maintained within the subject’s visual field. The technique is minimally invasive and therefore could easily be used in a wide variety of real and simulated environments without interfering with normal working practices. The need to press a button, however, does mean that a subject would be unable to respond if both hands are occupied, although initial tests have shown that it can be activated through a gown allowing it to be used during an aseptic procedure.

The consistency of individual response time shown in Figure 2 suggests that responses during baseline, unstressed periods vary between individuals, but are stable for individuals over time. This suggests that once an individual’s responsiveness has been established, changes due to increased workload are likely to be relatively easy to quantify.

Further, as the device provides a constant level of stimulus until the subject responds, a prolonged delay indicates that the subject is unresponsive to external factors for that period. The patterns of response times shown in Figure 3 suggest that mental workload varies rapidly and that subjects rapidly pass through cycles of responsiveness and overload, shown by rapid and prolonged responses. This is important as any attempt to measure workload at intervals of >60 s is likely to produce unrepresentative results. It may also explain the variability seen in human performance, in that crucial information may be accepted or ignored depending on whether it arrives at a peak or trough of responsiveness.

The finding that several subjects recorded greatly prolonged responses supports the concept that some individuals may be more susceptible to stress and becoming overloaded or ‘maxed out’. In the future, such methods could be used as one part of an assessment strategy to identify susceptible individuals for extended training or perhaps even exclusion from high workload areas.

Although the findings of this study are entirely consistent with other studies in aviation and anaesthesia, we accept that this method has limitations, in that it is possible to assign any delayed response to a variety of factors, such as the subject making a conscious decision to press the button more slowly. This is inherent in any technique that seeks to measure a cognitive process, in the same way that measuring the effect of an analgesic drug using a verbal rating scale may be questioned because patients may choose to provide ratings that are inaccurate. However, these methods are well established in other industries, for example, a recent study used the prolonged response time to pressing a button in response to a light stimulus to conclude that conversing on a mobile phone caused an increase in mental workload while driving and that using hands-free kit did not reduce this effect compared with a hand-held unit. Their conclusion was that the increased mental workload would result in a greater likelihood of driving errors when using a mobile phone. Similarly, the data presented here suggest that although the anaesthetists’ studied did not make observable errors, their high mental workload would make errors likely if they were exposed to similar conditions in clinical practice.

One study found that 64% of anaesthetists reported errors due to excessive workload. Without validated assessment tools, we cannot measure the performance of medical practitioners and therefore cannot properly evaluate strategies to make patient care safer. This may require the use of multiple, simultaneously measured variables, including subjective and objective methods, as used in other industries rather than the subjective opinions in current use.

In conclusion, the importance of human factors to the safe practice of medicine is becoming increasingly apparent. These data are consistent with the subjects developing cognitive overload during a simulated anaesthetic crisis, although further studies are required to define the relationship between the results of this method, primary task workload, other measures of mental workload, and, most importantly, clinical performance. However, if training, education, and equipment design can be shown to reduce mental workload, the likelihood of error and therefore patient harm may also be reduced.

References
Mental workload in anaesthetists

18 Casali JG, Wierwille WW. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Hum Factors 1983; 25: 623–41