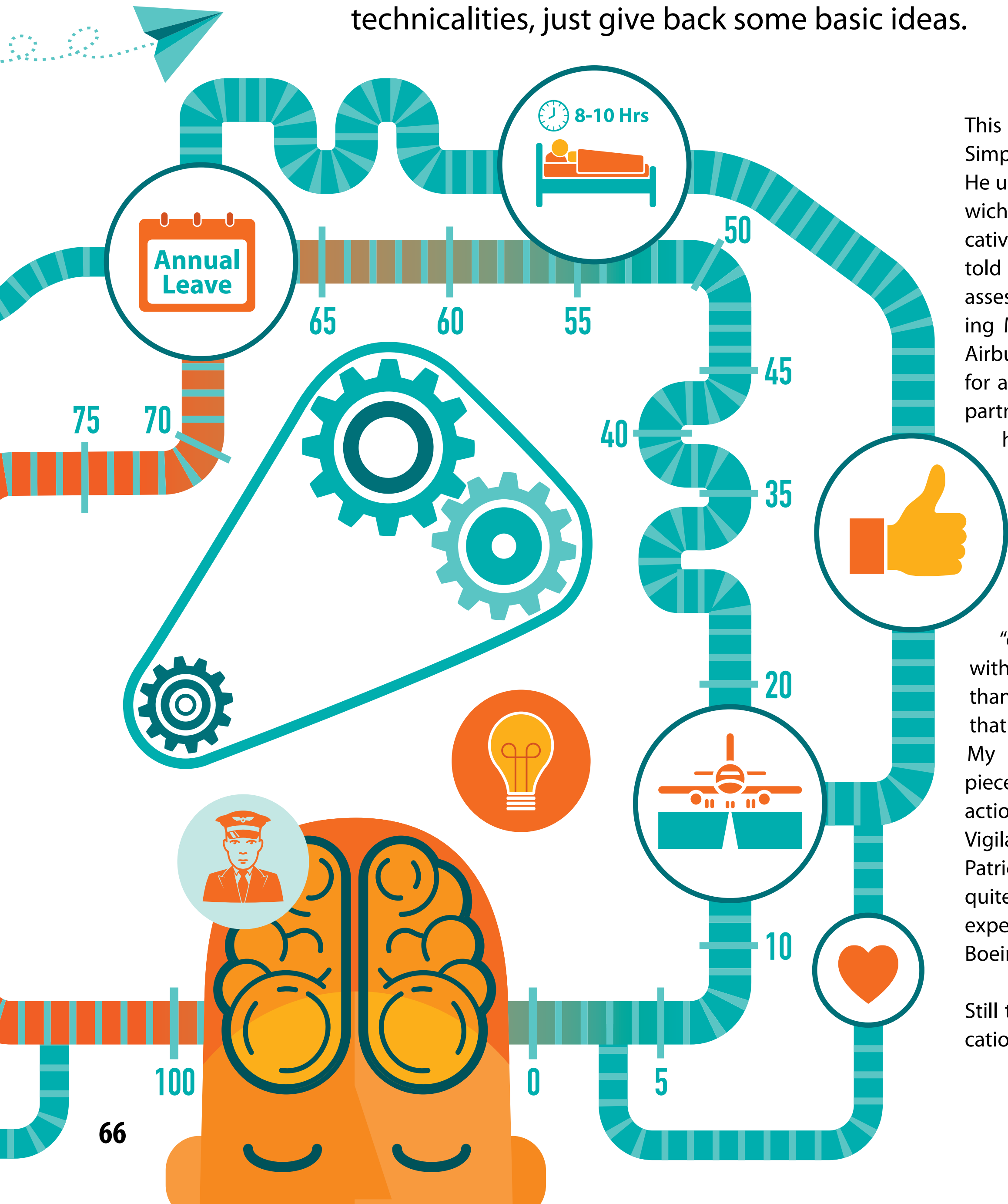
Can ATM learn from the experience of pilot workload measurement?

by Jean-Jacques Speyer

Hard-won experience from 35 years of experience with airline pilot workload measurement methods can perhaps shape the future of Air Traffic Controller workload measurement in the context of SESAR & ATM just by sparking some ideas…We don't want to bore you with any technicalities, just give back some basic ideas.

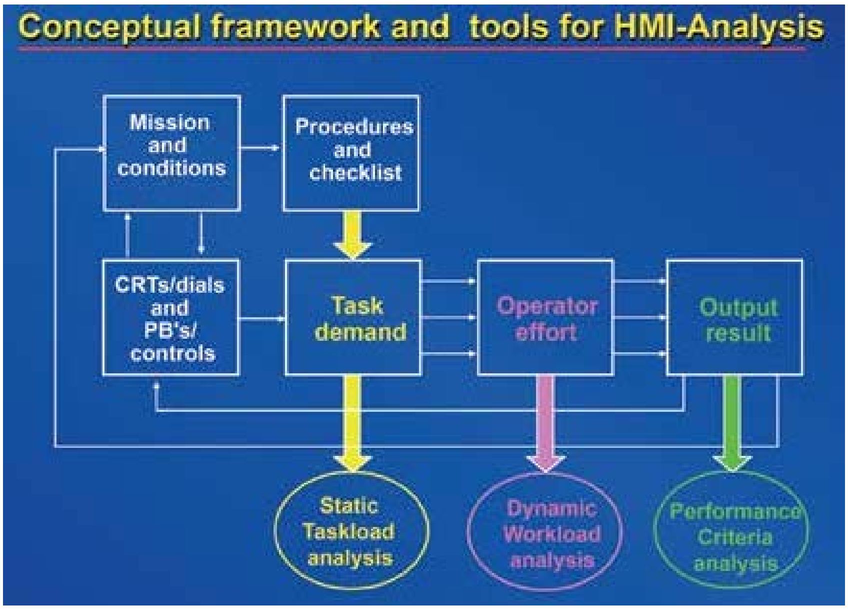


This all started with Professor Bob Simpson, my thesis supervisor at MIT. He used to meet his guys over a sandwich lunch and was pretty communicative about his own work. One day he told us about his review of workload assessment methods for the impending MD80 certification. Years later at Airbus, as I was being interviewed for a job in the Flight Operations department, I was asked "have you ever heard of pilot workload?" "No… oh, yes, now wait a minute, yes, now I remember…" Within days I was working on the design of the flight deck of the world's first two-crew wide-body aircraft, the Airbus A300FF, with a "carte blanche" to go and meet up with NASA's HF gurus who were more than happy to show headquarters that they were popular in Toulouse. My starting point was a landmark piece "A Simulator Study on the Interaction of Pilot Workload with Errors, Vigilance and Decisions", by the late Dr Patrick Ruffell Smith. His work coupled quite well with my personal practical experience of workload from time as a Boeing 707 Flight Engineer.

Still to this day minimum crew certification under FAR 25 and its Appendix

D stipulates both workload functions and factors that need to be assessed and documented but do not suggest any means of compliance. After proposing a simple framework to my management, which the airworthiness authorities readily accepted, we could start work in earnest to develop our own evaluation methods.

The overall idea was to compare a new aircraft to be certified for two-crew operation with existing ones – at the time already well – proven in actual airline service – to assess whether the new aircraft footprint would be within the envelope of the older design. Methods were developed just by "doing it" – by intuition.



Static Taskload Analyses considered the Normal, Abnormal & Emergency system procedural tasks that had in the past been carried out by a Flight Engineer. These indicated that taskload of the aircraft under evaluation for certification would be within the envelope defined by the reference aircraft (B737/DC-9), an early indication of acceptable two-person crew operations on the new aircraft. This work also enabled first hand task-sharing evaluations in mockups with early sets of procedures not yet subject to flight experience. The avionics smoke procedure analysis even triggered a redesign of the A300 electrical system.

Dynamic Workload Analyses compared the timelines of workload variations under demanding scenarios with subjective ratings from each pilot using an 8-point scale derived from Cooper-Harper's scheme to evaluate *Aircraft Handling Qualities*[1] with concurrent subject pilot and observer ratings from Airworthiness Authority pilots. These demonstrated that workload ratings measured for both the A300 and for the smaller A310 were within those recorded for B737/DC-9's. Overall, workload profiles were in the moderate range and crews never experienced workload levels becoming extreme and unacceptable to the point that errors became inevitable. Good convergence emerged between subject pilot and observer ratings, with about 75% of ratings being identical. Calibrated and validated, the proposed rating scale was deemed usable.

Certification reports – already as early as then – emphasized the need for pilot training to include full use of ECAM messages and FMS modes, specifically predicting the need to focus on conventional aircraft handling without resorting to automation so as to maintain pilot flying skills and avoid loss of situational awareness, years ahead of issues that would become top industry concerns,

Performance Criteria Analyses compared electronic flight instruments and flight management systems respectively to electromechanical instruments and conventional navigation systems (HSI, ADI), measuring pilot performance when executing a specified and demanding circuit on the flight test A300. Another experiment compared side-stick/Fly-By-Wire (FBW) versus conventional controls by removing the conventional controls at the left side of the test aircraft and replacing those with a side-stick. Experience was gained in assessing basic measures derived from dozens of aircraft performance parameters such as for example pitch, speed, elevator position and engine power lever angle. For the FBW experiments, smoothness, stability measures and rates of reversals showed a marked improvement over conventional control. When using the side-stick, pilot control inputs were reduced by 50% or more, releasing time for other flight management duties.

Following the positive outcome of these methods, a mathematical model, the ***Airbus Workload Model***, was developed in the wake of the A310 certification and fitted the subjective rating data well. This model was a statistical combination of aircraft flight parameter data such as airspeed and roll rate with heart rate variability data on both pilots and flight status measures such as phase of flight. The model predicted the rating a pilot would have given on the workload rating scale. The overall experimental methodology was sufficiently original for a joint patent to be obtained in France, ▸▸

1 - G. Cooper and R. Harper. The use of pilot rating in the evaluation of aircraft handling qualities. Technical Report TN D-5153, NASA, April 1969.

Can ATM learn from the experience of pilot workload measurement? (cont'd)

the USA and Europe on behalf of three parties concerned[2,3].

The model gave the "subjective" nature of ratings a solid foothold as subjective measures often sounded unusual and weird, eventually strengthening acceptance of the two crew certification process. The availability of continuous modeled graphs provided a unique opportunity to examine some issues such as possible associations between workload, automation and errors; there was an indication that severe errors may occur during periods of high and increasing workload; as well as the suggestion that these take place near surges and peaks of workload. Workload decreases noted thereafter appeared as pilots procrastinating, taking tactical postures, at times even shedding tasks. We also found that just as for errors, plotting workload graphs and automation levels pointed towards a classical inverse relationship between the two.

The model gave the "subjective" nature of ratings a solid foothold as subjective measures often sounded unusual and weird, eventually strengthening acceptance of the two crew certification process. The availability of continuous modeled graphs provided a unique opportunity to examine some issues such as possible associations between workload, automation and errors; there was an indication that severe errors may occur during periods of high and increasing workload; as well as the suggestion that these take place near surges and peaks of workload. Workload decreases noted thereafter appeared as pilots procrastinating, taking tactical postures, at times even shedding tasks. We also found that just as for errors, plotting workload graphs and automation levels pointed towards a classical inverse relationship between the two.

Step-by-step, the version of this model that had been developed for the A310 was used for the A320 and was expanded for airline applications on A310, A320, B767 and B747-400 aircraft and validated for the A340 certification. It was also able to help investigate the impact of automation on crew underload in long-range operations in parallel with dedicated **crew alertness measurements** to help formulate recommendations to cope with crew fatigue.

Although there had been much research on pilot workload, virtually all of this had been focused on overload. Hence we progressed to measuring crew alertness in actual airline operations to complement workload plots calculated from the model using DFDR data and pilot ECGs.

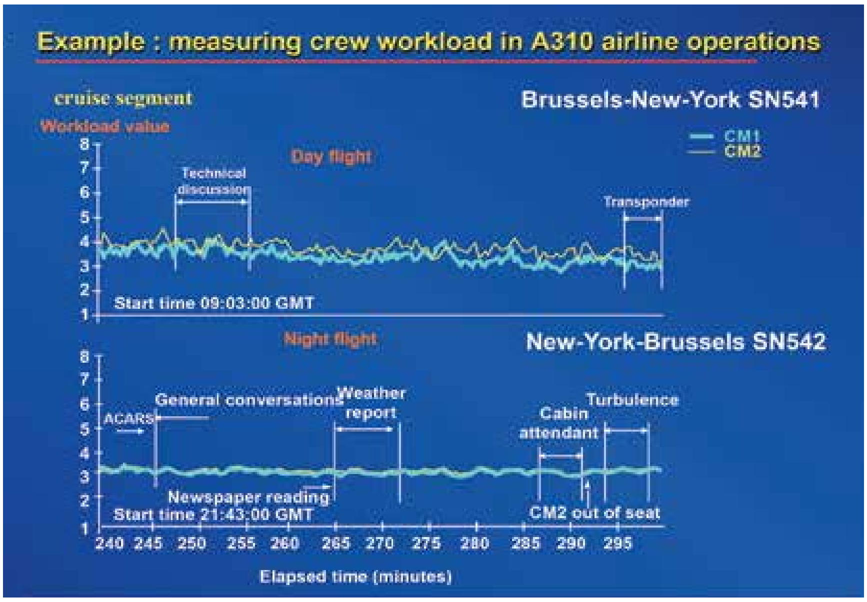
It was found that for short periods, both pilots would cooperate in shared tasks and that these would be separated by longer periods where pilots were effectively disconnected from one another. To preserve crew vigilance, the concept of pilot alternation emerged, crewmembers taking turns to enable napping "in the seat" to reduce sleep pressure. Pilot ECGs showed that low activity with relatively flat workload curves were typical of night flights whereas a saw-tooth pattern characterised daytime flights. This suggests that lower workload variability could act as a predictor of low vigilance or under-alertness. Unsurprisingly, high pilot workload was calculated when arriving at busy and unfamiliar airports after long overnight flights. One night we even recorded a precautionary engine shutdown after erratic oil pressure fluctuations just after a near miss coupled with ATC coordination problems – the activity ratings surged. This was seen as "proof of the pudding" as far as the model's reliability was concerned given that it had been formulated during certification testing where pilots knew in advance there would be problems to be faced.

The Dynamic Workload Scale

WORKLOAD ASSESSMENT		CRITERIA			APPRECIATION
		Reserve Capacity	Interruptions	Effort or stress	
LIGHT	2	AMPLE			VERY ACCEPTABLE
MODERATE	3	ADEQUATE	SOME		WELL ACCEPTABLE
FAIR	4	SUFFICIENT	RECURRING	NOT UNDUE	ACCEPTABLE
HIGH	5	REDUCED	REPETITIVE	MARKED	HIGH BUT ACCEPTABLE
HEAVY	6	LITTLE	FREQUENT	SIGNIFICANT	JUST ACCEPTABLE
EXTREME	7	NONE	CONTINUOUS	ACUTE	NOT ACCEPTABLE CONTINUOUSLY
SUPREME	8	IMPAIRMENT	IMPAIRMENT	IMPAIRMENT	NOT ACCEPTABLE INSTANTANEOUSLY

2 - Dunlap & Associates, the Cochin Laboratory of Adaptation Philosophy and Airbus.

3 - Speyer JJ, Fort A, Blomberg R.D , Fouillot JP,"Assessing Workload for Minimum Crew Certification", in AGARD AG-282, "The practical Assessment of Pilot Workload, Ed. By Dr A.Roscoe,, June 1987



Work on active/passive pilot alternation also informed a means to monitor the awake pilot during the other pilot's in-seat napping on long haul flights and warn him/her, if necessary, of his/her own impending sleepiness. This concept of electronic pilot activity and alertness monitoring depended on two inputs:

- pilot activity monitoring by detection of their interaction with different aircraft systems such as the FMS, ECAM, EFIS FCU and RMP ,
- pilot eye-tracking because of the possibility that pilots would still manipulate systems when in a low alertness state,

The concept also goes along with present philosophy on pilot flying/pilot monitoring.

Evaluation on selected flights showed that:

- Cameras could be used to obtain some measure of the cognitive part of pilot workload since filmed eyelid closures correlated with alertness measures as it would clearly be impracticable to 'wire-up' line pilots to EEG recorders in routine airline operations,
- Cameras could be used to record what instruments pilots were actually looking at – although this was and remains taboo for normal line flying purposes,
- EEG recordings confirmed that decreased alertness would get back to higher levels if a pilot took on a cognitive workload coupled with physical interface activity, within a limited time and sufficiently vigorously that situational awareness could be regained. A finding we could verify for the cruise parts.

Finally, we undertook Minimum Crew Certification for the A380. This was quietly and smartly performed by means of NASA's TLX index. Its 6 dimensions – mental, physical and temporal demand, frustration, effort and performance, were measured using hand held devices to capture pilot ratings. After calibration these could assess multi-attribute tasks involving tracking work, systems monitoring and scheduling work, communications, systems status work and resource management, so as to assess all elements of workload acceptability using just a handful of scenarios.

This process led to the finding that on all long-haul flights without augmented crews, taskload tended to rate higher since longer time on task would deplete more of their physical and mental reserves. It was also found that in the case of the A380, as with all the other types, CRM breaks down under extreme workload. When cognitive bias overpowers the PF, the other pilot has to take over. But what if both pilots are under pressure that narrows their workload capacity? How then can we avoid situational awareness sliding further away?

So, what does all this say for the world of the ATC Controller? Perhaps some ideas are that:

- Their work could be monitored; just as we had observer-pilots rating workload and describing flight scenarios, can we imagine a similar procedure for future ATM operational evaluation and certification using an equivalent of the **Aircrew Data Logging** software.
- Eye tracking could be used to see what the controller looks at and record eye movement sequences eyelid or eye-pupil devices to assess alertness & workload
- ATC strategies & tactics could be mapped, workload models would soon emerge and discriminations like taskload, workload & performance could be made!

Jean-Jacques Speyer

Jean-Jacques, MS El & Mech Eng, VUB, Brussels, Belgium, MS Aero & Astro, MIT, USA, DU HF's & Ergonomics, Univ René Descartes, Paris, France. After an airline start as B707 Flight Engineer, JJ Speyer moved on to Flight Operations Engineering & Certification at Airbus from where he retired as Senior Director Airline Consulting, Fuel & Flight Efficiency at the end of 2009. He is currently Professor in Aircraft Specification & Certification at the Vrije Universiteit Brussel.