

THE PLACE OF PSYCHOMETRIC TESTING IN PILOT SELECTION

Peter Clem

BSc, DipMilAvn, PostGradDipSci, MBBS,
PostGradDipAvMed, RACGP

Corresponding author:

Peter Clem

peterclem.feedback@gmail.com

ABSTRACT

Aim

Pilot training has always been a relatively expensive undertaking. So attempts to control these costs by predicting the likelihood of success or failure is a constant that is almost as old as aviation itself. Incorporation of Psychometric testing was made to pilot selection in the years between the first and second world wars. Despite the many changes that have occurred in this area, psychometric testing continues to feature in modern systems of pilot aptitude testing. This paper reviews some of the history of psychometric testing in pilot selection.

BACKGROUND

After World War One efforts to predict success in pilot training were occurring both in Europe and the United States. In Europe, the French and British approaches tended to focus on the Physiological challenges of aviation. In the United States, the approach tended to focus more on the psychological difficulties. This dichotomous approach continued well into the Second World War. It was not until the landmark Pensacola 1000 study in the United States in 1945 that demonstrated the superiority of the Psychological approach. (1)

900 US Navy flight training cadets were subjected to 60 different psychological, psychomotor and physiological tests. The Pensacola 1000 study determined that the physiological tests were not predictive of success more than chance. The study concluded that psychometric and

psychomotor tests were predictive of success on flight training.

The Pensacola 1000 study became the model for pilot psychometric testing from 1945 until the present day. The research led to the creation of the Naval Aviator Test Battery. The Naval Aviator Test Battery included the Wonderlic Personnel Test (a test of general ability or intelligence), the Bennett Mechanical Comprehension Test (a test of mechanical interest and skills), and the Purdue Biographical Inventory (a measure of morale, interest, and attitudes).

PSYCHOMOTOR TESTING

The Pensacola 1000 study demonstrated that psychomotor tests were predictive of success in pilot training. Despite this result, Psychomotor tests had problems with ease of use, reliability, and standardisation. As a result, psychomotor tests were omitted from the Naval Aviator Test Battery. Psychomotor tests fell out of use in the United States in the decade after the Second World War. (1)

Outside of the United States, the experience was different. The Royal Air Force (RAF) and the Royal Australian Air Force (RAAF) persisted with electromechanical psychomotor tests from the 1940s until well into the 1990s.

Pilot aptitude testing has also included various combinations of other measures. These have included previous flight experience, previous service experience, interview results, performance on work sampling, and results at flight screening. This has been an attempt to boost the relatively modest predictive power of psychometric and psychomotor testing. (2, 3)

DEFENCE MECHANISM TEST

In the 1970s and 1980s, Scandinavian countries developed the Defence Mechanism Test (DMT), which was a significant departure from the approach that had developed following the Pensacola 1000 study. The DMT is a projective test based on assumedly anxiety-provoking images. The images are exposed through a tachistoscope. The tachistoscope gradually increases the image exposure from 5 to 2000 milliseconds. The rationale for the test as a

selection instrument for stressful occupations is that psychological defences bind psychic energy necessary for coping with stressful situations. Furthermore, those subjects with maladaptive strategies for dealing with stress will perform worse on the test. The Scandinavians reported significant predictive ability for the DMT. This approach was trialled by Air Forces outside of Scandinavia, including the RAAF. The DMT failed to demonstrate the same results. A study by Ekehammar et al. in 2005 aimed to understand why this was so. They concluded that the DMT does not measure what it purports to measure. They found that a more plausible explanation was that DMT performance reflects information processing difficulty due to anticipatory or test anxiety. (4)

COMPUTER TECHNOLOGY

Since the 1980s computer technology has been introduced into pilot aptitude testing. Various systems have been deployed which combine psychometric and psychomotor tests into the one device. (1) It is important to realise that these machines are based on electronic versions of pre-existing psychometric and psychomotor tests. Because of this, Bartram et al. 1995 concluded that this technology is not expected to significantly enhance the prediction of success or failure on pilot training. (5)

A strength of computer-based devices is that the Psychomotor element doesn't suffer from the issues that plagued electromechanical devices of the past. This led to the United States military reintroducing psychomotor testing into their test batteries in the 1980s. Computer-based systems have thus enabled reintroduction and combination of psychometric and psychomotor testing into a single device. The widespread uptake of these devices by military and non-military users around the globe has helped to standardise the assessment process.

STATISTICAL ISSUES

Although there has been widespread and continuous use of psychometric testing over a very long period, the predictive abilities of these tests have always been modest.

A landmark meta-analysis study by Hunter and Burke in 1992 reported validity coefficients as a function of predictor type (table 1). (6)

Predictor Measure	Mean r
General Cognitive	0.1924
Personality	0.1168
Information Processing	0.2256
Job Sample	0.3272
Biographical Inventory	0.2646
Psychomotor Co-ordination	0.3035
Composite Battery	0.1934
Other	0.0889
Total	0.1973

Table 1. Validity Coefficients as a Function of Predictor Type

The researchers reported that in general, job sample measures were the best predictors of performance, followed by psychomotor co-ordination and biographical inventories.

Somewhat depressingly, Hunter and Burke reported that the analysis showed a decline in the mean validity correlations over the previous 50 years.

Another disappointing finding was that for the personality measures (mean correlation of 0.1168) the 95% confidence interval was +/- 0.2644. An interval which includes zero meaning that this measure is no more predictive than a coin toss. (6)

SO WHY ARE THESE TESTS NOT ABLE TO BE MORE PREDICTIVE?

In a 1996 paper by Damos et al, *Pilot Selection Batteries: Shortcomings and Perspectives*, the authors noted the low correlation between predictors and outcome criteria as described by Hunter and Burke. They noted that predictive validities based on intelligence tests and personality tests were in the range of 0.15 to 0.20. Damos noted that selection batteries that combined intelligence, psychomotor, personality and information processing tests could achieve predictive validities in the range of 0.20 to 0.40. (7)

Damos offered the following list of potential explanations for why these tests aren't more predictive: (7)

1. **Sudden workforce changes.** Leading to military altering criteria for pass-fail and thereby adversely affecting the correlations.

2. The use of pass-fail criteria. Dichotomising the criterion variable at the mean results in a 38% reduction of effective sample size when the correlation is between 0.20 and 0.50. A high success rate in pilot training effectively limits the biserial point correlation between the predictor and the criterion variable.

3. Test development. Historically the tests were not based on task analysis but were assumed a priori to have some validity for predicting success in pilot training. Many of the tests were based on psychological theories of human cognition and personality which may or may not play a significant role in performing such a complex task as flying an aircraft.

SO WHY DO MILITARIES AND AIRLINES CONTINUE TO UNDERTAKE THESE TYPES OF TESTS?

Although psychometric tests are unable to provide significant prediction in isolation, when they are combined into selection batteries, they provide increments in prediction that continue to be attractive to organisations that are responsible for candidate selection. (2, 3) A common and interesting observation is that correlations with success on pilot training are not reflected in success on operational training. (7)

As long as pilot training continues to be expensive and while there is a large number of applicants for a small number of training places it is likely that this approach will continue, despite its limitations. On the other hand, if there becomes a severe shortage of pilots (as predicted by ICAO), the limitations of this approach may become more apparent.

AIRLINE INVOLVEMENT

Many airlines around the world have incorporated psychometric testing into their selection processes. Somewhat paradoxically, United States airlines have not employed these tests as much as many overseas Airlines, due in part to the particular regulatory framework in which they operate. (7)

The International Aviation Transport Association (IATA) have published *Guidance Material and Best Practices for Pilot Aptitude Testing*. (8)

These guidelines make the following claims for pilot aptitude testing;

- “If correctly implemented, a pilot aptitude testing system can contribute to considerable cost savings for the airline as well as:
- Enhanced safety
- Decreased training costs
- Increased training and operational performance success rates
- More positive working environment
- Reductions in labor turnover
- Enhancement of the flight operations department and airline’s brand”

The guidelines are based on a large survey of the practices of member airlines. Although the guidelines make significant claims for the benefit of pilot aptitude testing, they do not address some of the limitations identified in research journals previously identified in this paper. The guidelines do note the following;

“Aptitude testing systems are not "perfect" in predicting the future performance of pilots. However, if they are developed and designed responsibly, they can offer valuable guidance to the operator. There is consensus amongst experts in the field of aptitude testing that performance of pilots can be reasonable well predicted employing testing. Opinions differ on a) how long the predictions are valid, b) which category of performance can be predicted best and c) how detailed the prediction can be”. (8)

As previously noted, research has indicated that correlations with success on pilot training are not reflected in success on operational training. So it’s worthwhile to consider to whom the airlines are applying these IATA guidelines. If they are applied to airline cadets then based on the military experience there they might predict success with training.

If the test batteries are being applied to pilots being recruited from the military, or other airlines, then this expectation is probably not realistic. Alternatively, airlines may expect that (for qualified pilots) the batteries may select candidates who are ‘safer’ or more compatible with the ‘culture’ of the company. The evidence base for this expectation is not currently well described in the scientific literature.

THE FUTURE

Since the 1980s there has been a convergence of psychometric and psychomotor testing in terms of their incorporation into computer-based devices.

REVIEW

At the same time computer technology has been incorporated into the cockpit with resulting automation of critical roles. As these processes continue the screening device, the simulator and the aircraft may converge to the point that they are largely identical from the view point of the "pilot". At this stage very little if any selection or training will take place in real aircraft.

A survey by financial services firm UBS estimated that moving from 2 pilots to 1 pilot for airline operations would yield a potential profit of \$15 billion. The study also noted that 70-80% of accidents are the result of human error and that 15-20% of those are due to crew fatigue. (10) It is likely that these drivers will result in greater automation of the cockpit to the point where a pilot's role may be unrecognisable from what it is today. The level of automation may result in pilot's tasks becoming more similar to that of a Main Control Room (MCR) operator in a nuclear power plant.

A study by Zhang et al. looked at any correlation between a Psychometric measure known as general mental ability (GMA) and the performance and safety compliance of main control room (MCR) operators in nuclear power plants. The study noted that GMA is the best single predictor of work performance with the criterion related validity as high as .51. (9)

In this context, it's interesting to consider that a change in the "task" might be the missing ingredient which finally delivers on the promise of psychometric testing.

BIBLIOGRAPHY

1. Griffin GR, Koonce JM. Review of psychomotor skills in pilot selection research of the U.S. military services. *Int J Aviat Psychol.* 1996;6(2):125-47.
2. Carretta T. Pilot Candidate Selection Method: Still an Effective Predictor of US Air Force Pilot Training Performance. *Aviation Psychology and Applied Human Factors.* 2017.
3. Weissmuller JJ, Damos DL. Improving the Pilot Selection System: Statistical Approaches and Selection Processes. *The International Journal of Aviation Psychology.* 2014.
4. Ekehammar B, Zuber I, Konstenius ML. An empirical look at the Defense Mechanism Test (DMT): reliability and construct validity. *Scandinavian journal of psychology.* 2005;46(3):285-96.
5. Bartram D. Validation of the Micropat Battery. *International Journal of Selection and Assessment.* 1995;3(2):84-95.
6. Hunter D. Predicting Aircraft Pilot- Training Success: A Meta-Analysis of Published Research. - [ijap_meta_analysis_of_pilot_selection.pdf](#). 1992.
7. Damos DL. Pilot selection batteries: shortcomings and perspectives. *Int J Aviat Psychol.* 1996;6(2):199-209.
8. Association IAT. *Guidance Material and Best Practice for Pilot Aptitude Testing.* Montreal Geneva: International Air Transport Association; 2012.
9. Zhang J, Yongjuan L, Changxu W. The Influence of Individual and Team Cognitive Ability on Operator's Task and Safety Performance: A Multilevel Field Study in Nuclear Power Plants. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084528>
10. Would you get on a pilotless plane? BBC 8 August 2017. <https://www.bbc.com/news/business-40860911>