PILAPT Science Fact Sheet

Agnes Kokorian
People Technologies
2010

www.pilapt.com

INTRODUCTION

The information contained in this fact sheet is organised in two parts. The first part presents background information on each of the **PIL**ot **APT**itude (**PILAPT**) tests. The second part presents key details of the psychometric properties of the tests.

PART 1: WHAT DOES PILAPT MEASURE?

In this section, the reader will find a description of the constructs and background on the models and theories used in the design of the individual PILAPT tests. The information is presented in two sections reflecting the structure of the PILAPT battery: PILAPT Base, a series of tests to assess the core abilities identified as consistent predictors of pilot training success, and Capacity, a mini-battery of three tests in one, to assess performance under increasing workload. More detailed information on the tests is available to the PILAPT clients.

PILAPT BASE

PILAPT Base measures core abilities related to flying performance shown to consistently predict pilot training performance in civilian and military settings. The current tests are Deviation Indicator, Concentration, Hands, Patterns and Trax. The administration time for the PILAPT Base battery is around 50 minutes. New tests have recently been added to PILAPT Base as part of the ongoing PILAPT R&D programme.

Deviation Indicator

The design of the Deviation Indicator (DI test) was based on classical designs of psychomotor tests for pilot selection dating back to the work undertaken by the Royal Air Force (RAF) and the United States Air Force (USAF) during the Second World War (see Hunter and Burke, 1994, for a review of early research in this area). Tests such as DI are sometimes referred to as simple psychomotor tasks as the candidate is required to compensate only for movements in a two dimensional display as compared to more complex measures of psychomotor ability in which information is presented in a three-dimensional display. Tests such as this have a strong history of validity in predicting pilot training success (Hunter and Burke, 1997). Deviation Indicator provides a measure of hand-eye coordination in a 2 dimensional environment. It is a compensatory tracking task requiring the use of a joystick.

Concentration

The design intent of Concentration was to combine elements of time sharing and selective attention since these are often involved in typical pilot and air traffic controller tasks. Time sharing is defined by Fleishman and Reilly (1992) as the ability to shift back and forth efficiently between two or more activities or sources of information. It involves using information from more than one source, such as monitoring multiple signals and targets. They contrast this ability to selective attention which they define as the ability to concentrate on a task while ignoring distracters, such as scanning a display for infrequent targets. In Concentration, the candidate is required to monitor multiple signals related to a primary task while also discarding incorrect or distracting information. As such, the cognitive processes involved relate to resource theories of attention such as those proposed by Kahneman (1973), Navon and Gopher (1984), and Norman and Brobow (1975), while also accepting Allport's (1980) view that attention may draw upon several resources and maybe subject to a number of "hardware" constraints.

Hands

Carretta, Perry and Ree (1996) report a strong relationship between spatial reasoning ability and working memory and the situational awareness of F15 pilots in the United States Air Force. Hands was designed to tap the spatial element associated with situational awareness, and one of the three key facets of spatial reasoning identified by Lohman (1996) linked to working memory and general intelligence. The test also builds on the work of Baddeley (2003) by presenting the candidate with an auditory rule followed by a visual task to be completed using that rule. As such, the test is intended to engage both the auditory and spatial components of working memory.

Patterns

Patterns was designed to tap into one of the three facets of spatial abilities identified by Lohman (1996) and hypothetised to be related to the ability to recognise and interpret critical information masked by other information (or background noise) under time pressure, as might be experienced by a trainee pilot at key stages in training where they are expected to assume greater control of an aircraft, and in situations such as emergencies. The content was modelled on tests designed to measure flexibility of closure (Carroll, 1993; Fleishman and Reilly, 1992) which is variously defined as the ability to identify or detect a known pattern (figure, word or object) that is hidden in other material. Those such as French (1954) and Thurstone (1944) have linked flexibility of closure to the ability to demonstrate selective attention. The appearance of the items is similar to that of embedded figures tests (e.g. Witkin et al. 1971) but the items are designed to be more complex than typical embedded figures tests to increase the cognitive load under time constraint.

Trax

The Trax test was designed to measure pursuit tracking which can be defined as the ability to accurately track a moving object. It was also designed to present the candidate with a simulation of a three-dimensional environment. This class of tests has a strong history of validity in predicting pilot training success (Hunter and Burke, 1994) from Fleishman's research with the United States Air Force following world War Two (Fleishman, 1956) through more recent research with the Royal Air Force (Burke, Hobson and Linsky, 1997).

New tests

The following tests have been developed more recently to strengthen the assessment of information processing and spatial abilities.

Sequences

The design of Sequences combined two constructs as defined by Fleishman and Reilly (1992) from which publication the following definitions are taken:

- Information ordering is the ability to correctly follow a rule or set of rules specifying how to arrange things or actions in a certain order
- Perceptual speed is the ability to compare letters, numbers, objects, pictures, quickly and accurately. The stimuli to be compared may be presented at the same time or in succession.

Fleishman and Reilly (1992) mention typical tasks associated with these two abilities as rapid identification of target aircraft, and following a check out procedure in operating equipment. Sequences was also designed with Air Traffic Controllers in mind as the cognitive processes associated with the tasks given by Fleishman and Reilly are frequently given in task analyses of ATC and pilot tasks.

Views

Views was designed to measure visualisation, one of the three facets of spatial abilities identified by Lohman (1996) and defined by Fleishman and Reilly (1992) as the ability to imagine how something will look when it is moved around or its parts are moved or rearranged. Views is the most recent PILAPT test and was added to strengthen the coverage of spatial ability as well as to provide a measure related to a trainee pilot's ability to develop the mental models required to acquire skills in positioning an aircraft and adjusting its position, as well as specific skills for pilots such as formation flying in military flying.

CAPACITY

A key area that has been emphasised in recent pilot selection research has been the need for measures of multiple task performance. The Capacity mini-battery has been developed to meet this need. The administration time for Capacity is around 15 minutes.

The design of Capacity was based on resource theories of attention such as those proposed by Kahneman (1973), Navon and Gopher (1984) and Norman and Brobow (1975), while also accepting Allport's (1980) view that attention may draw upon several resources and may be subject to a number of "hardware" constraints. Essentially, the Capacity test is framed in the context of a flying or psychomotor (primary) task to be executed alongside two additional tasks, one a visual recognition task (sharing similar though not necessarily identical resources with the psychomotor task) and a sequential auditory task (simulating the task of audio information received from air traffic controllers, forward air controllers, other crew members, other aircraft or ground personnel). The output of Capacity presents the performance of the candidate across the three conditions of single, dual and triple task.

Capacity was designed for use with military candidates to specifically identify potential for fast jet roles. In civilian pilot selection, it is widely used to assess more experienced (direct entry) candidates. Many clients use Capacity in conjunction with PILAPT Base to obtain a rounded picture of the candidate potential.

PART 2: PSYCHOMETRIC PROPERTIES OF PILAPT

In this section, the reader will find details of the reliability and validity of the tests from information gathered through a research and development programme that is now in its second decade and that encompasses a range of client sites including military and civilian pilot training programmes in Asia, Europe, Middle East and South America.

RELIABILITY

The reliability of a test or scale provides information relevant to how accurate a test or scale is by stating the proportion of variation (or variance) in test scores that can be attributed to the design and functioning of the test or scale. The score on a test will be a factor of how well a test functions as a measure of a specified construct (in the case of PILAPT, a specified ability relevant to effective performance as a pilot) and error. Errors may represent factors related to the person sitting the test (such as time of day, emotional condition when sitting the test), conditions of administration (such as the behaviour of an administrator, the environment in which the tests are administered) or poor test design. The higher the reliability of a test is, the greater the variation in test scores is a factor of effective test design and good measurement rather than errors or noise in test scores.

Reliability can be assessed in many ways, but two of the most widely used approaches are relevant to PILAPT tests. **Internal Consistency** estimates of reliability provide information on how consistent the elements of a test function together. It is appropriate when a test

contains discrete elements such as independent questions or items, and higher reliabilities indicate that the elements in the test provide a consistent contribution to the overall score on the test. **Stability** or **Test-retest Coefficients** provide information on how stable a score on a test is over time and indicates whether the ranking of candidates is consistent between two time periods. This index of reliability is also more appropriate when a test is not constructed from discrete elements or when some effect of learning of the tasks involved in the test is a factor of the test's purpose. For example, PILAPT has a number of psychomotor tests for which test-retest estimates of reliability are most appropriate.

What is an acceptable level of reliability? Generally, 0.7 (indicating that 70% of the variation of scores can be attributed to effective test design and consistent or stable measurement) is used a minimum for the use of a test in screening or selecting people for jobs. However, reliability is also a factor of test length and, to manage the candidate experience and factors such as fatigue, a compromise in reliabilities may be acceptable. For example, our general advice to clients is to use the overall PILAPT scores (the sum of standardised scores across the individual PILAPT tests) and to use individual scores to guide interpretations of that overall score. As such, and since the PILAPT tests are designed to be efficient and to contribute to a holistic assessment of an individual, the reader will find some reliabilities fall slightly below the 0.7 level for individual tests. In terms of the overall PILAPT score, as will be seen in the summaries presented below, the test-retest reliability exceeds 0.8.

PILAPT Test (Score)	Internal Consistency	N	Test-retest	N
Capacity			**0.72**	**146**
Concentration	**0.79 – 0.90**	**7,575**	**0.68**	**146**
Deviation Indicator			**0.72 – 0.80**	**255**
Hands	**0.88 – 0.94**	**12,261**	**0.74**	**146**
Patterns	**0.70 – 0.76**	**12,261**	**0.65**	**146**
Sequences	**0.65 – 0.77**	**1,635**	**No data yet available**	
Trax			**0.80 – 0.84**	**255**
Views	**0.63 – 0.79**	**1,800**	**No data yet available**	
PILAPT (overall score)			**0.82**	**146**

PILAPT RELIABILITIES

VALIDITY

The evaluation of validity is essentially concerned with establishing the scope for the use of test scores including evidence supporting claims for what the score measures and represents, how generalisable it is across different populations and settings, as well as evidence that the score or scores predict outcomes of value to the test user and can be shown to provide an appropriate basis for decisions such as selecting or rejecting candidates for employment or a training programme.

In this fact sheet, the reader will find evidence supporting the **construct validity** of the PILAPT tests (i.e. evidence supporting the interpretation of a score as offering a valid measure of a stated ability) and the **predictive validity** of the tests (i.e. that the scores are

related to more effective performance in pilot training programmes). The breadth of the data provided on predictive validity in terms of settings (civilian and military), and nationalities and languages also supports the **generalisability** of PILAPT test scores as an effective process for assessing pilot aptitude.

An example of evidence supporting the **construct validity** of PILAPT is shown in the table below. It relates to a Danish study with data collected in 2001 and involving four PILAPT tests - DI, Hands, Patterns and Trax - and a 15 test battery used to assess both aircrew and ATC aptitudes. Data were available across all 19 tests for a sample of 632 applicants. The content of the 15-test battery was classified according to test content in line with the classifications used by Hunter and Burke (1997) in their meta-analysis on pilot selection studies. This classification then provides a direct test of the extent to which PILAPT is measuring relevant predictor constructs. Hunter and Burke identified the following predictor constructs as being the most consistent and substantial predictor of pilot training success: numerical reasoning, mechanical reasoning, spatial reasoning, psychomotor and simulation based tests. The Danish data set did not contain psychomotor or simulation based tests, but the results clearly show that PILAPT is tapping the other predictor constructs identified by Hunter and Burke as critical to predicting pilot training success. Another construct validation including other psychomotor and simulation tests as represented in the RAF pilot selection battery was conducted with 382 applicants from the Royal Norwegian Air Force. The multiple correlation between the RAF battery and the PILAPT battery was 0.88.

Test Group	DI	Hands	Patterns	Trax	Overall
Mathematical Reasoning	***.12***	***.31***	***.37***	.06	***0.44***
Numerical Speed & Accuracy	***.11***	***.29***	***.25***	.03	***0.35***
Language	***.18***	***.14***	***.20***	.08	***0.29***
General Reasoning	***.18***	***.33***	***.51***	.11	***0.57***
Spatial	***.24***	***.38***	***.38***	***.17***	***0.53***
Mechanical	***.27***	***.35***	***.40***	***.29***	***0.55***
Memory	.05	***.23***	***.13***	-.09	***0.27***

Notes: Overall column gives the regression of the Test Group onto the 4 PILAPT tests
Correlations in bold and italicised are significant at the 0.01 level

EXAMPLE OF EVIDENCE SUPPORTING THE CONSTRUCT VALIDITY OF PILAPT

Evidence of the **predictive validity** of the tests is provided in the next table. Criterion validities are available for military pilot trainees (Brazil, Chile, Italy and UK) and for commercial pilot trainees (UK). The table shows relationships that are consistently above zero and support the generalisation of the criterion validity of PILAPT tests across settings and populations.

PILAPT Test (Score)	Criterion Validity	N
Deviation Indicator	**0.22 – 0.46**	**622**
Concentration	**0.24 – 0.31**	**390**

Hands	**0.16 – 0.29**	**622**
Patterns	**0.16 – 0.33**	**622**
Trax	**0.37 - 0.51**	**622**
PILAPT Base (overall score)	**0.40 – 0.55**	**622**
Capacity	**0.19 - 0.46**	**457**

PILAPT PREDICTIVE (CRITERION) VALIDITIES

REFERENCES

Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.) *Cognitive Psychology: New Directions*. London: Routledge and Kegan Paul.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, **4**, 829-839.

Burke, E., Hobson, C., and Linsky, C. (1997). Large sample validations of three general predictors of pilot training success. *The International Journal of Aviation Psychology*, **7**, 225-234.

Carretta, T. S., Perry, D. C., and Ree, M. J. (1996). Prediction of situational awareness of F-15 pilots. *The International Journal of Aviation Psychology*, **6**, 21-41.

Carretta, T., and Ree, M. J. (2000). General and specific cognitive and psychomotor abilities in personnel selection: the prediction of training and job performance. *International Journal of Selection and Assessment*, **8**, 227-236.

Carroll, B. S. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Chaiken, S., Kyllonen, P. C., and Tirre, B. (1999) Organisation *and components of psychomotor ability*. Report No. AFRL-HE-BR-TR-1999-0146. San Antonio Texas: Air Force Research Laboratory

Fleishman, E. A. (1956). Psychomotor selection tests: Research and application in the United States Air Force. *Personnel Selection*, **9**, 449-467.

Fleishman, E. A., and Hempel, W. E. (1954). Changes in the factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, **19**, 239-252.

Fleishman, E. A., and Reilly, M. E. (1992). *Handbook of human abilities: Definitions, measurements, and job task requirements.* Palo Alto, CA: Consulting Psychologists Press.

French, J. W. (1954). *Manual for kit of selected tests for reference aptitude and achievement factors.* Princeton, NJ: Educational Testing Service.

Hunter, D. R., and Burke, E. F. (1997). Predicting aircraft-pilot training success: A meta-analysis of published research. *The International Journal of Aviation Psychology*, **4**, 297-313.

Hunter, D. R., and Burke, E. F. (1994). *Handbook of pilot selection.* Ashgate Publishing.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Lohman, D. F., (1996). Spatial ability and G. In I. Dennis and P. Tapsfield (Eds). *Human abilities: their nature and assessment* (pp. 97 -116). Hillsdale, NJ: Erlbaum.

Navon, D., Gopher, D., Chillag, N., & Spitz, G. (1984). On separability of and interference between tracking dimensions in dual-task axis. *Journal of Motor Behavior*, **16**, 364-391.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource limited processes. *Cognitive Psychology*, **7**, 44-64.

Thurstone, L. L. (1944). A *factorial study of perception*. *Chicago, IL: Chicago university Press.*

Witkin, H. A., Oltman, P. K., Raskin, E., and Karp, S. A. (1971). *Manual for the embedded figures tests.* Palo Alto, CA: Consulting Psychologists Press