Paper presented at the 40th Applied Military Psychology Symposium, Oslo-May 2004

Pilot selection in the Norwegian Air Force: From paper-and pencil to computer-based assessment

Monica Martinussen
University of Tromsø and The Norwegian Defence Leadership Institute

Tore M. Torjussen
Olav Storsve
Olav Hjerkinn
The Norwegian Defence Leadership Institute

Abstract

The purpose of this paper was provide an overview over pilot selection in the Norwegian Air Force and to examine the predictive validity of a computer-based assessment battery currently used for pilot selection. The current test battery includes 13 tests measuring psychomotor coordination, information processing, and spatial ability. The tests were administered to a sample of applicants, and only a highly selected group was permitted to enter basic flying training, and thus constitute the validation sample (N = 108). Several criteria of pilot performance were included in this study, both pass/fail in training as well as instructor's ratings. Several test measures were significantly related to pass/fail in training and to the instructor's ratings. The combined test score was significantly correlated with both pass/fail in training (r = .26, uncorrected for range restriction) and instructors' ratings.

Introduction

Psychological tests have been used for pilot selection since the earliest days of aviation. Sustained interest in pilot selection is probably a function of the high costs associated with pilot training and operation of aircrafts, as well as concerns for safety. The results from two meta-analyses (Hunter & Burke, 1994; Martinussen, 1996) indicate that some tests are better predictors of pilot performance than others. The best predictors of pilot performance are tests measuring certain cognitive and psychomotor abilities, whereas tests measuring personality traits and general intelligence have lower predictive validity.

Military psychology in Norway dates back to World War II. The first pilots were selected for the Norwegian Air Force in 1946 (Riis, 1986). The specific tests and selection procedures have been modified over the years, and from 1998 computer-based assessment have been implemented (Martinussen & Torjussen, 1998; Torjussen & Hansen, 1999). Pilot selection in the Norwegian Air Force is conducted in several stages. The candidates who meet the formal criteria in terms of age, and who have at least 12 years of education, are admitted to the first step of the selection process. They are then tested on six paper-and pencil tests measuring cognitive abilities, mathematical and language skills. About 50% pass this stage of the selection process. The candidates are also subjected to a general physical examination (e.g., vision, hearing, height, and physical fitness). The computer-based tests are administered in

the second phase of the selection process, and it includes 13 tests. In addition, the candidates are given one Big-five measure (5PFmil; Engvik, 1993) and the Defence Mechanism test (Torjussen & Værnes, 1991). The candidates are also interviewed by a psychologist, and the last part of the selection process is a Board interview and a final medical examination. Only 10% of the candidates who are tested on the first stage are eventually admitted to Basic Flying School (BFS).

By replacing the existing mechanical and paper- and-pencil based test versions with one based on CBA technology we intended to improve the selection process both in terms of efficiency and quality. The purpose of this study was to perform the first evaluation of the predictive validity of the new computer-based tests.

Method

Participants

The participants for the validation study were 108 candidates who applied for aviation flight training. They were all between 19 and 24 years old. About half of the candidates were rejected from Basic Flying School at some time during training, primarily due to inadequate flying performance.

Procedure

All subjects were tested before entering the flight program, and the instructions were presented to the subjects on the computer before each task. The test battery took about three hours to administer. The order of the tests was the same for all subjects, and the subjects were given a short break, half-way through the testing session. All instructions were presented to the subjects on the screen before each task individually. The criteria were collected during the period with basic flying training and only for the selected group.

Tests

All the tests included in the test battery are computer-based, and programmed by Psytech LTD, London (Burke, Kitching, & Valsler, 1997). The system includes 12 tests, four from the PILAPT battery (Trax, Deviation Indicator, Hands, and Patterns) (PILAPT Handbook, 1999), five tests developed originally by the Royal Air Force (Control of Velocity, Sensory-Motor Apparatus, Instrument Comprehension, Digit Recall and Vigilance) and three tests based on paper-and pencil tests from the previous Norwegian pilot selection battery (Planes, Attention and Numbers) (Torjussen & Hansen, 1999). One computer-based test called Determinasjonsgerät (DTG) is administered in addition to the main tests, and on a separate computer. DTG is part of the Vienna Test System developed by Schuhfried.

Control of Velocity (CVT-600). The test is a 1-dimensional pursuit tracking task (psychomotor test). The candidate uses a joystick to move a pointer horizontally across the screen to hit targets as they descend from the top of the screen. The score is based on the number of hits.

Sensory Motor Apparatus (SMA-610). This is a two-dimensional compensatory tracking task. The candidate uses both a joystick and rudder pedals to move a pointer both horizontally and vertically to keep the pointer as close as possible to a cross in the middle of the screen. The pointer is during the test session moved away from the

middle and the candidate has to use the controls to compensate for this. The candidate's score is based on their deviation from the centre of the screen.

Deviation Indicator (620). The test is a compensatory tracking test requiring fine motor coordination. The task is to keep two lines, one vertical and one horizontal, centred on the screen. The candidate has to compensate for simultaneous horizontal and vertical deviations from the centre target. The score is based on the number of errors in controlling the task. The test-retest reliability reported in the test manual was .80 after a time interval of four months (PILAPT Handbook, 1999).

Trax (630). The test is a pursuit tracking task modelled on the idea of flying down through an ILS glide path. In contrast to other coordination tests, Trax requires adaptation to the dynamics of the three-dimensional flying environment. The test assesses a combination of psychomotor co-ordination, spatial and information processing abilities. The score is based on the number of errors made when controlling the task. The test-retest reliability reported in the test manual was .80 after a time interval of four months (PILAPT Handbook, 1999).

Planes (640). This is a spatial ability test involving visualization and mental rotation. A picture of an airplane is displayed in one of the corners of the screen either flying towards you or away from you and in four different orientations: with a left or right bank, straight level or upside down. A target is simultaneously shown in a diagonal corner, and the task is to decide as quick as possible what stick movement a phantom pilot will have to make in order to move the plane in the direction of the target, using the joystick. The score is number of correct answers, giving credit to short reaction times.

Instrument Comprehension (INSB-650). This test is a measure of general reasoning and spatial reasoning (altimeter, artificial horizon, velocity indicator, air speed, compass, turn and bank) all of which are fully explained to the candidate. In the first half of the test the candidate is presented with all six instrument dials and a number of verbal descriptions of an aircraft's orientation. The candidate is required to use the instrument readings to identify which of the descriptions reflects the aircraft's correct position. In the second half of the test, the candidate uses just two dials (artificial horizon and compass). Here the candidate must identify the correct orientation of the aircraft from a given set of pictures. The score is the number of correct answers and time used to complete the test.

Hands-660. The test requires the candidate to use an audio search message to scan visual objects and declare how many of those objects meet the search criteria. The candidate has to translate verbal material into visual material, and the task is essentially a basic working memory task testing how quickly someone can move from verbal to visual material and make an accurate decision. Each item comprises an auditory message (positive-negative, left or right, circle or square). The candidate has to mentally store spoken information and use it to interpret visual information. Responses (e.g., how many men are carrying a circle in the left hand?) are made through a keypad. The man can be displayed in four different orientations. The time allowed to respond to each item decreases throughout the series. The score is based on both speed and accuracy. The test-retest reliability reported in the test manual was .77 after a four months time interval (PILAPT Handbook, 1999).

Attention-670. This is an information-processing test, which requires rapid apprehension of the up-down and left-right dimension. A 4x4 matrix with quadrants filled with different colours and digits is presented together with an auditive command. The information in the command identifies according to certain rules, a) the target quadrant in the matrix, and b) which aspect of the quadrant (colour or digit) that should be focused on. The commands are given with accelerating speed and the matrixes are shown with increasingly shorter exposure times. The score is based on the accuracy of the responses.

Digit Recall-680. This is a test of short-term memory. The candidate is presented with a set of numbers of varying size. Once the numbers disappears, the candidate is required to type in the number. The score is based on the number of digits entered correctly.

Numbers-700. The test is measuring perceptual speed and is associated with quick, accurate and efficient use of visual perception as well as working memory. Numerous digits of varying size, spread over the screen is presented four times, and the task is within a limited time frame to hit as many digits as possible in each of four different sequences. The time at disposal decreases for each sequence. The score is the number of hits over the sequence.

Patterns-710. The test is a measure of perceptual closure requiring the candidate to identify a target shape in patterns, which represent visual clutter. The task also includes time pressure to provide an assessment of selective attention and decision making. The test comprises 10 items and both speed and accuracy is scored (PILAPT Handbook, 1999).

Vigliance-720. The test measures attention and the candidate is presented with a nine by nine matrix. Each cell in the matrix may be identified by the numbers running along the top and to the left of the matrix. The candidate has to attend to two distinct tasks, a routine task and a priority task.

Determinasjonsgerät (DTG)-690, (Vienna Test system, Schuhfried). Measures reactive stress tolerance, attention deficits and reaction speed in the presence of rapidly changing and continuous optical and acoustic stimuli. The test requires production of continuous, sustained rapid and varied reactions to rapidly changing stimuli.

Criteria

Several criteria were included in this evaluation, both pass/fail in basic training and instructor's ratings. The assessments were conducted using a behavioral anchored rating system. The abilities and skills rated for each candidate were selected based on a previous job-analysis of 43 experienced fast jet pilots from Canada, Norway, and the United States (Carretta, Rodgers, & Hansen, 1996). The purpose of the job analysis was to determine the relative importance of abilities and personality characteristics for fast jet pilot performance. A total of 27 abilities were identified and ranked. From these abilities, 17 abilities and personality characteristics mostly with the highest mean ratings were chosen. The range of each four-point scale was identified by two endpoints expressed in terms of observable behaviour. For example,

Situational awareness is identified by “attentive and alert to situational changes vs. inattentive and fails to notice relevant information”. Another example is Cooperative, which is defined by the endpoints “Cooperates well, willing to help others, a team player” versus “neither cooperatives nor helps others”.

Results

A principal component analysis (Varimax rotation) of the 17 ratings (BARS) was conducted and it resulted in two factors (eigenvalue > 1.0). The first factor consisted of items related to flying skills, whereas items with high loadings on the second factor were more related to personal skills like cooperation and leadership ability. The items were combined into two scales consisting on 12 and 5 items respectively. Cronbach’s alphas for the two scales were .97 and .85. The correlations between the pass/fail and the combined rating scales were high, r = .80 (Flying skills) and r = .82 (Personal skills).

To evaluate how well the tests predicted training performance, the bivariate correlations between the tests and the three criteria were computed. The correlations are presented in Table 1. In addition, the correlations for individual tests were corrected for multivariate range restriction using the computer program RangeJ (Johnson & Ree, 1994).

Table 1
Bivariate correlations between pass/fail in training (criterion), Instructors’ ratings (BARS) and the tests included in computer-based test battery

	Pass/fail N = 108		BARS Pilot skills N = 99		BARS Personal skills N = 99	
Single tests	*R*	r_c	*r*	r_c	*r*	r_c
CVT	.18*	.24	. 24*	.29	.13	.15
SMA	.17*	.22	.20*	.23	.19*	.20
Deviation	.01	.10	.20*	.30	.11	.20
Trax	.12	.17	.10	.18	.05	.11
Planes[a]	.01	.11	-.11	-.04	-.03	.00
Instrum. C.	.12	.21	.01	.08	.07	.08
Hands	.16*	.27	.05	.13	.14	.23
Attention[a]	.22*	.31	.29**	.42	.32*	.37
Digit recall	.09	.15	.09	.17	.09	.11
Numbers[a]	.00	.04	.00	.03	-.03	-.04
Patterns	-.03	.02	-.01	.05	.07	.05
Vigilance	.20*	.26	.09	.16	.10	.14
DTG[a]	.21*	.31	.24*	.33	.09	.08
Total index (All tests combined)	.26**	-	.20*	-	.22*	-

Note: [a]The sample size was smaller for some of the tests (*n* = 62-75) as some of the tests were introduced later in the test battery.
r_c = correlation corrected for range restriction
*p < .05; **p < .01 (one-tailed).

Discussion

The group of candidates included in the validation sample is a highly selected group. The observed correlations between tests and criteria are thus computed on a highly selected group of candidates and therefore will underestimate the predictive validity of the tests. The corrected correlations are better estimates of the true validity. The corrected correlations ranged from low to medium for the tests involved. The predictive validity results based on Pass/fail in training and Pilot skills were similar, whereas the predictive validity was lower for criterion Personal skills. This is perhaps not surprising as the computer-based tests are designed to measure cognitive and psychomotor abilities rather than personality traits like leadership and cooperation reflected in the criterion Personal skills.

The sample is still small and the findings need to replicated on a larger sample, however, the results are promising as several of the tests are significantly related to both pass/fail in training and instructors' ratings of performance.

References

Burke, E. F., Kitching, A. & Valsler, C. (1997). The Pilot Aptitude Tester (PILAPT): On the development and validation of a new computer-based test battery for selecting pilots. Proceedings of the Seventh International Symposium on Aviation Psychology. Colombus, OH: Ohio State University.

Carretta, T. R., Rodgers, M. N., Hansen, I. (1996). *The identification of ability requirements and selection instruments for fighter pilot training*. Euro-Nato Aircrew Human Factor Working Group. Technical report No. 2.

Engvik, H. (1993). Big-five på norsk. *Journal of the Norwegian Psychological Association, 30,* 884-896.

Hunter, D. R., & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *International Journal of Aviation Psychology, 4,* 297-313.

Johnson, J. T., & Ree, M. J. (1994). RangeJ: A Pascal program to compute the multivariate correction for range restriction. *Educational and Psychological Measurement, 54,* 693-695.

Martinussen, M., & Torjussen, T. (1998). Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. *The International Journal of Aviation Psychology, 8,* 33-45.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology, 1,* 1-20.

Riis, E. (1986). Militærpsykologien i Norge. *Journal of the Norwegian Psychological Association , 23 (Suppl. 1*), 21-37.

The Pilot Aptitude Tester (PILAPT) Handbook Edition 1.0. (1999). London: People Technologies.

Torjussen, T. M., & Værnes, R. (1991). The use of the Defence Mechanism test (DMT) in Norway for selection and stress reseach. I M. Olff, G. Godaert, & H. Ursin (Red.), *Quantification of Human Defence Mechanisms.* Berlin Springer Verlag.

Torjussen, T. M., & Hansen, I. (1999). The Norwegian Defence, best in test? The use of aptitude tests in the Defence with emphasis on pilot selection. *Journal of the Norwegian Psychological Association, 36*, 772-779.

Contact information:
Monica Martinussen
Department of Psychology
University of Tromsø
N-9037 Tromsø
Norway
Phone: +47 77644348
Email: monicam@psyk.uit.no