JCivilAvia Online First, published on 19th March 2015

Full Length Article

NOVA 1.01: a dynamic psychometric paratest for pilot candidates

Alev Banu BEKÖZ[1], Üner BEKÖZ[2], Batuhan ÇETİNKAYA[2]

1 Emergency Department, State
Hospital, Antalya, Turkey
2 ELBA Aviation Institute, Antalya, Turkey

Correspondence to
Dr Üner Beköz, ELBA Aviation Institute, Guzeloba mah.
2284 sok, 3/6, Antalya 07100, Turkey;
unerbekoz@gmail.com

Received 21 Oct 2014
Accepted 19 Mar 2015

To cite: Beköz AB, Beköz Ü, Yilmaz E, et al. JCivilAvia Published Online First:
[Day MonthYear] 2015

ABSTRACT

A dynamic psychometric test item, NOVA 1.01, is introduced to recruit better candidates for pilot training. This study revealed that the scores of the motor-mental constructs measured by the static operational computer-based tests might differ under dynamic conditions. The study also aimed to compare the score skewness of measurements in dynamic conditions to develop an understanding and to introduce dynamic scores that could be used in proposed pilot candidate selection.
Three dimensional random motion was generated by an omni vectoral random motion (ovram) generator. 41 subjects were used to calibrate the testing system and 64 volunteer pilot candidates were tested. Results suggested that this paratest is reliable and valid and can be used as a part of pilot candidate selection protocols.

Background

Today, most flight schools accept pilot candidates after they complete one of the available psychometric tests, such as those available in Turkey. Similarly, airline companies employ pilots only after they obtain a suitable report from one of the known psychometric test protocols. To some extent, this may seem sufficient to reassure schools and airline companies, however, in reality, being a pilot is complex and requires considerable skills. Above all, this action takes place in an unnatural environment for humans. As the aircrafts are developing swiftly and require the pilots to undertake increasingly challenging mental and physical tasks, conventional psychometric tests are much less able to successfully recruit candidates with the expected motor-mental qualities for piloting. This conclusion was reached after evaluating not only the common test protocols but also the background of the candidates who have passed these tests. It should be noted that the use of formative indicators for construct measurement in empirical studies is still scarce (Diamantopoulos, 2008). Almost without exception, the existing psychometric test protocols are designed to assess the level of performance at a static point (Bartram, 1984). On the other hand, according to information about human performance documented in comprehensive aviation sources, the dynamic nature of flights needs to be integrated into psychometric tests in order to recruit more suitable candidates for pilot training.
Psychometric consideration in standard tests demonstrates the existence of at least two spatial factors: visualization and orientation, and predictive validity in various aspects of perceptual-cognitive functioning (e.g., mathematics and field independence) (McGee, 1979) (Pellegrino, 1984). Apparent differences between these factors are verified by process analyses of individual differences in spatial relations and visualization ability. Information processing studies suggest multiple sources of individual differences, such as process execution speed, quality and capacity of

representation, process coordination, and problem solving strategies.
Humans are mentally and physiologically organized for standard, two dimensioned ground conditions, within a tiny layer of pressure tolerance where the vector of gravity rules everything. Rationally, mental and related motor “normals” are ordinary in this environment. For the majority of the training, flight schools acclimatize the candidates to the seemingly “unearthly” nature of the flight.
Item generation is becoming increasingly feasible for many cognitive tests; they seemingly conflict with the well-established principle of measuring persons from items with known psychometric properties (Embretson, 1999). For the psychometric tests that are used to analyze the individual differences in spatial relations and visualization ability for piloting, a dynamic item is to be established into conventional steps of the tests. Overall, the apparatus tests measuring psychomotor abilities and multiple task performance of pilot candidates turned out to be the most important predictors.
Having a normal or “over the bar” type of performance in psychometric tests does not prove the same performance in no man's land. Above all, even standard static test protocols have shown distinctions when applied to the pilot candidates from different cultural backgrounds and required revisions (Hoermann HJ, 2002). Parsons et al. argue: “While standard neuropsychological measures have been found to have adequate predictive value, their ecological validity may diminish predictions about real world functioning” (Parsons, 2008).
The mental and motor capability of a person should keep performing in the appointed course under such circumstances, and clarification should be useful for justifying the use of a test for this particular purpose (Sireci, 2007). This is not to say that the psychometric tests for recruiting pilot candidates should be held in space simulated conditions or under certain G forces; rather, they ought to be held under dynamic conditions to such a point that their predictive values remain valid.

Material and Method

The purpose of this study was first to show that the scores of the motor-mental constructs measured by the static operational computer-based tests might differ under dynamic conditions. Second, assuming that psychometric testing is the process of measuring a candidate's relevant strengths and weaknesses, we aimed to compare the score skewness of measurements in dynamic conditions to develop an understanding, and to introduce dynamic scores that could be used in proposed pilot candidate selection.
To generate random three-dimensional motion, we built an omni vectoral random motion (ovram) generator. This motion platform is commonly known as a “human gyroscope” with two gimbals. Our platform generated motion on three independently rotating gimbals with a pilot seat and computer equipment (see photograph 1).

Photograph 1

The software of the test block used in this dynamic exercise was exclusively programmed for this research. The software comprised of a simple body of psychomotor coordination in multiple tasks settings, mental concentration, and spatial abilities. As general validity criteria or the standards of performance assessments, the content, substantive, structural, and consequential aspects of construct validity were highlighted (Messick, 1995).
Construct validation was accomplished through the scores of 3 world famous aerobatic pilots with extensive piloting skills, 9 student pilots, 5 military fighter pilots, 5 airline pilots, 9 general aviation pilots (one of which was a certified flight instructor), and 10 non-pilot individuals, without regard to gender differences. Experimental tests were administered to 64 volunteer pilot candidates (43 males and 19 females, aged between 17 and 32); they had all completed the conventional psychometric tests as mandated by Turkish Civil Aviation

regulations, and had been accredited for flight training.
The software included a wide blue screen with a randomly moving red spot. The subjects were expected to pursue the red spot using two joysticks and ensure the distance to the spot was maintained in an appointed range. The distance of the spot reads on the far left side of the screen. The program then launched a series of tasks, such as simple calculations, flashing photographs, or numbers on the peripherals of the screen; the subjects were expected to answer the questions while pursuing the red spot.
Before the test began, each subject was given a briefing about the exercise. Next a Holter instrument with a pulse oxy-meter was attached to the subject and they were fastened into the ovram chair (see photograph 2).


Photograph 2

The test comprised of two general steps. The first step, a static stage, has three phases where the subject is introduced to the software and practices with the left and right hand joysticks for one and half minutes to understand the basics of the pursuit; the subject is also informed about how the apparatus works by an operator. They are then tested on their pursuit of the red spot for exactly one minute; in the third phase, the subject is tested for the cross control and the time is recorded. The subject is then alerted about the second and dynamic step along which second and third phases of the first step are repeated with different (although similar) questions. The second phase of the second step has a limited duration of one minute and the subject is expected to be swift. Thus, the total dynamic step is no longer than 2.5 minutes while the first step may last between 4–6 minutes. During the second step, the ovram is activated. The ovram and the software operate independently (see photograph 3).


Photograph 3

The motion of the ovram is uncontrollable by the subject (although the ovram is a motor driven device we used manual power for safety reasons). For the second step, as the ovram may generate jerky movements, joystick loads and sensitivity tolerances are altered accordingly.
The software detects all the variables during the test and logs them to be used in a composite score, derived from the variables to indicate the probability of success of the candidate in pilot training. Essentially, the measurement model is a set of theoretical hypotheses about the conceptual meaning of the dynamic period and its relationship to its measures (MacKenzie, 2003). Details of the software and calibration algorithm cannot be discussed in this paper for commercial reasons. However, any researcher is free to use the system.
It is evident that pilots who acquired higher scores on recruiting tests were found to have experienced fewer hazardous events (Goeters, 1993) (Hunter, 2003). This is why the construct validity of any test to recruit a pilot candidate ought to be adequately supportive.
First and foremost, the construct validation of test scores differed at certain points for 41 subjects. To our surprise, vital output was not one of them. All the subjects screened almost the same vital output values; therefore, we concluded that the effect of the dynamic condition has a very dense effect on vital output values, however, this does not (or at least must not) correlate with the psychomotor performance. The vital outputs of aerobatic and fighter pilots recorded slightly higher scores with better test results. We found no

correlation between the vital outputs and test results of the other 33 subjects.
Test results of the 10 non-pilot individuals were balanced with the ones obtained from 3 aerobatic pilots, 9 student pilots, 5 military fighter pilots, 5 airline pilots (none of which held former military posts), and 9 general aviation pilots; accordingly, a base corridor was established for the test.
Subsequently, the 64 volunteer pilot candidates (43 males and 19 females, aged between 17 and 32) were tested.

Results and Conclusion

The 64 pilot candidates who volunteered for the test had already obtained a test result from a well-known test protocol for their psychomotor and intellectual suitability for piloting. Fifty-three pilot candidates (35 males and 18 females) completed the test with similar vital outputs and reasonable corridor scores. Two of the 64 subjects exposed scores that were over the corridor. On the other hand, 9 of the subjects failed to enter the previously set corridor; three of these subjects (2 males, 1 female) had very low blood oxygen levels with high heart rates.
We followed 7 of the 9 failed candidates to their initial flight training. Two of these candidates had to stop training as they vomited each time they flew; three progressed very slowly, with one of them still not being released for their first solo flight at 17 hours of flight training.

Post-test follow-up of the candidates may demonstrate that our test is reliable and valid and can be used as part of pilot candidate selection protocols.
Thus, it seems necessary to construct test batteries within specific selection conditions to ensure they are robust.

References

Bartram, D. a. (1984). Automated testing: Past, present and future7. *Journal of Occupational Psychology , 3*.

Diamantopoulos, A. P. (2008). Advancing formative measurement models. *Journal of Business Research , 61* (12).

Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika , 4*.

Goeters, K.-M. B. (1993). The construction of personality questionnaires for selection of aviation personnel. *The International Journal of Aviation Psychology , 2*.

Hoermann HJ, L. X. (2002). Empirical evaluation of a selection system for Chinese student pilots. *Space Med Med Eng , 1*.

Hunter, D. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *International Journal of Aviation Psychology* .

MacKenzie, S. B. (2003). The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science , 3*.

McGee, M. G. (1979). Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological bulletin , 5* (889).

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist , 9*.

Parsons, T. D. (2008). Virtual environment for assessment of neurocognitive functioning: virtual reality cognitive performance assessment test. *Studies in Health Technology and Informatics* .

Pellegrino, J. W. (1984). Understanding spatial ability. *Educational Psychologist , 4*.

Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher , 8*.