

Paper presented at the 20th International Symposium on Aviation Psychology. Dayton/OH, May 7-10, 2019.

THE USE OF A PERCEPTUAL SPEED TEST IN CIVILIAN PILOT SELECTION

Hans-Juergen Hoermann, Ph.D.
German Aerospace Center (DLR)
Institute of Aerospace Medicine, Department of Aviation and Space Psychology
Hamburg, Germany
Diane L. Damos, Ph.D.
Damos Aviation Services, Inc.
Gurnee, Illinois

Perceptual speed is an important attribute for success as a pilot and has been assessed in many pilot selection batteries. The Tabular Speed Test (TST), a paper-and-pencil test of perceptual speed, was administered to 227 ab initio pilots and 152 licensed pilots who applied for employment at a large European airline. The airline had a multi-stage selection process including a computerized battery assessing quantitative and spatial abilities, short-term memory, psychomotor performance, multi-tasking, and personality. The number of correct responses (NC) had significant positive correlations for both ab initio and licensed pilots with measures of quantitative and spatial abilities, visual memory, multi-tasking, and psychomotor performance, but not with personality. The number of incorrect responses (NW) was correlated with Emotional Instability and Openness for licensed pilots but not with measures of cognitive ability.

Perceptual speed was identified by Thurstone (1938) as one of the seven primary human abilities comprising intelligence. During World War II, the U.S. Army Air Forces Aviation Psychology Program found that flight instructors often cited slow perception as a cause of failure in flight training (Guilford & Lacey, 1947). Consequently, several perceptual speed tests were developed and administered to pilot candidates. The results from two of these tests, Table Reading and Dial Reading, were combined. The composite was found to have a validity of 0.28 to pass/fail from primary flight training. This composite also had the distinction of being the best single predictor for any aircrew specialty (pilot, navigator, bombardier, etc.). The Table Reading Test had loadings on a perceptual speed factor, a numerical factor, and a spatial-relations factor (p. 813-815). Perceptual speed tests are still included in the U.S. Air Force pilot selection battery because they contribute unique variance to the prediction equation.

Mount, Oh, and Burns (2008) reviewed the literature on perceptual speed tests and determined that NC and NW assess different attributes. Mount et al. hypothesized that NC assesses task performance, whereas NW assesses rule compliance. The NW showed incremental validity for predicting rule compliance for warehouse workers beyond that contributed by general mental ability, conscientiousness, extraversion, and emotional stability. Thus, NW reflects a different attribute.

The TST was developed to assess perceptual speed in civilian pilots, not in the population as a whole. Like traditional perceptual speed tests, it is speeded. It has been used at a U.S. university in the professional pilot curriculum, where it was found to predict pass/fail for the private pilot's license flight test (Mekhail, Niemczyk, Ulrich, & Karp, 2010).

The study reported below was conducted at a major European airline and had three major goals. The first goal pertained to the use of the TST as a selection instrument. A good pilot selection tool assessing an attribute should not be affected by flight time, age (within normal pilot hiring limits), or gender. The second goal was to determine the relation between scores on the TST and other tests of cognitive ability, psychomotor performance, and personality traits. If, as Mount et al. (2008) suggest, NC on perceptual speed tests assesses a cognitive ability, then NC on the TST should be unrelated to personality traits or psychomotor skills but positively related to scores on other cognitive tests. NW, in contrast, should only be related to personality traits that relate to rule compliance.

The third goal pertained to the predictive validity of the airline company's current assessment battery. Traditionally, selection tests have been validated by correlating scores on the instrument with scores obtained during flight training. However, because of the pilot shortage, airlines are less willing to fail pilots or students during training. Some airlines are providing remedial training for weak candidates before they report for their initial airline training. Other airlines are providing additional ground school and simulator sessions to increase the likelihood that a weak candidate will successfully complete initial training. The use of remedial training with or without extra sessions during initial training makes pass/fail from training questionable as a criterion. Additionally, large airlines with subsidiaries may send low-time pilots to their subsidiary airlines, which may make data collection difficult. Ab initio candidates may require several years to complete flight training, delaying the collection of criterion data and raising issues pertaining to internal validity (Campbell & Stanley, 1963).

This paper takes a different approach to validity. Pilot hiring is an expensive process, particularly when it involves simulator evaluations, interviews by line-qualified captains, and evaluations by multiple assessors. Most airlines place the most expensive selection instruments toward the end of the selection process to minimize late-stage failures. Such failures represent substantial financial losses for the airline. Thus, any instrument that can be administered in the early stages of selection and can predict late-stage results is valuable for an airline. For the purposes of this paper, we examine how well scores on the TST predicted success at various stages of an airline selection process.

Methods

Participants

Data were collected on 227 ab initio candidates and 152 pilots holding at least a European commercial license. Most were German nationals. The ab initio sample consisted of 35 women and 192 men with ages ranging from 17 to 41. Most had no flight time although a few had up to 170 hours. None of the ab initio candidates had a commercial pilot's license. The experienced pilot sample consisted of 18 women and 134 men ranging in age from 20 to 41.

Their flight time ranged from 130 to 6250 hours. All of the licensed pilot candidates were native German speakers. The minimum education level for all candidates was passing the university entrance examination or the equivalent.

Procedure

The TST was included as part of a selection battery administered to both ab initio and experienced pilots who applied for a position at a major European airline. The airline used a four-stage, multi-hurdle procedure for both the ab initio candidates and the experienced pilots. Elimination was possible after each stage. Testing was conducted on three days, which could be separated by two to four months.

The four stages were 1) computerized aptitude testing, which included psychomotor tests and personality tests, 2) assessment center evaluation, 3) simulator evaluation, and 4) panel interview. Ab initio candidates performed the stages in this order. Stages 2 and 3 were reversed for the experienced pilot candidates. The aptitude testing was done on Day 1. The TST was administered after all of the aptitude tests were completed. Stage 2 testing was done on Day 2. Stages 3 and 4 were done on Day 3. Instructions for the TST were given in English for all candidates. Instructions for all of the other tests were given in German. Day 1 required about 9 hours for ab initio candidates and about 8 hours for experienced pilots.

Tests

Aptitude tests. The aptitude battery assessed five cognitive abilities plus a test of psychomotor coordination and a test of multi-tasking. Five cognitive abilities were assessed: 1) quantitative, 2) spatial, 3) attention, 4) perceptual speed, and 5) memory span (aural and visual). Additionally, the candidates were given a test of written English. Experienced pilots also received tests of aeronautical knowledge and left-right discrimination, whereas the ab initio candidates were given a test of mechanical aptitude.

Personality tests. The Temperament Structure Scale (TSS) (Mittelstaedt, Pecena, Oubaid, & Maschke, 2016) was administered to all candidates. This test requires approximately 40 min. and has 11 scales. Two of the scales were not administered to licensed pilots because they were concerned with willingness to travel and other lifestyle choices that were less relevant. The cockpit management attitudes questionnaire (FMAQ) was administered only to the licensed pilots and required about 35 min. Both personality tests were administered after the aptitude tests. (Merritt, Helmreich, Wilhelm, & Sherman, 1996)

TST. The TST is a paper-and-pencil test that requires 15 min. to administer with 9 min. of testing time.

Criterion Measures

Simulator evaluation. The ab initio candidates performed simple instrument flying tasks in a low-fidelity training device. The instrumentation was similar to that of a single-engine aircraft. In contrast, the experienced pilots used a high-fidelity simulator for their evaluation.

Assessment center. During the assessment center testing candidates were asked to perform three different types of exercises. One involved communication with a partner; the second, conflict resolution; and the third, group problem solving. Candidates were scored by a team of trained observers on traits such as “rule compliance” and “leadership.”

Interview. The interview was conducted by a board consisting of one airline captain and two psychologists for the ab initio candidates and by two airline captains and two psychologists for the experienced pilots. The interviews for both types of candidates took approximately 60 min. The interviews used a semi-structured format, i.e. the topics were identical for each type of candidate, but the number and type of follow-up questions could differ from candidate to candidate. Results from the FMAQ guided some of the questions for the licensed pilots.

Results

Because some of the test results (especially NW) had very skewed distributions, the Spearman rho (r_s) correlation, a non-parametric measure, was used to assess association.

TST as a Selection Instrument

Ab initio candidates. Significant male-female differences were found ($t(225) = 3.43, p = 0.018$) for NC but not for NW. Female candidates scored slightly higher. Neither NC nor NW was significantly related to either age or total flight hours.

Licensed pilots. There were no significant male-female differences for licensed pilots for either NC or NW. Age was not significantly related to NC or to NW. The NC was significantly related to the number of flight hours ($r_s(151) = -.178, p = .029$); NW was not.

TST and Cognitive, Psychomotor, and Knowledge, Measures

Table 1 shows the significant correlations between TST scores and performance on the Day 1 computerized aptitude and knowledge tests.

Table 1.
Spearman Rho Correlations Between TST Scores and the Aptitude and Knowledge Test Scores

Selection Test	<u>Ab Initio Candidates</u>		<u>Licensed Pilots</u>	
	TST NC	TST NW	TST NC	TST NW
Quantitative				
Mental Arith.	.31**		.16*	
Mental Arith (NW)	-.27**			
Math Reason.	.37**		NA	NA
Math Reason (NW)	-.28**		NA	NA
Perceptual Speed				
Optical PS	.28**		.19*	
Optical PS (NW)	-.25**			

Spatial				
Mental Rot.	.19**		.22**	
Mental Rot. (NW)	-.19**		-.22**	
Cube Rot.	.20**			
Cube Rot. (NW)		.14*		
Left-Right Discrim	NA	NA	.33**	
Left-Right (NW)	NA	NA	-.22*	
Memory				
Visual	.27**		.33**	
Visual (NW)	-.30**		-.31**	
Running Span			.24**	
Attention				
Concentration	.18**			
Concent. (NW)	-.16*		-.28**	
Psychomotor				
Multi-task	.21**		.25**	
Mechanical Comp				
Test 1	.17*		NA	NA
Test 1 (NW)	-.15*		NA	NA
Test 2	.22**		NA	NA
Test 2 (NW)	-.19**		NA	NA
Knowledge				
English	.15*		.17*	
English (NW)	-.16*		-.18*	
Aeronautics	NA	NA		
Aeronautics (NW)	NA	NA		

Note. Only significant results are shown. N = 227 for the ab initio candidates. N = 152 for the licensed pilot candidates. * $p < .05$; ** $p < .01$. NA is not applicable. NW is number wrong. NC is number correct.

TST and Personality

For both the ab initio candidates and the licensed pilots, none of the correlations between scores on the TSS scales and NC were significant. NW correlated significantly with Openness ($r_s(151) = .179, p = .027$) and with Emotional Instability ($r_s(151) = .208, p = .010$) only for licensed pilots.

Criterion Measures

Most of the candidates who completed Day 1 testing had not completed Day 2 at the time of the analyses. NC had a significant correlation with pass/fail for Day 1 ($r = 0.187, p < .01, N = 227$) for the ab initio candidates. It was not significantly correlated with pass/fail for the Assessment Center ($p > .05, N = 45$) or with pass/fail for the simulator evaluation ($p > .05, N = 28$). No analyses were conducted on the final interview because only 18 candidates completed this stage. The log (NW) was not significantly correlated with pass/fail from Day 1, the Assessment Center, or with the simulator evaluation. For licensed pilots, NC correlated

significantly with pass/fail from Day 1 ($r = .28, p < .01, N = 152$). No significant correlations were found for Day 2 or Day 3 ($N = 85$ and $N = 52$, respectively).

Discussion

The analyses of TST as a selection instrument showed mixed results. Age had no effect. NC was related to the number of flight hours, but only for the licensed pilots. Gender differences were found but only for ab initio candidates and only on NC. Because of the small number of women in both samples, the gender results should be viewed with caution.

NC was significantly correlated with many measures of cognitive aptitude, psychomotor performance, and knowledge. These results may indicate that the TST is g-saturated and assesses several attributes. The lack of correlation between NC and measures of personality was expected assuming NC is a measure of performance. NW was related only to the NW on the cube rotation test. The correlation between NW and Emotional Stability was larger than Mount et al. (2008) found but not unexpected. Thus, more errors on the TST are correlated with a higher level of emotional instability. We have no explanation for the correlation with Openness.

NC correlated significantly with pass/fail for Day 1 for both groups of pilots. Validation results for the ab initio candidates for Day 2 and Day 3 were limited by the small number of candidates who had completed these phases of testing and should be regarded with caution.

References

- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Printed classification tests. Report No. 5*. Washington, D.C.: U.S. Government Printing Office.
- Mekhail, A., Niemczyk, M., Ulrich, J., & Karp, M. (2010). Using the Table Reading Test as an Indicator for Success in Pilot Training. *Collegiate Aviation Review*, 28(1), 101-114.
- Merritt, A. C., Helmreich, R. L., Wilhelm, J. A., & Sherman, P. J. (1996). *Flight Management Attitudes Questionnaire 2.0 (International) and 2.1 (USA/Anglo)*. (Aerospace Crew Research Project Technical Report 96-4). Austin, TX: University of Texas:
- Mittelstaedt, J. M., Pecena, Y., Oubaid, V., & Maschke, P. (2016). Construct validity of the Temperament Structure Scales within the Big Five Framework in Aerospace Selection. *Aviation Psychology and Applied Human Factors*, 6, 68-80.
- Mount, M. K., Oh, I.-S., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, 61, 113-139.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, 1, ix + 121.