KNOWLEDGE TESTS FOR AB-INITIO PILOT SELECTION IN CIVIL AVIATION

Oliver Zierke
German Aerospace Center (DLR)
Hamburg, Germany

A validity study was conducted for two purposes: to evaluate the importance of knowledge tests for pilot selection and to validate the knowledge tests of the German Aerospace Center. The two criteria were a theoretical test given in the Flight Training School and the overall graduation rate. Regression models were calculated to measure the incremental validity of knowledge tests beyond ability tests. A further question addressed the comparison of the predictive validity of knowledge tests and school grades. Knowledge tests contributed 12% incremental validity and thus yielded similar results to school grades (17%). Both regression and dropout analyses revealed knowledge tests to be a better predictor of Flight Training School success than ability tests.

Test batteries often contain knowledge tests, as is the case for the Air Force Officer Qualifying Test (AFOQT, Carretta & Ree, 1996), the Armed Services Vocational Aptitude Battery (ASVAB, Sands, Waters, & McBride, 1997), or the test battery of the German Aerospace Center (DLR, Maschke, 2004). What we refer to as knowledge tests are methods used to reveal the amount of knowledge, skills, and experience of an individual in a given field. Knowledge tests measure aspects of crystallized intelligence in terms of Cattell's theory of intelligence (Cattell, 1987). Meta-analytic research on predictors for job success suggests that knowledge tests show good prognostic validities in general, and particularly for jobs in aviation. Schmidt and Hunter (1998) found a mean correlation of $r = .48$ with job success and an incremental validity beyond cognitive measures of .07 for many jobs. Dye, Reck, and McDaniel (1993) also found a mean correlation of $r = .47$ between knowledge tests and job performance. For the field of aviation, Olea and Ree (1994) reported a mean correlation of $r = .24$ between knowledge tests and practical training success for pilots. In meta-analytical studies, Martinussen (1996) found prognostic validities of $r = .24$ for tests of aviation information and in Martinussen (1998) the same magnitude was reported for tests of mechanical comprehension and aviation information for pilot selection in the Norwegian Air Force. As a result of her meta studies, Martinussen explicitly recommended that pilot selection batteries should include knowledge tests about aviation (Martinussen, 1996, p. 16). The reasons for using knowledge tests for the selection of pilot applicants are different for ready-entry and for ab-initio applicants. Ready-entry applicants are required to have prior job knowledge (Ree, Carretta, & Teachout, 1995), which can be defined as "the cumulation of facts, principles, concepts, and other pieces of information that are considered important in the performance of one's job" (Dye et al., 1993, p. 153). It is assumed that job knowledge has a positive impact on job performance. In contrast, ab-initio applicants are generally tested for more basic knowledge in mathematics, physics, and technical comprehension. This corresponds to the distinction of Caretta and Ree (2000, p. 2): "When selecting applicants for ab-initio training, indicators of ability (i.e. trainability) are emphasized. When selecting from experienced pilots, commercial carriers tend to emphasize indicators of prior experience (…) and flying competence (…)." This study concentrates on the impact of knowledge tests for

ab-initio pilots. These tests measure specific knowledge in domains which are important prerequisites for the theoretical training, that is mathematics, physics, technical comprehension and English language.

The following questions will be addressed:
1. Do knowledge tests contribute any unique predictive validity to cognitive ability tests?
2. How valuable are knowledge tests compared to school grades as predictors of success in the flight training school?
3. Which tests best predict training dropouts?

Methods

Subjects

The criterion data were collected from 402 student pilots (88% male, 12% female) in their theoretical ab-initio pilot training at the flight school of a major European airline. All students had graduated from high school with a degree suitable for enrolling at a university. The subjects were about 21 years old on average ($SD = 2.5$, range = 18 to 29).

Measures

The predictor data emerged from the first phase of the DLR selection approximately 8 months before the students began their pilot training. The DLR test battery was administered by computer and included seven cognitive ability tests as well as four knowledge tests and two psychomotor ability tests. The cognitive ability tests covered concentration, memory capacity, quantitative ability, and spatial orientation. The knowledge tests covered the topics of English language, mathematics, technical comprehension, physics, and technical basics. Furthermore, school grades in subjects related to theoretical pilot training were recorded. These subjects are English, mathematics, and physics.

The first criterion is a composite score of two test results from a theoretical training first phase (Starter Course) at the flight school. This Starter Course covers basic mathematical and physical knowledge which is considered to be prerequisite for the following theoretical training for the Air Transport Pilot License (ATPL). The second criterion is the pilots' successful completion of the flight training program. It is a simple pass/fail-criterion. The failed applicants were defined as one group called dropouts.

Analyses

Bivariate correlations between predictors and criteria were calculated. Correlation coefficients were corrected for range restriction and are reported in both forms, corrected and uncorrected. In order to estimate the incremental validity of knowledge tests, multiple regressions were carried out. Furthermore, a dropout analysis with independent-samples t-tests was done in order to find out which tests were most predictive.

Results

1. Knowledge tests and school grades were more strongly related to success in the Flight School Starter Course than the cognitive ability tests (see Table 1, column Correlation).
2. Knowledge tests added 12% to the explained variance of the criterion beyond the cognitive tests. School grades added 17% to the explained variance of the criterion.
3. The most successful tests in predicting the dropouts were the psychomotor tests and school grades followed by the knowledge tests. The cognitive ability tests were least predictive (see Table 1, column Effect Size).

Discussion

The knowledge tests turned out to predict both an early and a late criterion well. This result is in line with most of the meta-analytical results mentioned in the introduction. In fact, their predictions of different criteria were even better than those of cognitive ability tests. However, this was surprising in light of research on ab-initio pilots by Olea and Ree (1994) and Ree, Carretta, and Teachout (1995) who demonstrated the predictive power of general intelligence (g) for the AFOQT. These studies did not differentiate between fluid intelligence (Gf) and crystallized intelligence (Gc). Instead, they defined every test loading on a first factor as g, regardless of whether it was a reasoning test or a knowledge test. Actually, about half of the tests they used were knowledge tests in terms of the Gf-Gc theory, as for example Mathematical Knowledge, Mechanical Comprehension, and Aviation Information. These tests correlated well with practical training results. It seems that Olea and Ree's g includes both Gc and Gf to the same degree because their test battery, the AFOQT, contains elementary cognitive tasks as well as knowledge tests. Therefore, their results do not contradict the present study.

The present study is consistent with results in the area of admissions to graduate programs. A meta-analysis by Kuncel, Hezlett, and Ones (2001) referring to the Graduate Record Examination included more than 1700 independent samples and demonstrated that domain-specific knowledge tests are better predictors for successful graduation than the ability tests. The same line of results is shown in research on learning and instruction as well as research on expertise. Weinert and Helmke (1995) demonstrated that previous knowledge predicted school grades better than intelligence. Grabner, Stern & Neubauer (2007) showed that the strongest predictor of the attained level of chess experts was deliberate practice and the accumulation of a broad knowledge base rather than general intelligence. Based on this evidence, knowledge tests should receive more attention in aviation psychology.

Table 1

Correlations of the Predictor Variables with the Criterion Starter Course Test and Effect Sizes for Prediction of Dropouts

Variable	N	Validity	
		Correlation *r*	Effect Size *d*
Cognitive Tests			
Concentration			
OWT	402	.08	0.03
SKT	402	-.02	0.34
Memory Capacity			
MST	402	.07	0.30
RMS	402	.15**	0.01
Quantitative Ability			
KRN	402	.20**	-0.07
Spatial Orientation			
PPT	402	.19**	0.43
ROT	402	.08	0.08
Psychomotor Ability			
MIC	402	.10*	0.49
SIM	402	.18**	0.69
Knowledge Tests			
ENS	402	.09	-0.10
RAG	402	.32**	0.33
TEC	402	.33**	0.31
TVT	402	.23**	0.69
School Grades			
English	374	.17**	0.21
Mathematics	375	.30**	0.59
Physics	366	.35**	0.60

Note. r = observed correlation, **p < .01, one-tailed.

References

Carretta, T.R. & Ree, M.J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology*, 8, 29-42.

Carretta, T.R., & Ree, M.J. (2000). *Pilot selection methods.* Technical report. Dayton: U.S. Air Force Laboratory.

Cattell, R.B. (1987). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Dye, D.A., Reck, M., & McDaniel, M. (1993). The validity of job knowledge measures. *International Journal of Selection and Measurement*, 1, 153-162.

Grabner, R.H., Stern, E., & Neubauer, A.C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, 124, 398-420.

Kuncel, R.K., Hezlett, S.A., & Ones, D.S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations. *Psychological Bulletin*, 127, 162-181.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *The International Journal of Aviation Psychology*, 6, 1-20.

Martinussen, M. (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.

Maschke, P. (2004). The acceptance of ab initio pilot selection methods. *Human Factors and Aerospace Safety,* 4, 225-232.

Olea, M.M., & Ree, M.J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology*, 79, 845-851.

Ree, M.J., Carretta, T.R., & Teachout, M.S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology*, 80, 721-730.

Sands, W.A, Waters, B.K., & McBride, J.R. (Eds.) (1997). *Computerized adaptive testing: from inquiry to operation.* Washington: APA.

Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

Weinert F.E. & Helmke, A. (1995). Interclassroom differences in instructional quality and interindividual differences in cognitive development. *Educational Psychologist*, 30, 15-20.