Fluid intelligence and spatial reasoning as predictors of pilot training performance in the South African Air Force (SAAF)

Authors:
François de Kock[1]
Anton Schlechter[1]

Affiliations:
[1]Department of Industrial Psychology, Stellenbosch University, South Africa

Correspondence to:
François de Kock

e-mail:
fsdk@sun.ac.za

Postal address
Department of Industrial Psychology, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa

Keywords:
pilot selection; flight training; intelligence; spatial ability; incremental validity

Dates:
Received: 25 July 2008
Accepted: 02 Oct. 2008
Published: 30 Apr. 2009

How to cite this article:
De Kock, F., & Schlechter, A. (2009) Fluid intellingence and spatial reasoning as predictors of pilot training performance in the South African Air Force (SAAF). *SA Journal of Industrial Psychology/ SA Tydskrif vir Bedryfsielkunde, 35*(1), Art. #753, 8 pages. DOI: 10.4102/sajip.v35i1.753

This article is available at:
http://www.sajip.co.za

ABSTRACT

Pilot selection is a form of high-stakes selection due to the massive costs of training, high trainee ability requirements and costly repercussions of poor selection decisions. This criterion-related validation study investigated the predictive ability of fluid intelligence and spatial reasoning in predicting three criteria of pilot training performance, using an accumulated sample of South African Air Force pilots (N = 108). Hierarchical multiple regression analyses with training grade achieved as criterion were performed for each of the phases of training, namely practical flight training, ground school training, and officers' formative training. Multiple correlations of 0.35 ($p < 0.01$), 0.20 ($p > 0.05$) and 0.23 ($p > 0.05$) were obtained for flight, ground school and formative training results, respectively. Spatial ability had incremental validity over fluid intelligence for predicting flight training performance.

INTRODUCTION

Military pilot selection has traditionally been heavily researched, in part because pilots play a key role in modern warfare and training them is costly in terms of both finances and time (Hunter & Burke, 1994). In the United Kingdom, the estimated unit cost of training a fast-jet pilot is more than £3.7 million. One fighter pilot in the South African Air Force (SAAF) takes at least 5 years to train, which makes training failures costly. Dropout rates are high in the United States Air Force (20%) and Australian and Canadian programmes (30%) (Bourn, 2000). Aircraft accidents are also expensive in human, financial and psychological terms. Since choosing pilots represents such a high-stakes selection scenario, the military continues to research effective pilot selection measures.

Two issues seem to dominate current pilot selection research. Firstly, the fact that Spearman's general cognitive ability (g) plays such a central role in predicting pilot success has raised the question of whether it really makes sense to also assess specific intelligences (Carretta & Ree, 1989; Ree & Carretta, 1996; 2002). Proponents of this argument argue that most specific intelligences are so saturated with g that it rarely adds any incremental validity to batteries already containing measures of g. This argument would make sense, assuming Carroll's (1993) model of cognitive ability of a hierarchy of factors with g at its apex and group factors at successively lower levels to be true. For purposes of this paper, fluid intelligence and g were assumed to be theoretically congruent (for a summary, see Alderton & Larson, 1990) and, therefore, are used interchangeably as in other cited literature. The second issue relates to the so-called criterion problem. Most validity studies cite the lack of meaningful, quality criteria to validate predictors against a weakness (Hunter & Burke, 1994; Hunter & Schmidt, 1990). This problem is especially prominent in pilot selection research (Damos, 1996). Gaining a better exposition of the performance domain takes us some way towards a better understanding of ability-performance linkages (Schmitt & Chan, 1998). The present study sought to address both of these issues by assessing both measures of g and specific intelligences, and examining how well this combination predicts multiple criteria of pilot training performance. It also reports on the incremental validity of measures of specific intelligence over and above fluid intelligence, which remains a contested topic in pilot selection research.

Validation

The validation of selection procedures is necessary for various reasons. The pragmatic perspective views human resources, where the individual and his/her output is key, as critical to success in any organisation. Gatewood and Feild (1998, p. 3) state that '…the performance of employees is a major determinant of how successful an organisation is in reaching its strategic goals and developing a competitive advantage of rival firms'. Selecting people that are likely to perform effectively is a key responsibility of the human resource function, which by implication includes developing and validating effective selection procedures (Campbell, McCloy, Oppler & Sager, 1993; Milkovich & Boudreau, 1997).

The selection process must be reliable and it needs to make valid claims. According to internationally accepted principles and guidelines (American Psychological Association, 2003; United States Department of Labor, 1978) a sound selection procedure is one that allows valid inferences to be made regarding future job behaviour from available measure scores. Likewise, the Guidelines for the Validation and Use of Assessment Procedures for the Workplace (Society for Industrial and Organizational Psychology, 1998, p. 1) concur by stating that the evaluation of any assessment procedure should be 'based on the fact that sufficient proof can be found that the procedures used are indeed relevant to the position or work concerned'.

The 'proof' referred to above can be termed 'validity', and it refers to the '…degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test' (American Educational Research Association, APA & National Council for Measurement inEducation, 1999, p. 184). Validation, therefore, involves the accumulation of evidence – content, criterion or construct-related – to provide a sound scientific basis for the proposed score interpretations (APA, 2003).

From a legal perspective, validation is required by law in South Africa, as stipulated in the Employment Equity Act (RSA, 1998, p. 10): Psychological testing and other similar assessments of any employee are prohibited unless the test or assessment being used has been scientifically shown to be valid, reliable; can be applied fairly to all employees; and is not biased against any employee or group.

During the selection process a choice is made about desired qualities and traits. This choice rests upon a predictive hypothesis that is formulated after considering the demands and context of the job (Guion, 1965). The focus of selection research is then to test the predictive hypotheses that certain qualities and traits predict certain desirable behaviour. In this sense, validation is seen as a process of traditional hypothesis testing (Binning & Barrett, 1989; Landy, 1986).

Traditionally, validation research has received considerable attention in the military (Cook, 1999; Rumsey, Walker & Harris, 1994; Schmitt & Borman, 1993). A survey of contemporary literature (e.g., Hilton & Dolgin, 1991; Hunter & Burke, 1994) reveals a thorough understanding of the link between tasks, demands and knowledge-skills-abilities (KSAs) for the job of the military pilot, which is simulated fairly well in the training task.

Determinants of pilot training success

There is general consensus that the determinants of pilot success resort in three main domains, namely intelligence and aptitude, psychomotor coordination and personality (Carretta & Ree, 1989). A broad summary of research on each predictor area is provided next, even though not all of these were assessed in this study.

Intelligence and aptitude

Hilton and Dolgin (1991, p. 94) remark that '…there is little doubt that above average intelligence is necessary to master military pilot training.' They also characterise intelligence as the best and most stable predictor of flight training success, in their summary of pilot selection research during the last century.

Intelligence is a broad concept, and is sometimes defined more specifically. For instance, Ree and Carretta (1996) make a useful distinction between two types of intelligence. They use Spearman's (1904) two-factor theory of cognitive ability and argue that intelligence can be seen as general cognitive ability (g) on the one hand, or in terms of specific abilities (s_n) on the other. The factor g is a general factor that is obtained through factor analysis and is thought to underlie most of the other intellectual abilities (Plug, Meyer, Louw & Gouws, 1989). The construct g is synonymous with fluid intelligence.

The predictive validity of these types of intelligence appears to differ. Hunter and Burke (1994), in their meta-analysis of predictors of pilot success, found that general intelligence was not generalisable across studies as a predictor; at most it had an influence moderated by other variables. However, general cognitive ability has consistently been shown to predict pilot training success, showing average statistically significant correlations of 0.33 (Ree & Carretta, 1996).

General intelligence in other guises has also been shown to predict pilot training success. Cattell's concept of fluid intelligence (Cattell, 1987; Raven & Court, 1998) is defined as intellectual abilities that are determined primarily by genetic factors, as opposed to cultural or environmental factors (Plug *et al.*, 1989). Some evidence has been found that information processing capability, an important indicator of fluid intelligence, predicts pilot training success (Damos, 1996). A more recent South African study found that pilots could be differentiated from non-pilots on the grounds of rate of information processing (Barkhuizen, Schepers & Coetzee, 2002). It can be argued that fluid intelligence and information processing capability are two factors of intelligence that drive transfer and automatisation of learning during the flight training task. With regard to specific intelligence (s_n), a multitude of abilities have been found to predict pilot training success, among others verbal, quantitative, spatial, and mathematical ability, as well as perceptual speed and instrument comprehension (Burke, Hobson & Linsky, 1997; Carretta & Ree, 1996).

The relative importance of g and s_n in predicting pilot training success remains a controversial issue. On the one hand, some authors (e.g., Burke, Hobson & Linsky, 1997; Carretta, Perry & Ree, 1996; Ree & Carretta, 2002) maintain that g remains a better predictor of pilot success than specific abilities. Other authors (e.g., Hunter & Burke, 1994; Martinussen, 1996) come to different conclusions and report – as a result of their meta-analyses – that measures of general intelligence had low mean validities compared to more specific measures of intelligence.

Carretta, Perry & Ree (1996) shed light on this apparent contradiction with their view that the inclusion of specific abilities (s_n) adds little to the ability to predict criteria (see also Ree & Carretta, 1996), since many of the additional measures that are used are saturated with g and do not represent unique abilities. Some authors (e.g., Martinussen, 1996) disagree and demonstrate that the inclusion of specific abilities indeed had incremental validity over and above measures of g. Clearly, the debate on the role of intelligence and aptitude in the prediction of pilot training success is still very active and can be interpreted as an attestation of its dominance in pilot selection batteries.

Psychomotor coordination

Psychomotor skills research has a long history in pilot selection (Griffin & Koonce, 1996). The term 'psychomotor' denotes a combination of physical and psychological activities (Plug *et al.*, 1989). Measures of psychomotor coordination or hand-eye coordination as it is sometimes referred to are commonly included in selection batteries for two apparent reasons, being (a) they have an obvious relation to the task and (b) the results of validation research support their inclusion in selection batteries (Hilton & Dolgin, 1991).

In their study, (Burke, Hobson & Linsky, 1997) found that psychomotor tests were predictive of pilot training success and that its validity generalised across samples. They used Validity Generalisation Analysis (VGA) with three samples from different national air forces, with a large sample (N = 1760). A continuation of these authors' findings is the fact that various studies report that measures of psychomotor abilities were able to increase predictive validity of a battery already measuring g (Ree & Carretta, 1996). For instance, in one study when psychomotor tasks were added to a USAF selection battery already including the Air Force Officer Qualifying Test (AFOQT) scores, the predictive validity of the battery increased from 0.168 to 0.207 (Damos, 1996).

New developments in psychomotor predictors also abound. Various studies have illustrated the role of situational awareness in pilot functioning (Carretta *et al.*, 1996). Therefore, it can be expected that this construct might prove useful in future pilot selection batteries.

Personality

Personality can be defined as those aspects of individuals that make predictions about their behaviour in specific situations possible (Plug *et al.*, 1989). Contrary to expectation, most studies report that personality adds little to the prediction of pilot success (Carretta, Perry & Ree, 1996; Hunter & Burke, 1994; Retzlaff & Gibertini, 1987; Turnbull, 1992). However, some studies did in fact report that certain aspects of personality had incremental predictive validity in traditional batteries, for instance attitude to risk (Ree & Carretta, 1996). In another study,

Article #753
SA Journal of Industrial Psychology

Carretta (2000) found that a measure of conscientiousness incremented the multiple correlation coefficient of a battery measuring general mental ability from 0.51 to 0.60.

Despite the generally weak ability of personality to predict pilot training success, it is often used in pilot selection. For instance, certain militaries use personality as a screening variable to identify clinical dysfunction and other undesirable traits. It also appears that personality is receiving increased attention in the important areas of stress tolerance and motivation (Hilton & Dolgin, 1991).

A study that compared the personality profiles of pilots to those of college students through cluster analysis, found that pilots had distinct personalities that distinguished them from non-pilots (Retzlaff & Gibertini, 1987). A similar finding was obtained by a study comparing the personality profiles of student naval pilots with normative data (Lambirth, Dolgin, Rentmeister-Bryant & Moore, 2003). Ashman and Telfer (1983) found pilots to be more achievement oriented, outgoing, active, competitive, dominant and less introspective, emotional, sensitive and self-effacing than a sample of non-pilots.

In another study, pilot trainees completed a personality inventory measuring five dimensions thought to be associated with flight training performance. After their training was completed, three of the measures were in fact related significantly to training outcome, namely hostility, self-confidence and values flexibility. Disappointingly, incremental validity analysis did not indicate that the inventory could enhance a selection model already containing traditional aptitude scores (Siem, 1992).

Meta-analyses of predictors of pilot training performance

The meta-analysis of Hunter and Burke (1994) of 68 published studies, with a total of 437 258 combined cases using the method proposed by Hunter and Schmidt (1990), concludes that not one predictor conclusively generalised in terms of predictive validity across samples. However, a number of variables had generalisable validity moderated by various factors, including decade of the particular study, aircraft type, arm of service and nationality. The variables that had generalisable validity (with mean sample-weighted correlations indicated) included job sample (0.34), gross dexterity (0.32), mechanical ability (0.29), reaction time (0.28), biodata inventory (0.27), aviation and general information (0.22), perceptual speed (0.20), spatial ability (0.19) and quantitative ability (0.11). Validities that could not be generalised across samples were verbal ability (0.12), fine dexterity (0.10), age (-0.10), education (0.06) and personality (0.10).

Similar results are reported by Martinussen (1996) in a meta-analysis of 66 independent samples from 50 studies (combined N = 17900) from 11 nations, also using the Hunter and Schmidt (1990) meta-analysis method. She found the best predictors of pilot performance to be – with mean corrected validities indicated – a combination of cognitive and psychomotor tests (0.37), previous training experience (0.30), cognitive abilities (0.24), psychomotor/information-processing abilities (0.24), aviation information (0.24) and biographical inventories (0.23). Similar to the findings of Hunter and Burke (1994), certain factors were found to have low mean validities, including personality (0.14), intelligence or g (0.16) and academic tests (0.15).

In a smaller follow-up meta-analysis of four studies (combined N = 973), again using the Hunter and Schmidt (1990) method, Martinussen and Torjussen (1998) found that the best predictors of success in pilot training were instrument comprehension (0.29), mechanical principles (0.23) and aviation information (0.22).

Clearly, predictors vary across time frames, technology and development in the nature of the task of the military pilot. This underscores the importance of validation within the particular context of use of a selection battery. As Huysamen (1994, p. 31) caveats, 'it is therefore more appropriate to refer to the validity of a test for a particular application than to speak of the validity of a test.'

There is general consensus that the ability to predict pilot training success is inadequate. Obtained multiple correlations are still low (Damos, 1996), largely because of the choice of criterion and unique problems associated with pilot selection research, such as small selection ratios and severe restriction of range (Burke, Hobson & Linsky, 1997; Carretta, 1992a; Hilton & Dolgin, 1991). More recently, it has been proposed that more valid and reliable criterion measures be developed and that research into new models of personality be conducted (Damos, 1996).

The SAAF, like other military and civilian organisations utilising pilots in their service, continuously attempts to improve its ability to predict successful pilot training performance (Aspeling, 1980; Croucamp & Bolton, 2002; Smit & Bielfeld, 2001). If so much rests upon the quality of decisions made in pilot selection, it is critical that the relationship between the constructs measured by the assessment battery and different criteria of flight training performance be investigated. Moreover, the additional gain in predictiveness from measures of special intelligence should be investigated.

Research objectives

In light of the above, the objectives and hypotheses of this study were formulated. In the first instance, to establish the extent to which the predictors in this study relate to pilot training performance, it was hypothesised that statistically significant relationships would exist between predictors and criteria of pilot training success. Secondly, in order to establish the extent to which a battery consisting of fluid intelligence and spatial ability is a valid predictor of multiple criteria of flight training performance, it was hypothesised that the pilot training performance of SAAF pilots could be predicted from a battery consisting of measures of fluid intelligence and spatial ability. Thirdly, to establish if adding a measure of specific intelligence to a battery already containing a measure of g, improves the ability to predict pilot flight training performance, it was hypothesised that spatial ability would have incremental validity over and above fluid intelligence for predicting pilot training performance.

RESEARCH DESIGN

Research method

A predictive criterion-related validity design was used to investigate the relationship between predictors and identified criteria (Schmitt & Chan, 1998). The adoption of quantitative methods and the use of statistical analyses allowed the researchers to compare results with earlier research findings.

Participants

The sample consisted of five full annual intakes of SAAF pilots (N = 108) who successfully completed officer's formative training, ground school training and practical flight training. In order to control for possible differences in training and evaluation content across training cycles, the researchers extracted data for the total population of pilots who qualified from 1997 to 2001, since training and evaluation was held constant in this period. Since foreign instructors were used to train and evaluate pilot trainees in training cycles after this period, the researchers decided to limit the analyses only to these comparable groups (i.e., 1997–2001). The annual intake sizes were relatively well distributed (e.g., 24.1%, 24.1%, 23.1%, 13% and 15.7%, from 1997 to 2001, respectively). The ranks of participants upon entering training ranged from candidate-

SA Journal of Industrial Psychology Article #753

officer to major, where most (85.1%) resorted in the former category. In terms of gender, 101 of the pilots were men and six were women. The pilots were all under the age of 25 upon entering the training programme. All pilots had completed at least Grade 12. The distribution of the gender and ethnic groups in the sample is shown in Table 1.

Measuring instruments

Criterion measures

The criterion for this study was subjects' performance during the total pilot training process. Therefore, instructors' ratings of practical flight performance, training grades for ground school flight training and scores on officers' formative training were considered as measures of the dependent variable. Evidence of construct validity of the three measures of training performance was found in the present study and is reported later. The reliability of the criterion measures could not be investigated, which is a common weakness of pilot validation studies (Hunter & Burke, 1994; Martinussen, 1996). One Norwegian study estimated the reliability of its criteria (theoretical tests and pass/fail measures of training success) to be 0.90 (Martinussen & Torjussen, 1998).

Predictor measures

The predictor measures used in this study were selected from a larger set of assessment instruments that are used by the SAAF for selection of military pilots. Due to the fact that the composition of selection batteries varied over the five-year period of this study, only the predictors included in all cycles were studied.

Raven's Advanced Progressive Matrices

The Raven's Advanced Progressive Matrices (APM) is a nonverbal measure of general cognitive ability or *g* (Alderton & Larson, 1990; Jensen, 1998), as well as fluid intelligence (Cattell, 1987; Raven & Court, 1998). The APM was designed to differentiate between people of superior intellectual ability, such as students for advanced scientific or technical studies (Raven & Court, 1998). Various authors (e.g., Alderton & Larson, 1990; Arthur & Woehr, 1993) have confirmed that the APM measures *g* as a unidimensional construct. Reliability, construct and predictive validity of the instrument have been established in numerous studies (e.g., Bors & Stokes, 1998; Martinussen & Torjussen, 1998; Rushton, Skuy & Fridjhon, 2003).

Blox Test of Spatial Ability

The Blox Test (Lombard, 1980) is a test of spatial relations, orientation and visualisation, or spatial ability as it is commonly referred to in literature. This test assesses the ability to recognise three-dimensional objects which have been rotated in space and which are represented two-dimensionally, as in technical drawings. The Blox Test has been shown to yield acceptable reliability estimates of scores for various South African cultural groups, namely for Black Xhosa men (KR-20 = 0.89), Coloured Afrikaans Men (KR20 = 0.82), Indian Males (KR21 = 0.79) and Black Zulu Males (KR21 = 0.77). Studies in the engineering and trade environment illustrate adequate construct and predictive validity (Lombard, 1980). For instance, Van der Merwe (2002) showed that the Blox Test predicts success in skilled, technical jobs.

Procedure

The psychometric test scores of all participants were collected during their selection for the pilot training programme and combined with the training evaluation scores achieved after completion of training and subsequently screened for inadequate data. Cases with missing data on the primary criterion of flight training evaluation scores were excluded from the study. The validation design took the form of a predictive criterion-related validation study (Schmitt & Chan, 1998). In order to control for possible multi-year effects in criterion performance, the researchers compared the annual cycles in terms of the most important criterion measure for this study, namely flight training performance. An analysis of variance showed that the effect of year group was non-significant, $F(4, 95) = 2.03$, $p = .096$. Considering the minimum (65.9) and maximum (90.1) scores obtained on this criterion measure, the annual means ranged between 77.63 and 80.5. The Levene test for equality of variances between these groups was not significant for *year group* ($p = .208$). In addition to the fact that the content and method of flight training and evaluation remained constant, these results support the use of multi-year data in the subsequent analyses, since there is no indication of annual differences in average pilot flight training performance.

TABLE 1
Participants' biographical detail

POPULATION GROUPS	MALE	FEMALE		
	n	*n*	*N*	%
African	7	1	8	7.5%
Coloured	6	0	6	5.6%
Indian/Asian	5	0	5	4.7%
White	83	5	88	82.2%
Total	101	6	107*	
Percentage	94.4%	5.6%		100%

Note. One case had neither gender indicated ($N = 108$).

Statistical analysis

The statistical techniques included descriptive statistics, Pearson Product-Moment Correlation Analysis and Hierarchical (Sequential) Multiple Regression Analysis (Tabachnick & Fidell, 2001). Correlation analysis was used to determine individual ability-performance relationships. Hierarchical Regression was used to determine if the addition of information regarding specific abilities such as spatial ability, improved prediction of criteria of pilot training success beyond that afforded by variance in fluid intelligence. Analyses were performed using SPSS REGRESSION and SPSS FREQUENCIES, by the Statistical Package for the Social Sciences (SPSS), for evaluation of assumptions (SPSS, 2006). An alpha level of 0.05 was used for the determination of significance levels for all tests, unless stated otherwise. Using the tables of Cohen (1988), statistical power for this study was estimated at 0.87 ($N = 108$; estimated effect size $d = 0.30$).

RESULTS

Preliminary Analyses

Analyses were performed using SPSS REGRESSION, SPSS DESCRIPTIVES and SPSS FREQUENCIES were used for the evaluation of assumptions underlying the statistical techniques employed. These results led to transformations of the variables to reduce skewness and improve normality, linearity, and homoscedasticity of residuals. Inverse square root transformations were used on spatial ability, fluid intelligence and ground school performance scores. In most cases skewness was reduced with transformation, but normality was not significantly improved as judged by the respective Kolmogorov-Smirnov test statistics, which tests the hypothesis that a sample comes from a normal distribution. Therefore, transformations were not retained due to the consequent complication of interpretability of results and the fact that multiple regression analysis is believed to be fairly robust against moderate violations of the assumption of normality resulting from skewness (Tabachnick & Fidell, 2001). With the use of a $p < 0.05$ criterion for Mahalanobis distance, no outliers among the cases

Article #753
SA Journal of Industrial Psychology

were identified. A few cases had missing data, which were deleted pairwise, $N = 108$.

Correlations between predictor and criterion measures (Hypothesis 1)

Table 2 depicts correlations (Pearson) between the scores on predictor measures and the three measures of pilot training success. Based on the survey of literature and reasoning followed, it was expected that the first hypothesis would be supported, i.e. intercorrelations between predictors and criteria would be statistically significant ($p < 0.05$).

The results show that the intercorrelation between the two predictors, namely fluid intelligence and spatial ability, was moderate, positive and highly statistically significant ($r = 0.415$; $p < 0.001$).

Fluid intelligence was positively associated with two (of three) criteria for pilot training performance, namely with flight and formative training performance ($r = 0.248$; $r = 0.216$ respectively; $p < 0.05$). On the other hand, spatial ability was positively associated with flight training performance ($r = 0.336$; $p < 0.001$), but not with ground school training performance ($r = 0.138$; $p > 0.05$) and officers' formative training ($r = 0.033$; $p > 0.05$). Judging by these correlations, the first hypothesis regarding predictor-criterion relationships was supported for fluid intelligence as a predictor of two of the three criteria of flight training performance, but for spatial ability, only in the case of actual flight training performance as a criterion.

Hierarchical Multiple Regression results (Hypotheses 2 and 3)

To determine the validity of the battery to predict pilot training success, the regression of the various measures of pilot training success on the scores on the psychometric instruments was computed. Hierarchical Regression Analysis was performed for each criterion, since they represent distinctly different aspects of the training process that were of interest to the researchers. Since theory suggests that general cognitive ability underlies measures of specific intelligence, fluid intelligence was entered into the equation first, followed by spatial ability (Carretta, Perry & Ree, 1996). Table 3 displays the correlations between the variables, the unstandardised regression coefficients (B) and intercept, the standardised regression coefficients (β), the semipartial correlations (sr_i^2), and R, R^2, and adjusted R^2 after entry of both IVs.

For flight training performance, R was significantly different from zero at the end of each step. After step 2, with both predictors in the equation, $R = 0.354$, F (2, 93) = 6.66, $p < 0.01$. After step 1, with fluid intelligence in the equation, $R^2 = 0.061$, F_{inc} (1, 93) = 6.143, $p < 0.05$. After step 2, with spatial ability added to the prediction of flight training, $R^2 = 0.125$ (adjusted $R^2 = 0.106$), F_{inc} (1, 93) = 6.799, $p < 0.05$. The addition of spatial ability to the equation with fluid intelligence resulted in a significant increment in R^2. Collinearity diagnostics did not provide conclusive evidence of collinearity. Using the criteria for multicollinearity suggested by Belsley, Kuh and Welsch (1980), no roots had conditioning indices greater than 0.30 for a given dimension, although some dimensions had more than one variance proportion greater than 0.50. None of the tolerances (1 - SMC) approached zero. Coupled with low variance inflation factors (VIF), it indicated no serious cause for concern regarding multicollinearity (Tabachnick & Fidell, 2001).

For the other two criteria of pilot training performance, R was not significantly different from zero. For brevity's sake, their results will not be reported in separate tables but only in the text. For ground school training performance it was found that, with both IVs in the equation, $R = 0.196$, F (2, 93) = 1.855, $p > 0.05$. After step 1, with fluid intelligence in the equation, $R^2 = 0.038$, F_{inc} (1, 93) = 3.731, $p \leq 0.05$ (obtained p was marginal at 0.056). After step 2, with spatial ability added to the prediction of ground school training, $R^2 = 0.038$ (adjusted $R_2 = 0.018$), F_{inc} (1, 93) = 0.019, $p > 0.05$. The addition of spatial ability to the equation with fluid intelligence did not result in a significant increment in R^2. Similar results were found for officers' formative training performance, where R was significantly different from zero only after step 1, with fluid intelligence in the equation, $R^2 = 0.047$, F_{inc} (1, 88) = 4.327, $p < 0.05$. After step 2, with spatial ability added to the prediction, $R^2 = 0.054$ (adjusted $R^2 = 0.032$), F_{inc} (1, 87) = 0.627, $p > 0.05$. With both IVs in the equation, $R = 0.232$, F (2, 87) = 2.468, $p > 0.05$. Although officers' formative training could be predicted from fluid intelligence, the addition of spatial ability to the equation did not result

TABLE 2
Intercorrelations (Pearson) between predictor variables and criteria of pilot training performance

	VARIABLE	1	2	3	4	5
1.	Fluid intelligence	-	0.42**	0.22*	0.20	0.25*
2.	Spatial ability		-	0.03	0.14	0.34**
3.	Officer's formative training			-	0.13	0.05
4.	Ground school training				-	0.42**
5.	Flight training					-

Note. In tables, values were rounded to the second decimal, but in the text, the third decimal was used. *$p < .05$. **$p < 0.01$ (2-tailed).

TABLE 3
Summary of hierarchical regression analysis for variables predicting flight training performance

VARIABLE	MEAN	SD	B	β	Sr^2	t	p
Fluid intelligence	29.76	2.93	0.22	0.13	0.06*	1.24	0.22
Spatial ability	37.93	3.87	0.35	0.28	0.06*	2.61	0.01
Flight Training Performance (DV)	78.56	4.84	Intercept = 58.85				
			ANALYSIS OF VARIANCE				
			Source	**df**		**Sum of Squares**	**Mean Square**
Multiple R	0.35**		Regression	2		278.92	139.46
R^2	0.13		Residual	93		1947.36	20.94
Adjusted R^2	0.11		Total	95		2226.28	
SE of Estimate	4.58						
			F (2, 95) = 6.66; $p < 0.01$.				

Note. $N = 96$. In tables, values were rounded to the second decimal, but in the text, the third decimal was used. *$p < 0.05$. **$p < .01$ (2-tailed)

in a significant increment in R^2. The results indicate that the second hypothesis was supported in the case of flight training performance, but not for the other criterion measures. Similarly, the third hypothesis – that spatial ability has incremental validity over and above *g* – was supported only in the case of the prediction of flight training performance.

As is common in most pilot selection validation studies, due to range restriction (Thorndike, 1949), obtained correlations or validities will tend to underestimate the true validities of predictors in the battery simply because the full range of ability is not present in the validation sample (Hunter & Schmidt, 1990). Since selection data for the unselected group were not available, corrections for multivariate restriction of range and unreliability in the criteria could not be computed (Guilford, 1954).

Using the Cattin (1990) formula to estimate population cross-validity, which is the value one could expect if an infinite number of samples were available upon which to estimate cross-validity and one computed the average value of cross-validity across these samples for flight training performance, population cross-validity is estimated at 0.32 (Schmitt & Chan, 1998).

The relationship between the various criterion measures is depicted in Table 2. Criterion convergence is apparent in these correlations, since pilot flight training and ground school training were strongly correlated and highly statistically significant ($r = 0.424$; $p < 0.001$), which serves as evidence of construct validity of the criterion. On its part, officers' formative training was not related to the other two criteria, thereby indicating that it measures aspects of training performance that are not necessarily related to the flying task.

DISCUSSION

The primary aim of this study was to determine the regression of pilot training performance during flight, ground school and officers' formative training on the scores on the psychometric instruments. The researchers expected that individuals with higher levels of fluid intelligence and spatial ability would achieve better training scores in pilot training. It was hypothesised that a significant relationship exists between pilot training performance, fluid intelligence and spatial ability. Furthermore, it was expected that spatial ability would add incremental validity to fluid intelligence in predicting pilot training performance. Partial support was found for these hypotheses.

Interpretation

Two (of three) criteria of pilot training performance were significantly (positively) associated with fluid intelligence, in line with earlier research on the prominent role of general cognitive ability (*g*) in predicting pilot training performance (Damos, 1996; Hilton & Dolgin, 1991; Ree & Carretta, 1996). Assuming that *g* and fluid intelligence are theoretically congruent, an explanation for this finding can be taken from Thorndike (1949, 1986) and Schmidt and Hunter (1998), who stated that *g* is central in predicting training and job success across hundreds of occupations.

The association between spatial ability and flight training performance mirrored the results of Carretta, Perry & Ree (1996), which confirms that spatial relations and orientation play an important part in the actual task of flying an aircraft. Spatial ability was not related to ground school and officers' formative training; there is also no apparent theoretical link to be made between these constructs.

In general, the results of this study are consistent with previous research on the prediction of pilot training success in two ways: (a) fluid intelligence remains one of the best predictors of flight training performance, and (b) the obtained correlations between predictors and criteria are still only moderate at best (Burke, Hobson & Linsky, 1997; Carretta, Perry & Ree, 1996; Damos, 1996; Hilton & Dolgin, 1991; Hunter & Burke, 1994).

A unique finding of this study was that spatial ability could indeed add incremental predictive validity to a battery already containing a measure of fluid intelligence (or *g*), contrary to the results of previous studies (e.g., Carretta, Perry & Ree, 1996; Ree & Earles, 1991). Although not a unique finding (see Martinussen, 1996), it is generally believed that little value is gained from measuring anything else but *g* in pilot selection, which is obviously not a generalisable conclusion. Even though our data support the view of Carretta, Perry and Ree (1996) that specific abilities (e.g., spatial ability) are saturated with *g*, this study shows that in some cases measures of specific forms of intelligence (s_n) could significantly improve the ability to explain variance in pilot training success.

Conclusion and recommendations

Since we have shown that sometimes special intelligence (s_n) has incremental validity over g, the implication is that, practically, such an (small) increase in predictive validity translates into significant utility yields when considering the low selection ratios, high costs associated with training and low base rates typical in pilot selection (Murphy & Davidshofer, 2005). Moreover, selection decision-errors in pilot selection can be catastrophic and therefore these can be minimised by using more accurate selection procedures that capture more of the factors that 'cause' performance in the cockpit.

One explanation for the imperfect prediction of criteria relates to the so-called criterion problem. Criteria in validation studies often do not receive the same attention as predictors, especially with regard to adequate choice, reliability and construct validity (Burke, Hobson & Linsky, 1997). Poor reliability of selection instruments affects predictive validity (Huysamen, 1996). Some American pilot selection studies (e.g., Carretta, 1989; 1992b) show that using computerised psychometric testing tends to increase reliability and validity.

Limitations

There were limitations to this study. For one, the absence of psychometric data from non-successful applicants makes estimates of the population statistics impossible, which is a requirement for the computation of adjustments to the validity coefficients for restriction of range and unreliability in the variables. Most pilot selection studies, such as that of Burke, Hobson and Linsky (1997), report substantial improvements in validity coefficients when adjusted for restriction of range and unreliability of criteria. Secondly, the psychometric characteristics of the criterion measures could not be investigated, apart from some evidence of convergent validity in this study. It is suggested that future studies provide a detailed analysis of the reliability of instructors' ratings, training score results and cockpit ratings of flying performance. Thirdly, a reviewer of this paper suggested the possibility that a method effect, caused by similarity in presentation of stimuli in both predictor measures, may have confounded the results under review. Although this remains a possible explanation for the research findings, the fact that spatial ability could explain additional variance in flight training performance suggests that the measures did indeed measure distinct constructs. However, future studies should consider the possible confounding influence of method effects in predictor measure choice.

These shortcomings highlight the fact that selection programmes should include validation considerations

Article #753
SA Journal of Industrial Psychology

such initiatives. Validation is research and should not be seen as a statistical post-mortem, but rather as an integral part of any personnel selection project.

Suggestions for further research

The results of a local validation study should be interpreted with caution, as validity coefficients can fluctuate from one sample to the next, especially where sample sizes are small (APA, 2003). Therefore, the results of this study should be cross-validated in a future study. Since sufficiently-sized validation samples in the SAAF must accumulate with time to allow for appropriate statistical analyses, collaboration with similar institutions in the private and non-governmental sectors should allow sharing of data for validation purposes (Sackett & Arvey, 1993).

However, a stamp-collecting approach to validation that exaggerates emphasis on statistical validities obtained is also undesirable (Landy, 1986). Unfortunately, most selection procedures involve small *N*-settings, which mean that a reliance on empirical results is not always an option when deciding on the suitability of a predictor measure. Sometimes, professional judgement should serve as sufficient evidence for a predictor's inclusion in a selection procedure, followed by a long-term effort aimed at subsequent empirical investigation (Schmitt & Chan, 1998).

An analysis of the criteria in pilot training selection in terms of relevancy, deficiency and contamination is essential to future pilot selection studies. It is clear that the current battery is deficient in the sense that it does not include personality, as suggested by the literature study. A promising research avenue is an investigation of the incremental validity of measures of personality in pilot selection, such as the five-factor model of personality (Costa & McCrae, 2003), since dimensions such as conscientiousness could be expected to relate to success in pilot training. Whether personality would help to explain pilot performance better than measures of *g* already do, is surely an important question to consider.

In summary, this research confirms the widely-held belief that measures of general cognitive ability remain as stalwarts in any selection programme for military pilots. The unique contribution of this study to aviation psychology in South Africa is twofold. Firstly, it was shown that spatial ability can significantly enhance the ability to predict pilot flight training success. Measures of specific forms of intelligence (s_n) such as spatial ability could have incremental validity over *g* and should therefore not be discarded in favour of measures of general cognitive ability or fluid intelligence. Other measures of s_n should be investigated in terms of their possible incremental validity. Moreover, it was shown that the use of multiple criteria of performance has value since it provides broader evidence of criterion construct validity and it facilitates a more complete understanding of ability-performance relationships (Schmitt & Chan, 1998).

REFERENCES

Alderton, D.L., & Larson, G.E. (1990). Dimensionality of Raven's Advanced Progressive Matrices items. *Educational and Psychological Measurement*, *50*(4), 887–900.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (1999). *Standards for educational and psychological testing*. New York: McGraw Hill.

American Psychological Association. (2003). *Principles for the validation and use of personnel selection procedures* (4th edn.). New York: APA.

Arthur, W., & Woehr, D.J. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement*, *53*(2), 471–478.

Ashman, A., & Telfer, R. (1983). Personality profiles of pilots. *Aviation, Space, and Environmental Medicine*, *54*(10), 940–343.

Aspeling, E.G. (1980). *Vlieënierskeuringstrategie vir die jare 1980 tot 1990* (Special Report U/Pers 109). Braamfontein: Nasionale Instituut vir Personeelnavorsing (WNNR).

Barkhuizen, W., Schepers, J.M., & Coetzee, J. (2002). Rate of information processing and reaction time of aircraft pilots and non-pilots. *Journal of Industrial Psychology*, *28*(2), 67–76.

Belsley, D.A., Kuh, E., & Welsch, R.E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley & Sons.

Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*(3), 478–494.

Bors, D.A., & Stokes, T.L. (1998). Raven's Advanced Progressive Matrices: Norms for first year university students. *Educational and Psychological Measurement*, *58*(3), 382–398.

Bourn, J. (2000). *Training new pilots, report HC 880 session 1999–2000*. London: National Audit Office, UK Ministry of Defence.

Burke, E., Hobson, C., & Linsky, C. (1997). Large sample validations of three general predictors of pilot training success. *International Journal of Aviation Psychology*, *7*(3), 225–234.

Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco: Jossey-Bass.

Carretta, T.R. (1989). USAF pilot selection and classification systems. *Aviation, Space and Environmental Medicine*, *60*(1), 46–49.

Carretta, T.R. (1992a). Understanding the relations between selection factors and pilot training performance: Does the criterion make a difference? *International Journal of Aviation Psychology*, 2(2), 95–106.

Carretta, T.R. (1992b). Recent developments in U.S. Air Force pilot candidate selection and classification. *Aviation, Space and Environmental Medicine*, *63*(12), 1112–4.

Carretta, T.R. (2000). U.S. Air Force pilot selection and training methods. *Aviation, Space and Environmental Medicine*, *71*(9), 950–956.

Carretta, T.R., & Ree, M.J. (1989). Pilot-candidate selection method: Sources of validity. *International Journal of Aviation Psychology*, *4*(2), 103–117.

Carretta, T.R., & Ree, M.J. (1996). U.S. Air Force pilot selection tests: What is measured and what is predictive? *Aviation, Space and Environmental Medicine*, *67*(3), 279–283.

Carretta, T.R., Perry, D., & Ree, M.J. (1996). Prediction of situational awareness in F-15 pilots. *International Journal of Aviation Psychology*, *6*(1), 21–41.

Carroll, J.B. (1993). *Human Cognitive Abilities*. Cambridge: Cambridge University Press.

Cascio, W.F. (1993). Assessing the utility of selection decisions: Theoretical and practical considerations. In N. Schmitt & W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 310–340). San Francisco: Jossey-Bass.

Cattell, R.B. (1987). *Intelligence: Its structure, growth, and action*. New York: Elsevier.

Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology*, *65*, 407-414.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd edn.). Hillsdale: Erlbaum.

Cook, M. (1999). *Personnel selection*. (3rd edn.). Chichester: Wiley.

Costa, P.T., & McCrae, R.M. (2003). *A comprehensive and detailed assessment of adult personality based on the Five-Factor Model: NEO Personality Inventory-Revised (NEO PI-R)*. Lutz: Psychological Assessment Resources.

Croucamp, Y., & Bolton, S. (2002). *Pilot selection, statistical report*. Pretoria: Department of Defence (RSA), Military Psychological Institute.
Damos, D.L. (1996). Pilot selection batteries: Shortcomings and perspectives. *International Journal of Aviation Psychology*, *6*(2), 199–209.
Employment Equity Act of the Republic of South Africa. (1998). *Government Gazette* (No. 28858).
Gatewood, R.D., & Field, H.S. (1998). *Human resource selection.* (4th edn.). Orlando: Harcourt Brace.
Griffin, G.R., & Koonce, J.M. (1996). Review of psychomotor skills in pilot selection research of the U.S. military services. *International Journal of Aviation Psychology*, *6*(2), 125–147.
Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill.
Guion, R.M. (1965). *Personnel testing*. New York: McGraw-Hill.
Hilton, T.F., & Dolgin, D.L. (1991). Pilot selection in the military of the free world. In R. Gal & A.D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 81–101). New York: Wiley.
Hunter, D.R., & Burke, E.F. (1994). Predicting aircraft pilot training success: A meta-analysis of published research. *International Journal of Aviation Psychology*, *4*(4), 297–313.
Hunter, J.E., & Schmidt, F.L. (1990). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills: Sage.
Huysamen, G.K. (1994). *Methodology for the social and behavioural sciences*. Halfway House: Southern.
Huysamen, G.K. (1996). *Psychological measurement*. (3rd edn.). Pretoria: Academic.
Jensen, A.R. (1998). *The g factor*. Westport: Praeger.
Lambirth, T.T., Dolgin, D.L., Rentmeister-Bryant, H.K., & Moore, J.L. (2003). Selected personality characteristics of student naval aviators and student naval flight officers. *International Journal of Aviation Psychology*, *13*(4), 415–427.
Landy, F.J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, *41*(11), 1183–1192.
Lombard, R.B. (1980). *BLOX test administrator's manual (A/80).* Pretoria: CSIR.
Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology*, *6*(1), 1–20.
Martinussen, M., & Torjussen, T. (1998). Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. *International Journal of Aviation Psychology*, *8*(1), 33–45.
Milkovich, G.T., & Boudreau, J.W. (1997). *Human resource management*. (8th edn.). Chicago: Irwin.
Murphy, K.R., & Davidshofer, C.O. (1988). *Psychological testing: Principles and applications*. Englewood Cliffs: Prentice Hall.
Plug, C., Meyer, W.F., Louw, D.A., & Gouws, L.A. (1989). *Psigologie woordeboek*. Johannesburg: Lexicon.
Raven, J.C., & Court, J.H. (1998). *Manual for the Ravens Advanced Progressive Matrices.* London: H.K. Kewis.
Ree, M.J., & Carretta, T.R. (1996). Central role of g in military pilot selection. *International Journal of Aviation Psychology*, *6*(2), 111–123.
Ree, M.J., & Carretta, T.R. (2002). g2K. *Human Performance*, *15*(1), 3–23.
Ree, M.J., & Earles, J.A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*(2), 321–332.
Retzlaff, P.D., & Gibertini, M. (1987). Air force pilot personality: Hard data on the "right stuff". *Multivariate Behavioral Research*, *22*(4), 383–399.
Rumsey, M.G., Walker, C.B., & Harris, J.H. (Eds.). (1994). *Personnel selection and classification*. Hillsdale: Lawrence Erlbaum.
Rushton, J.P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence*, *31*(2), 123–137.
Sackett, P.R., & Arvey, R.D. (1993). Selection in small N settings. In N. Schmitt & W.C. Borman (Eds.). *Personnel selection in organizations* (pp. 418–447). San Francisco: Jossey-Bass.
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*(2), 262–274.
Schmitt, N., & Borman, W.C. (Eds.). (1993). *Personnel selection in organizations*. San Francisco: Jossey-Bass.
Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks: Sage.
Siem, F.M. (1992). Predictive validity of an automated personality inventory for air force pilot selection. *International Journal of Aviation Psychology*, *2*(4), 261–270.
Smit, C., & Bielfeld, R. (2001). *Progress report of pilot selection data*. Pretoria: Department of Defence, Military Psychological Institute.
Society for Industrial and Organisational Psychology of South Africa. (1998). *Guidelines for the validation and use of assessment procedures for the workplace.* Auckland Park: SIOPSA.
Spearman, C. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.
SPSS Inc. (2006). *Base 15.0 applications guide*. Chicago: SPSS.
Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. (4th edn.). Needham Heights: Allyn & Bacon.
Thorndike, R.L. (1949). *Personnel selection: Test and measurement techniques*. New York: Wiley.
Thorndike, R.L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior*, *29*(3), 332–339.
Turnbull, G. (1992). A review of military pilot selection. *Aviation, Space and Environmental Medicine*, *63*(9), 825–830.
United States Department of Labor. (1978). *Uniform guidelines on employee selection procedures.* Washington, DC: US Department of Labor.
Van der Merwe, R.P. (2002). Psychometric testing and human resource management. *SA Journal of Industrial Psychology*, *28*(2), 77–86.

Article #753
SA Journal of Industrial Psychology