2007

Prediction Differs from Description: General Versus Specific Intelligence Testing for the Selection of Ab Initio Air Traffic Controllers

Stefan Hoeft

Marc Damitz

André Beauducel

PREDICTION DIFFERS FROM DESCRIPTION: GENERAL VERSUS SPECIFIC INTELLIGENCE TESTING FOR THE SELECTION OF AB INITIO AIR TRAFFIC CONTROLLERS

Stefan Hoeft
German Aerospace Center (DLR), Dep of Aviation and Space Psychology
Hamburg, Germany

Marc Damitz
Eurocontrol, Human Resources Directorate
Brussels, Belgium

André Beauducel
University of the Armed Forces, Dep of Psychology
Hamburg Germany

Meta-analyses and recent large-scale primary studies concerning the importance of intelligence indicate that for almost all jobs general mental ability (g) alone predicts performance well. However, there is a controversy concerning the question whether specific abilities (s) are needed to predict job or training performance. In the present study performance test data of 5223 applicants from the DLR program for selection of ab initio air traffic controllers at DFS Deutsche Flugsicherung GmbH were analyzed. Results of different approaches based on exploratory factor analysis and structural equation modeling to estimate g- and s-intelligence components were compared. In addition to this, the criterion-related validities of different g- and s-measures were tested using training performance criteria from 282 DFS trainees. It is argued that the preference for an intelligence model depends in part on the utilized theoretical approach and the objectives of the diagnostician (description vs. prediction).

Introduction

Meta-analyses concerning the importance of intelligence indicate that for almost all jobs general mental ability (g) alone predicts performance well (Hunter & Schmidt, 1998). However, there is a controversy concerning the question whether specific abilities (s) are needed to predict job or training performance. Especially in studies on the selection of aviation personnel it was argued that not much more than g was needed to predict different performance criteria (e.g., Ree & Carretta, 1998). Recently, the debate was refreshed by a paper published by Brown, Le and Schmidt (2006). Based on a large data set of U.S. Navy job groups they argued that specific aptitude theory and its implied test approach are not tenable, because they do not improve the prediction of training performance beyond the variance already explained by general mental ability.

Aim of our study

In the present study we investigate different statistical approaches to operationalize g- and s-components of mental ability tests. Further on, we investigate the relevance of these different g- and s-components for the prediction of job performance.

We hypothesize that different approaches to measure g and s lead to different prognostic validities of these components.

Theoretical background:

Different approaches to operationalize g and s

According to Spearman's approach, every measure of ability can be split up in a general (g) and a specific component (s). In case of a selection battery g is often interpreted as the first component of an unrotated Principle Component Analysis. Specific abilities are represented by the remaining components.

Modern hierarchical approaches (e.g., Carroll, 1993; Bucik & Neubauer, 1996) use a structural equation modeling approach (confirmatory factor analysis, CFA) and interpret g as a general second order factor with loadings on specific primary first order factors. In this approach, specific abilities are represented by the residual influences on the primary factors.

The main difference between these approaches is based on the allocation of residual variance representing specificity: In Spearman’s approach it is more or less homogenously distributed across all low-ranking components and loosely connected to specific manifest variables. In CFA, residual influences representing specificity are connected to meaningfully grouped manifest variables (e.g., the residual of a first order factor loading on two spatial orientation tests represents aspects of spatial orientation not explained by g).

Method

Subjects. The descriptive analyses are based on data of N=5223 applicants for ab initio air traffic controller positions at DFS. The predictive analyses are based on sub-sample data of N=282 DFS trainees. No differentiation was made for gender and age groups.

Instruments. Analyses are based on mental ability tests executed in the pre-selection phase of the DLR program for the selection of ab initio air traffic controllers at DFS Deutsche Flugsicherung GmbH.

The specific tests are described in table 1.

Table 1. Description of the DLR aptitude test battery analyzed

	test description	**content**
KRN	basic mental arithmetic tasks with auditory stimulus	*mental arithmetic*
RAG	Arithmetic tasks in applied settings	
ARA	Path figure drawing according to "left" and "right" oral instructions	*spatial orientation*
FPT	Mental rotation of an airplane shape according to written instructions	
BOU	Marking of critical letters in a nonsense text of random letters	*attention and concen-tration*
KBT	summation of numbers represented by different symbols	
WSB	Subtraction of fronts and backs of shown dices	
VIG	Longtime monitoring of sparse visual and auditory stimulus constellations	*vigilance*
CLE	Comparison of memorized letter combinations with similar auditory stimuli	*Memory*
MEK	Memorized number-symbol-associations have to be reproduced given the symbol alone	
ACT	Basic approach control test: Planes have to be guided safely and efficiently to their final destination	*basic work sample*
ENA	Acoustic English test for comprehension and translation	*English proficiency*
ENS	English test for word fluency, grammar and meaning	

Successful applicants (subjects with at least average scores in all domains) had been invited to the main selection phase with extensive work sample tests, assessment center exercises and a semi-structured biographical interview.

Criteria. Seven different training performance criteria are investigated in dthe predictive analysis, namely results of four final theoretical exams (basics, navigation, meteorology, aeronautics) and three results of practical training (radar center, coordination center, radar approach).

Results

Descriptive analyses

The following analyses are based on the total scores of the thirteen different metal ability tests.

Principal component analysis (PCA). The PCA reveals a strong first component (Eigenvalue 4.92, 37.9%) followed by two additional components with substantial Eigenvalues (>1). All thirteen components (g + 12 s) had been extracted for further analyses.

Confirmatory factor analysis (CFA). A CFA model with a general 2nd order factor and seven content specific 1st order factors gains a definitely better fit (χ^2=2126.78 with df=60, GFI=.936, AGFI=.903, RMSEA=.085) than a simple general 1st order factor (χ^2=7238.83 with df=65, GFI=.835, AGFI=.768, RMSEA=.140). However, a CFA model with intercorrelated specific 1st order factors produces a even better fit (χ^2=1500.77 with df=48, GFI=.954, AGFI=.912, RMSEA=.080). Although this model fits best, on the basis of conceptual reasons (separation of general and specific influences) the 2nd order model was chosen for the predictive analyses.

Detailed information for this model concerning factor loadings and amount of specific variances are given in figure 1.

Intercorrelation of PCA and CFA factor scores. The g scores seem to be identical (intercorrelation r=.991), but the specific factors reveal a

heterogeneous overlap. Only three of the seven CFA specificities show a remarkable intercorrelation (| r | >.5) with only one PCA low-ranking component. All other specific factors correlate substantially with two or three PCA components simultaneously.

Predictive analyses

The prognostic validity is analyzed by means of multiple regressions with g- and s-components serving as predictors and performance indices as criteria. The results are presented in table 2.

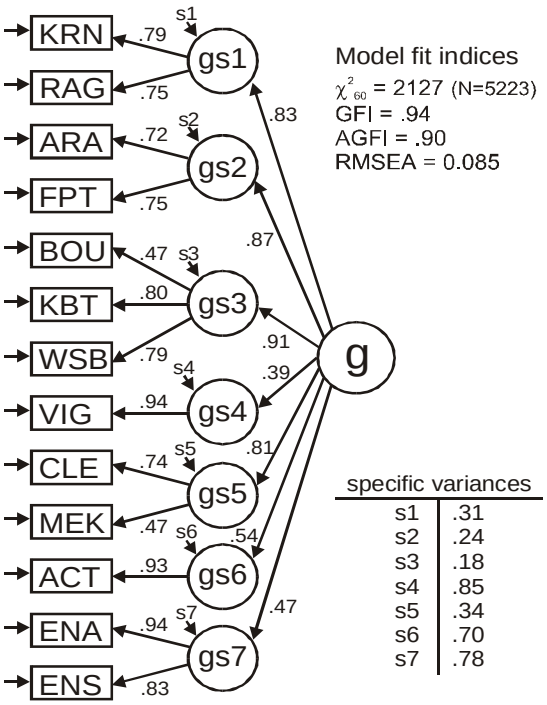

Figure 1. Graphical representation of the 2nd order CFA model. *Note.* g = general 2nd order factor; gs1-7 = 1st order factor representing confounded g and s; s1-7 = residual representing specific influences independent from g

The g component is entered in the first step; all specific components follow in the second step. All estimates are corrected for variance restriction using multivariate correction (ref. to Lawley, 1943).

Remarkable increases are noticeable especially for the performance criteria (e.g., coordinator center criterion). The increase is somewhat smaller for the CFA specificities.

Table 2. Regression results for PCA and CFA predictors

	PCA-predictors (corrected)			CFA-predictors (corrected)		
Criterion	r_g	R_{g+s}	ΔR	r_g	R_{g+s}	ΔR
Theoretical training						
Basics	.27	.36	.09	.29	.37	.08
Navigation	.37	.45	.08	.36	.46	.10
Meteorology	.25	.33	.08	.27	.31	.04
Aeronautics	.20	.32	.12	.23	.30	.07
Practical training						
Radar Center	.22	.32	.10	.24	.29	.05
Coordinator Center	.02	.31	.29	-.03	.19	.19
Radar Approach	.26	.42	.16	.25	.33	.08

Conclusions

In this paper we compared PCA and CFA approaches to differentiate general and specific aspects of mental ability tests in relation to their descriptive and predictive utility.

For descriptive purposes, the utilized CFA model with content-specific primary factors and a general 2nd order factor seems to be preferable to the Spearman model: The goodness-of-fit statistics indicate an acceptable model fit and the residual factors are interpretable in a meaningful way equivalent to the diagnostic usage of these terms.

For predictive purposes, the CFA specificities seem to be inferior compared to their PCA equivalents. This could be a combined result of the smaller number of factors (7 vs. 12) and the exclusive measurement of content- but not test-specific aspects. Nonetheless, both approaches (PCA and CFA) gain substantial incremental validity for specific criteria.

It has to be kept in mind that the actual choice of predictors and criteria represents a major influence for this kind of research:

The preference for a special intelligence model depends on different aspects of predictor selection (results of job analysis, test choice, chosen statistical procedure to measure g and s). The predictive validity of (g and/or s) ability aspects is also moderated by influences like the actual definition of job performance, availability of criteria and scope of analysis (bandwidth and fidelity).

References

Brown, K. G., Le, H., & Schmidt, F. L. (2006). Specific aptitude theory testing revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment, 14,* 87-100.

Bucik, V. & Neubauer, A. C. (1996). Bimodality in the Berlin model of intelligence structure (BIS): A replication study. *Personality and Individual Differences, 21,* 987-1005.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge, UK: Cambridge University Press.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62,* 28-30.

Ree, M. J. & Carretta, T. R. (1998). General cognitive ability and occupational performance. In C.L.Cooper & I. T. Robertson (Eds.), *International Review of industrial and organizational psychology, Vol. 13* (pp. 159-184). New York: John Wiley & Sons.

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.