ASVAB: E Pluribus Unum?

Martin J. Ippel, Ph.D.
CogniMetrics Inc.,
San Antonio, TX

&

Stephen E. Watson, Ph.D.
Navy Selection and Classification (CNO N132)
Washington, DC.

Paper presented at the 50th Annual Conference of the
International Military Testing Association
Amsterdam
September 29 – October 3, 2008

Summary

Is the meaning of the Armed Services Vocational Aptitude Battery (ASVAB) tests invariant throughout the full range of test scores? The answer to this question has important practical implications. It concerns the issue of the population of interest in selection models for the U.S. Armed Services. Should it be the general population, or should different subpopulations be tested with different selection batteries?

Recent studies with comprehensive tests of basic abilities demonstrated two empirical phenomena that seem to cast doubt on the assumption of measurement invariance: (a) the decrease in positive manifold of cognitive variables in samples with higher levels of "g", (b) lower loadings of "g" on cognitive tests in high-g samples. This paper argues that these phenomena were effects of selection on a latent variable (i.e., the latent variable "g") and can be derived from the classic Pearson-Lawley selection rules. The present study tested the selection-effects hypothesis by formulating a multi-group common factor model for samples that systematically differ on the latent variable "g" (i.e., g-lo, g-av, and g-hi). These samples ($N_{g\text{-}lo}=N_{g\text{-}av}=N_{g\text{-}hi}=600$) were systematically sampled from a parent population (N=48,222) of Air Force recruits using "g" as the selection variable. The factor loadings appeared invariant across the g-samples, but the mean structure of the tests changed in the selected samples, that is, the intercepts of the linear equations modeling the regression of the test scores unto the latent variables differed significantly between the three g groups. Implications for the ASVAB and the selection and classification process will be discussed.

Is the meaning of the Armed Services Vocational Aptitude Battery (ASVAB) tests invariant throughout the full range of test scores? The answer to this question has important practical implications. It concerns the issue of the population of interest in selection models for the U.S. Armed Services. Should it be the general population, or should different subpopulations be tested with different selection batteries? Intuitively it is hard to imagine that a candidate for the highly complex training course for Navy Electronics Technician-Submarine (ET SS_N) can be considered sampled from the same population of interest as a candidate for the Navy Engine Man (EN) course.

Recent studies with comprehensive tests of basic abilities demonstrated two empirical phenomena that seem to cast doubt on the assumption of measurement invariance: (a) the decrease in positive manifold of cognitive variables in samples with higher levels of "g" (Abad, Colom, Juan-Espinoza, Garcia, 2003; Deary, Egan, Gibson, Austin, Brand, and Kellaghan, 1996; Detterman and Daniel, 1989; Lynn, 1992). A similar effect was found with the Swedish Enlistment Battery (Carlstedt, 2001); (b) lower loadings of "g" on cognitive tests in high-g samples. At least two studies with the ASVAB report lower correlations between the tests at higher levels of g (Evans, 1999; Legree, Pifer & Grafton, 1997).

This paper argues that these phenomena were effects of selection on a latent variable (i.e., the latent variable "g") and can be derived from the classic Pearson-Lawley selection rules. Using these selection rules, Meredith (1964) showed that both the covariance structure and the mean structure in the selected samples are expected to change as a function of selection based on one or more latent variables. In a later paper, Meredith (1965) presented procedures to derive a single best fitting factor pattern (thus: invariant factor pattern) from a set of factor solutions obtained on populations differing on a latent variable (see also: Harman, 1970, p 268). Jöreskog (1971) formalized this viewpoint as an extension of the common factor model for a parent population to multiple groups generated from this population based on one or more latent variables in the model.[1] The present study tested the selection-effects hypothesis put forward by Meredith (1964) by formulating a multi-group common factor model for groups generated from a parent population based on a latent variable "g".

Measurement Invariance

We will first present a more precise definition of measurement invariance. Measurement invariance becomes an issue when we compare groups, or individuals from different groups. The requirement of measurement invariance is based on a simple and intuitive notion, namely, that the expected value of test scores of a person of a given level of ability should be independent of membership of these groups (Mellenbergh, 1989). In formula:

$$f\,(Y \mid \eta, v) = f\,(Y \mid \eta), \qquad [1]$$

where Y is the (manifest) test score, η is a given set of (latent) abilities and v refers to a set of groups for which equation [1] holds. The formulation of this model is very general. The specific form of the function f depends on the measurement model of choice. For example, a linear regression model would relate Y and η as follows:

$$y_{1ij} = \tau_{1i} + \lambda_{1i}\,\eta_{ij} + \varepsilon_{1ij} \qquad [2]$$

for a test score y_1 of person j in group i. The requirement of measurement invariance of the score y_1 imposes constraints on the model parameters τ_{1i}, λ_{1i}, and ε_{1ij}. In the most stringent form of measurement invariance these parameters should be across-group invariant for all groups $i = 1, \ldots , v$. Ideally, the test scores should only vary with the level of η, that is, the ability that the test is supposed to measure. The λ parameter expresses the relationship between the test score y and the latent variable η. The

---

1 Muthén (1989) derived comparable results for observable variables (see also Dolan & Molenaar, 1994).

intercept in Equation [2] represents all influences other than the factors modeled. Notice that, even when $\lambda_{1m} = \lambda_{1n}$ for every factor η_m and η_n in the model, differences between the intercepts (i.e., $\tau_{1m} \neq \tau_{1n}$) can invalidate across-groups comparisons.

We tested this selection-effects hypothesis in a statistical experiment by defining a latent variable "g" on a sample drawn from a database of scores on ASVAB tests of 48,222 Air Force recruits (i.e., the parent sample) and then drawing three samples from this parent population, based on their scores on "g", while systematically manipulating $E[\eta_i]$ and controlling $V[\eta_g]$ of each of these samples. The expectation was that if these samples were correctly considered samples from the same parent population and only differing in "g", a multi-group confirmatory factor analysis would show λ_{11}, ε_{1ij},and τ_{1i} to be invariant.

Defining the latent variable "g"

The Armed Services Vocational Aptitude Battery (ASVAB) is a test battery consisting of nine tests, which in different configurations are used in the recruitment, selection and classification of the Armed Forces. Table 1 lists the ASVAB tests and their measurement claims. The ASVAB does not deliver a comprehensive score other than the AFQT score, which is basically a measure of scholastic ability based on a subset of tests WK, PC, AR and MK. Other composites scores (e.g., the Navy's Minimum ASVAB Selection Composite scores) are also based on linear combinations of subsets of tests and are used to assign new recruits to military occupations (Sellman, 2004).

[Insert Table 1 about here]

Factor-analytic studies support a Holzinger-Spearman Bi-Factor Model for the ASVAB tests. Each ASVAB test loads on the general factor and a group factor. Four group factors have been identified and could be replicated in various studies, viz., Verbal Ability (Verbal), Quantitative Ability (Quantitative), Speed, and Technical Knowledge (Kass, Mitchell, Grafton, & Wing, 1983; Welsh, Kucinkas, & Curran, 1990). One test diverged from this pattern. General Science (GS) has split loadings on the Verbal and Technical Knowledge factors (Kass et al., 1983). In later versions of the ASVAB the Speed factor has been taken out and the Speed tests have been replaced by a single test Assembly of Objects (AO), a Visualization test.

The first analysis in this paper is designed to derive factor scores for the general factor, which then can be used to sample recruits from a parent database.

Method

Participants

Participants were 48,222 U.S. Air Force recruits whose data records were made available for analysis. The recruits were ranging in age from 17 to 22 years. The tests were administered during their Basic Military Training at Lackland AFB, TX, as part of a routine assessment of cognitive skills.

Instruments

The analysis involved data collected on nine subtests from the Armed Services Vocational Aptitude Battery (ASVAB). The tests and their measurement claims are listed in Table 1.

Procedure

The intention was to generate three samples (Ns = 600) from the parent population such that in each sample the latent variable “g” had an identical probability distribution, except for the first moment of the distributions. The means of the samples were (1) 1.5 standard deviation above the mean of the parent population (g-hi sample), (2) identical to the parent population (g-av sample) and (3) 1.5 standard deviation below the population mean (g-lo sample). Since the samples represent more homogeneous sub populations of the parent population with respect to "g", it was decided that the sample variance of this latent variable should be smaller than the population variance. Therefore, the sample standard deviations were (arbitrarily) set equal to 0.5 of the population standard deviation.

There is no readily available software for drawing samples with specified distributional characteristics from databases, and a statistical program in MS Visual Basic was written for this purpose. To produce an "ideal" distribution with specifiable characteristics, the program exploits the MS Excel utility to generate random numbers according to a specified distribution model (e.g., Normal, Binomial, Poisson) and parameters (e.g., mean, standard deviation, number of observations). The program was set to generate 600 random numbers according to a normal distribution with mean equal to zero and standard deviation equal to one. It then sorts these random numbers in descending order. The purpose of the program is to select cases from the database that closely match the elements of this "ideal" sample.

In order to accomplish this, the program arranged the cases in the database in descending order according to their standardized score on the latent variable (in this case a standardized g-score). A case will be selected if it matches an element from the "ideal" distribution with a difference smaller than 0.005.

Models

The factor structure of the ASVAB has been investigated over the years by a number of researchers. Two features characterize the ASVAB factor structure: (1) it has a dominant general factor accounting for approximately 60 percent of the variance (Kass, Mitchell, Grafton, & Wing, 1983; Welsh, Watson, & Ree, 1990), (2) in a number of studies group factors have been identified and could be replicated, viz., verbal ability (Verbal), quantitative ability (Quantitative), Speed, and Technical Knowledge (Kass et al., 1983; Welsh, Kucinkas, & Curran, 1990).

[Insert Figure 1 about here]

To implement the first characteristic we developed and tested two different models: (1) a hierarchical model in which the general factor is represented as a second-order factor (see Figure 1), and (2) a "g as first principal" model in which the general factor is realized in a way roughly similar to the first principal component in an exploratory factor analysis: it directly loads on every test in the battery. In this model "g" is independent of the other factors (see Figure 2).

[Insert Figure 2 about here]

The difference between the models is the following. In model 1 "g" explains the correlation between the first-order factors. The tests (with the exception of the AO tests) are subsumed under the first-order factors. In model 2 all factors are directly involved in explaining the covariance pattern among the tests. The factors Verbal, Quantitative and Technical Knowledge explain covariance that is not explained by "g". In model 2 we also implemented the split loadings of General Science (GS) on the Verbal and Technical Knowledge factors (Kass, Mitchell, Grafton, & Wing, 1983).

Factor Model Tests

The purpose of fitting the different models of "g" to a random sample from the Air Force database was to calculate g factor scores to define three different levels of g in the Air Force database. Since a linear structural equation modeling analysis cannot be meaningfully done on the complete database (N = 48,222), we decided to take random samples. In order to ensure that the sample used for the estimation of the g factor scores was maximally representative for the entire database, we first calculated the eigenvalues of the correlation matrix of the entire database and then compared the results with six different random samples of approximately 1000 recruits. Table 1 shows the results.

[insert Table 2 about here]

Notice that analysis of the eigenvalues of the database correlation matrix showed that the first factor, on which all tests had a significant loading, explained 43.67% of the total variance. Furthermore, only three eigenvalues of the database correlation matrix were larger than one. Based on a comparison of six samples from the Air Force database, we decided to use sample 1 for our analysis. In this sample the first eigenvalue corresponded with 44.12% of the total variance.

Selection Effects on the latent variable "g"

Table 3 summarizes distributional characteristics of the different g samples. Notice the almost perfectly identical distributional features of the three different g samples. Except for the difference in standard deviation the distributional differences between the random sample and the g-av sample were minimal.

[Insert Table 3 about here]

Analysis of the eigenvalues of the matrix of correlations between the nine ASVAB tests suggested a decrease in the level of average correlation in the higher g samples, viz., 0.221, 0.204, and 0.177 for the g-lo, g-av and g-hi samples, respectively. Altogether, the average correlations in the g samples were much lower than the level of average correlation in the random sample ($r_{..}$ = 0.371).[2] The decrease in average level of correlations was an almost perfect linear function of the reduction in test variances within the samples (R^2 = 0.997). Compared to the random sample, the average reduction in test variance was 23.1%, 29.4% and 41.8% in lo-g, av-g and hi-g samples, respectively. Notice that these differences in test variance reduction occurred despite the fact that the samples were selected such that the second (i.e., variance), third (i.e., skewness), and fourth moments (i.e., kurtosis) of the probability distributions of the g factor were identical (see Table 3). Only the first moments (i.e., the means) were to differ.

The decision to set the sample standard deviations at one half of the standard deviation of the random sample implied a reduction of g variance compared to the random sample of the parent population. In this case the reduction was 76.5% of the variance in g. Since the g factor explained 43.67% of the total variance in the random sample (see p.10), the reduction in g variance amounts to a reduction of the total variance of approximately 33% in the three g-samples, assuming all tests were affected equally by the different g levels in the samples.[3] This appeared not to be the case. In the sequel we analyze the differential effects of sampling on g on the means and variances of the ASVAB test scores.

---

[2] We applied Kaiser's (1968) formula: $\mathbf{r}_{..} = (\lambda - 1)/(p - 1)$, where λ stands for the first eigenvalue of the correlation matrix and p for the number of variables (i.e., tests).

[3] Notice that the latter value is quite close to the average reduction in test variance in the three g samples (i.e., 31.4%).

Insert Figure 3 about here

Figure 3 shows the differences in mean test scores of the ASVAB tests in the three different g samples. Tests loaded by the Quantitative Ability (Q) factor, that is, Mathematical Knowledge (MK) and Arithmetic Reasoning (AR), followed by Mechanical Comprehension (MC), a test loaded by the Technical Knowledge (TK) factor, showed the largest differences in means. Figure 4 displays the differential effects of sampling on g on the test variances (i.e., the reduction in test variances compared to the random sample). Again the tests Arithmetic Reasoning (AR), Mathematical Knowledge (MK) and Mechanical Comprehension (MC) were most affected, especially in the g-hi sample. Figure 4 suggests that in particular the g-hi sample is homogeneous with respect to these tests.

Insert Figure 4 about here

We tested the invariance of, λ_{1i}, ε_{1ij}, and τ_{1i} for the ASVAB tests against the three data sets by formulating a sequence of MGCFA models in which step by step the invariance constraints where further relaxed.

Insert Table 5 about here

Table 5 summarizes the results of the model tests. Model 1 is the model in which all invariance constraints were implemented: the factor loadings matrix Λ, vector ν, and the diagonal vector Θ were all set invariant across samples. This model is usually referred to as a Full Measurement Invariant (Full MI) model. Testing this model investigated whether individual differences within-group and between-group differences in test scores can be considered differences on the same latent dimension (i.e., generated by the latent factor "g" as defined in the parent population). The model was rejected (see Table 5). A slightly weaker version of the model, which defined residual variances as invariant within each sample resulted in a significant improvement of the model fit (χ^2 difference = 2396.991, df (difference) = 3, $p < 0.001$). The third model, in which the residual variance parameters were left free to vary, showed a further significant improvement in fit (χ^2 (difference) = 454.974, df (difference) = 24), $p < 0.001$). The PCLOSE index in this analysis suggested that the model fit was fairly close (discrepancy with the sub population data was certainly less then 5 percent). In fact, the RMSEA index was at 0.034 (i.e., 3.4 percent).

Insert Table 5 about here

A further relaxation of the model, by letting the intercepts free to vary, could only partially be tested, given the objective of estimation of group factor means, which requires the matrices Λ_i constrained to invariance and the invariance of at least some elements of the intercept vectors. The fourth row in Table 5 reports the model fit when the intercepts of WK, PC, AO and GS were free to vary. The intercepts of the tests PC, AR, MK, AS and EI were kept invariant across the samples to keep the model identified.

The result was a further significant improvement of the model fit (χ^2 (difference) = 45.22, df (difference) = 8, $p < 0.001$). The test outcome suggested that the invariance of intercepts was an untenable constraint on the model, a suggestion that seemed numerically plausible given the considerable differences in mean observed test scores between the samples (see Figure 3). As a result of relaxing the constraint on the intercepts, the residual means were much smaller for the model with (at least some) free intercepts. We accepted this model version for further analysis.

One outcome of a multi-group confirmatory factor analysis (mcfa) is a single set of best fitting regression estimates (i.e., factor loadings) that holds for all samples in the analysis. These results, together with their standard errors of measurement and significance levels are presented in Table 6. The shaded areas in the table contain the estimates for the same λ parameters in the aselect sample from the parent population.

All g-factor loadings in the g-samples were significant at 0.01-level, except the g-loading for Electronic information (EI), which was non-significant, and Math Knowledge (MK) and Auto Shop (AS), which were only marginally significant ($p < 0.05$). EI appeared to be almost exclusively a technical factor with some influence of verbal ability. MK had rather high loadings of the quantitative factor.

Due to differences in the variances of tests scores and latent variables in the three samples the standardized regression estimates may vary per sample. These results are reported in Table 7. We report the standardized regression estimates (or factor loadings) per g-sample because that is what is reported in the factor analytic studies that preceded the present study.

Insert Table 6 about here

Table 7 presents a comparison of factor loadings of the ASVAB tests in the aselect sample from the parent population and the three g-samples. The loadings of the g-factor were much lower in the three g-samples than in the a-select sample from the parent population, while the g-factor loadings in the g-hi sample were higher than in the other g-samples. In all g-samples the factor loadings on the group factors (i.e., Verbal Knowledge, Quantitative Knowledge and Technical Knowledge) were equal to or higher as compared to the a-select sample from the parent population.

General Discussion

We will first discuss the two empirical phenomena that were mentioned at the outset of this paper. First, the decrease of positive manifold of cognitive variables in samples with higher levels of g. This fact was replicated in this study. Moreover, it could be shown to be strongly related to reduction in test variance in higher g-samples -– a selection phenomenon.

Second, the lower loadings of g on cognitive tests in high-g samples can be understood in similar way. Figure 4 shows that the reduction in test variance from the random sample to the g-hi sample is strongest on tests with the highest relevance for g, viz., Arithmetic Reasoning (AR), Mechanical Comprehension (MC) and Object Assembly (AO), in this order. The g-hi sample obviously is more homogeneous with respect to performance on those tests. This would imply a reduced factor loading of g on such a test in higher g samples.

At a more conceptual level the issue is the measurement invariance of the ASVAB test battery throughout the range of test scores. Meredith (1993) proposed a concept of measurement invariance (MI) that provides the necessary and sufficient conditions to determine whether a set of measures has the same underlying factors in several groups. In his conception MI implies that the only difference between groups involve the relation of the observed scores to the latent variables, not the factor means and the factor covariances. To conclude that the tests (or items) measure the same across groups the sets of τ, λ, and ε parameters need to be identical across groups. Thus, if Eq. (1) holds across groups with identical parameter values, and even though the mean and the covariances of the factors η may differ, it can be concluded that the same factors are measured across groups. If the constraints on all three sets of parameters hold, then a test battery is considered to satisfy the condition of "strict factorial invariance".

Looking at the results of our analysis, the matrix Λ, representing the factor loadings λ, appeared to be invariant across the different levels of "g". However, the error term in Eq [2], that is, the vector ε could not be constrained to be equal. Notice that the error term of Eq (1), that is, vector ε, not only contains a random error component, but also an unique variance component (i.e., dependable test variance not explained by the factor model). Conceptually, this is not the most serious threat to MI provided that the test communalities are fairly high. This is, however, not the case (see Table 6).

A further step of model testing involved the parameters of the intercepts vector τ**.** Although the model in which the test intercepts were set equal across the samples, showed an acceptable fit, relaxing this constraint significantly improved the model fit. Therefore, it was concluded that the intercepts in Eq. (1), most likely, were not invariant across samples. This is a serious violation of measurement invariance, which becomes clear if we consider the practical implication of a systematic difference in intercept between two g-levels. Even if the factor loadings matrix (i.e., matrix Λ) and the residuals matrix Θ would be invariant across samples, it would still be the case that individuals from one sample (most likely individuals from the higher g-sample) consistently score higher given identical scores on the latent variables. In other words, the expected value of the observed scores would depend on the subpopulation from which the individual is drawn.

Acknowledgements

The opinions expressed in this paper are those of the authors, are not official, and should mot be construed as reflecting the views of the U.S. Department of Defense.

Research for this paper was performed under contract number FA 3089-04-F-430.

References

Abad, F. J., Colom, R., Juan-Espinoza, M., Garcia, L.F. (2003). Intelligence Differentiation in Adult Samples. *Intelligence, 31*, 157-166.

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of levelof general intelligence: A latent variable approach. *Multivariate Behavioral Research, 36,* 589-609.

Deary, I.J., Egan, V., Gibson, G.J., Austin, E.J., Brand, C.R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*, 105 – 132.

Detterman, D.K., & Daniels, M.H. (1989). Correlations of mental tests with *each other and with cognitive abilities are highest for low IQ groups. Intelligence*, *16*, 349-359.

Dolan, C.V. and Molenaar, P.C.M. (1994). Testing Specific Hypothesis Concerning Latent Group Differences in Multi-Group Structural Analysis with Structured Means. *Multivariate Behavioral Research, 29*, 3, 203-222.

Evans, M.G. (1999). On the asymmetry of g. *Psychological Reports, 85*, 1059-1069.

Harman, H.H. (1970). Modern Factor Analysis. Chicago: The University of Chicago Press.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several *populations. Psychometrika, 36*, 409-426.

Kass, R.A., Mitchell, K.J., Grafton, F.C., & Wing, H. (1983). Factorial Validity of the Armed Service Vocational Aptitude Battery (ASVAB), Forms 8, 9, and 10; 1981 Army applicant Sample. *Educational and Psychological Measurement, 43*, 1077-1087.

Larson, G. E., & Wolfe, J. H. (1994). Validity results for g from an expanded test base. *Intelligence, 20*, 15-25.

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

Legree, P.J., Pifer, M.E., Grafton, F.C. (1997). ASVAB Correlations are lower for higher aptitude groups. *Army Research Institute Report nr. A925823*. (20 pp).

Lynn, R. (1992). Does Spearman's g decline at high levels? Some evidence from Scotland. *Journal of Genetic Psychology, 153*, 229-230.

Lord, F.M. & Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley Publishing Company.

Meredith, W. (1964). Notes on Factorial Invariance. *Psychometrika, 29*, 177-185.

Meredith, W. (1965). A method for studying differences between groups. Pychometrika, 30, 15-129.

Meredith, W. (1993). Measurement Invariance, factor analysis, and factorial invariance. *Psychometrika, 58*, 525-543.

Muthen, B.O. (1989). Factor structure in groups selected on observed scores. *British Journal of Mathematical and Statistical Psychology, 42*, 81-90.

Sörbom, D. (1974). A general method for studying differences in factor means and facture structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 99-115.

Welsh, J.R., Kucinkas, S.K., & Curran, L.T. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies* (AFHRL-TR-90-22). Brooks, AFB, TX: U.S. Air Force Human Resources Laboratory.

Welsh, J.R., Watson, T.W., & Ree, M.J. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Predictiing military criteria from general and specific abilities* (AFHRL-TR-90-63). Brooks, AFB, TX: U.S. Air Force Human Resources Laboratory.

Table 1: ASVAB tests and measurement claims

- **General Science (GS)**: a 25 items knowledge test of physical and biological sciences.
- **Arithmetic Reasoning (AR)**: a 30 items arithmetic word problem test.
- **Word Knowledge (WK)**: 35 items testing knowledge of words and synonyms.
- **Paragraph Comprehension (PC)**: 15 items testing the ability to extract meaning from short paragraphs.
- **Auto and Shop Information (AS):** a 25 items knowledge test of automobiles, shop practices, tools and tool use.
- **Mathematical Knowledge (MK):** a 25 items test of algebra, geometry, fractions, decimals, and exponents.
- **Mechanical Comprehension (MC):** a 25 items test of mechanical and physical principles and ability to visualize how illustrated objects work.
- **Electronics Information (EI)**: a 20 items test measuring knowledge about electronics, radio, and electrical principles.
- **Assembling Objects (AO)**: a 16 items spatial visualization test.

Table 2: Eigenvalues of samples drawn from the parent population of Air Force recruits

		samples					
eigenvalue	**database**	**1**	**2**	**3**	**4**	**5**	**6**
1	3.930	3.971	3.876	2.475	3.888	3.811	4.208
2	1.328	1.324	1.314	1.241	1.337	1.386	1.212
3	1.039	1.008	1.061	1.164	1.077	1.039	0.974
4	0.700	0.701	0.713	0.922	0.702	0.689	0.683
N=	48222	1000	1000	1000	966	954	1015

Table 3: Distributional properties of samples generated from the parent population based on a latent variable "g"

sample	N	m	s.d.	skewness	kurtosis
random	1000	41.03	3.70	0.04	-0.68
g-hi	600	44.76	1.73	0.12	-0.11
g-av	600	41.19	1.73	0.12	-0.11
g-lo	600	37.61	1.73	0.12	-0.11

Table 4: Sequence of MCFA model fits and goodness of fit indices

model	df	par	X2	X2 diff.	sign.	CMIN/DF	RMSEA	PCLOSE	NFI	CFI
1. Full MI model	109	53	3106.26	-		28.498	0.124	0.000	0.959	0.961
2. RESVAR group-invariant model	106	56	709.27	2396.99	p < 0.001	6.691	0.056	0.004	0.991	0.992
3. RESVAR free model	82	80	254.29	454.97	p < 0.001	3.101	0.034	1.000	0.997	0.998
4. Model 3 & some intercepts free	74	88	209.08	45.22	p < 0.001	2.825	0.032	1.000	0.997	0.998

Table 5: Non-standardized MCFA factor loadings

	G				Verbal				Quantitative				Technical Knowledge			
	parent	mcfa			parent	mcfa				mcfa				mcfa		
	estimate	estimate	s.e.	sign.	estimate	estimate	s.e.	sign.	estimate	estimate	s.e.	sign.	estimate	estimate	s.e.	sign.
WK	**0.543**	0.501	0.115	0.00	**1.000**	1.000	0.000	0.00								
PC	**0675**	0.484	0.044	0.00	**0.461**	0.484	0.044	0.00								
MK	**1.039**	-0.317	0.162	0.05					**1.000**	1.000	0.000	0.00				
AR	**1.338**	0.453	0.110	0.00					**0.243**	0.390	0.063	0.00				
GS	**1.093**	0.339	0.145	0.02	**0.673**	0.839	0.073	0.00					**0.335**	0.328	0.050	0.00
MC	**1.346**	0.941	0.261	0.00									**1.000**	1.000	0.000	0.00
AS	**0.715**	0.568	0.249	0.02									**1.497**	1.800	0.115	0.00
EI	**0.960**	0.113	0.175	0.52	**0.257**	0.420	0.061	0.00					**0.962**	1.059	0.087	0.00
AO	**1.000**	1.000	0.000	0.00	**-0.207**	-0.259	0.092	0.00								

Table 6. Standardized MCFA factor loadings.

	G				Verbal				Quantitative				Technical Knowledge				Communality			
	aselect	high	average	low	aselect	high	average	low	aselect	high	average	low	aselect	high	average	low	aselect	high	average	low
g																				
verbal	**0.000**	0.000	0.000	0.000																
quant.	**0.000**	0.000	0.000	0.000	**-0.251**	-0.003	-0.103	-0.571												
TK	**0.000**	0.000	0.000	0.000	**0.180**	0.158	0.157	0.210	**-0.054**	-0.430	-0.505	-0.573								
WK	**0.468**	0.306	0.240	0.254	**0.798**	0.841	0.808	0.817									**0.856**	0.802	0.710	0.732
PC	**0.482**	0.269	0.199	0.212	**0.304**	0.370	0.335	0.342									**0.325**	0.393	0.433	0.435
MK	**0.656**	-0.158	-0.117	-0.137					**0.618**	0.934	0.832	0.603					**0.772**	0.897	0.706	0.382
AR	**0.766**	0.242	0.187	0.224					**0.129**	0.389	0.361	0.269					**0.603**	0.210	0.165	0.122
GS	**0.633**	0.154	0.126	0.133	**0.361**	0.523	0.526	0.529					**0.179**	0.185	0.220	0.200	**0.586**	0.373	0.378	0.382
MC	**0.673**	0.382	0.301	0.325									**0.463**	0.506	0.577	0.539	**0.667**	0.402	0.423	0.396
AS	**0.379**	0.197	0.146	0.172									**0.733**	0.779	0.835	0.850	**0.681**	0.646	0.718	0.751
EI	**0.528**	0.046	0.035	0.039	**0.131**	0.232	0.217	0.232					**0.490**	0.532	0.587	0.569	**0.559**	0.393	0.433	0.435
AO	**0.531**	0.431	0.333	0.324	**-0.102**	-0.154	-0.145	-0.135									**0.292**	0.210	0.132	0.123

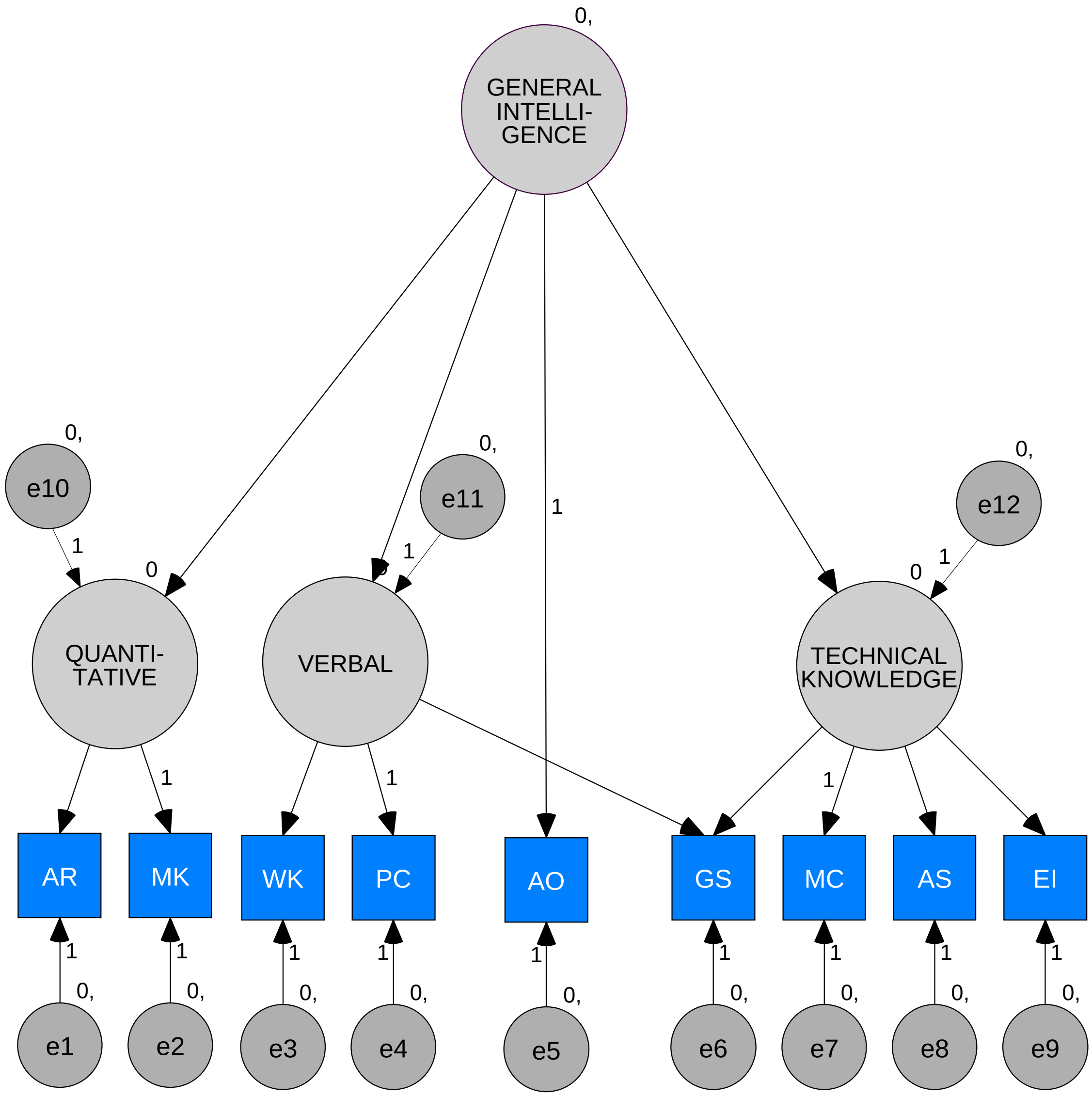

Figure 1: Model 1: A hierarchical model of "g"

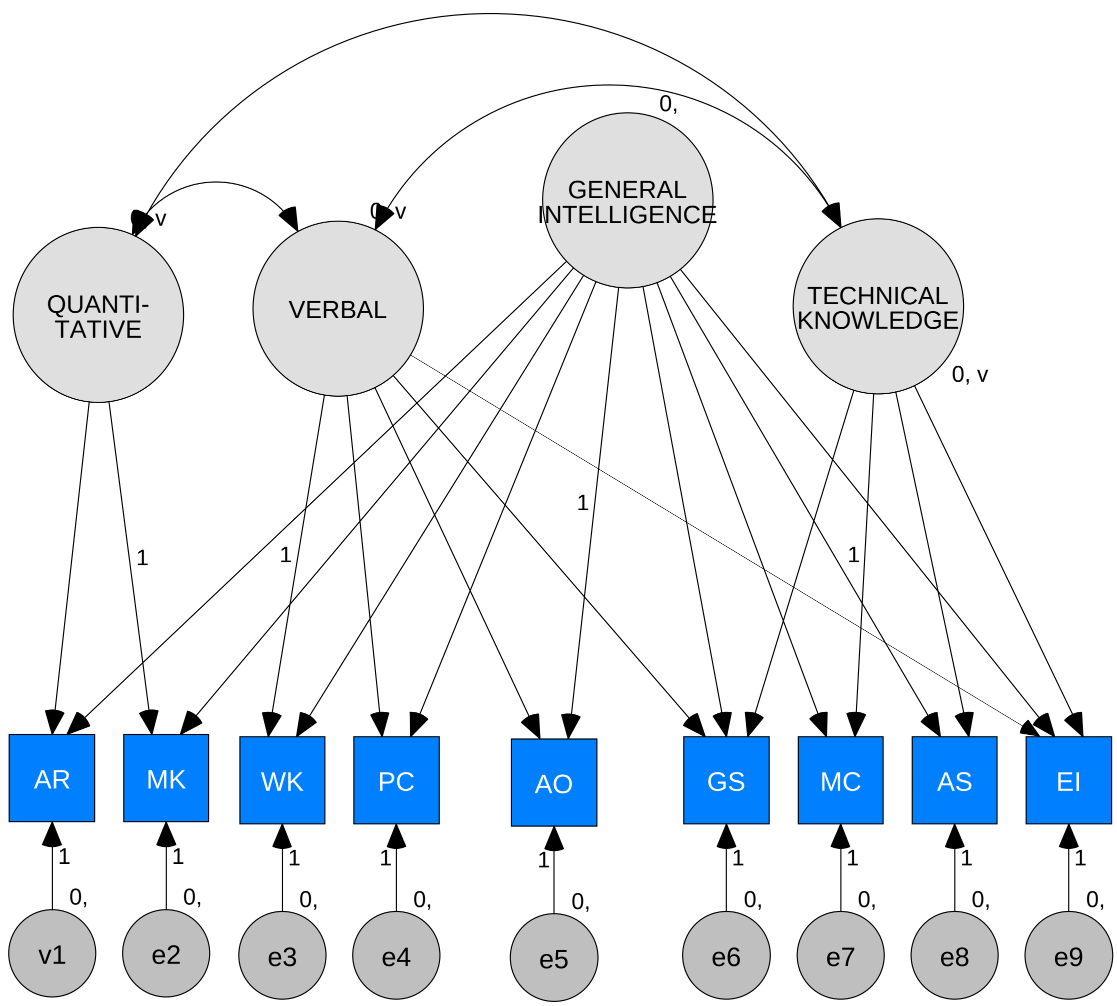

Figure 2: Model 2: A "G as first principal" model

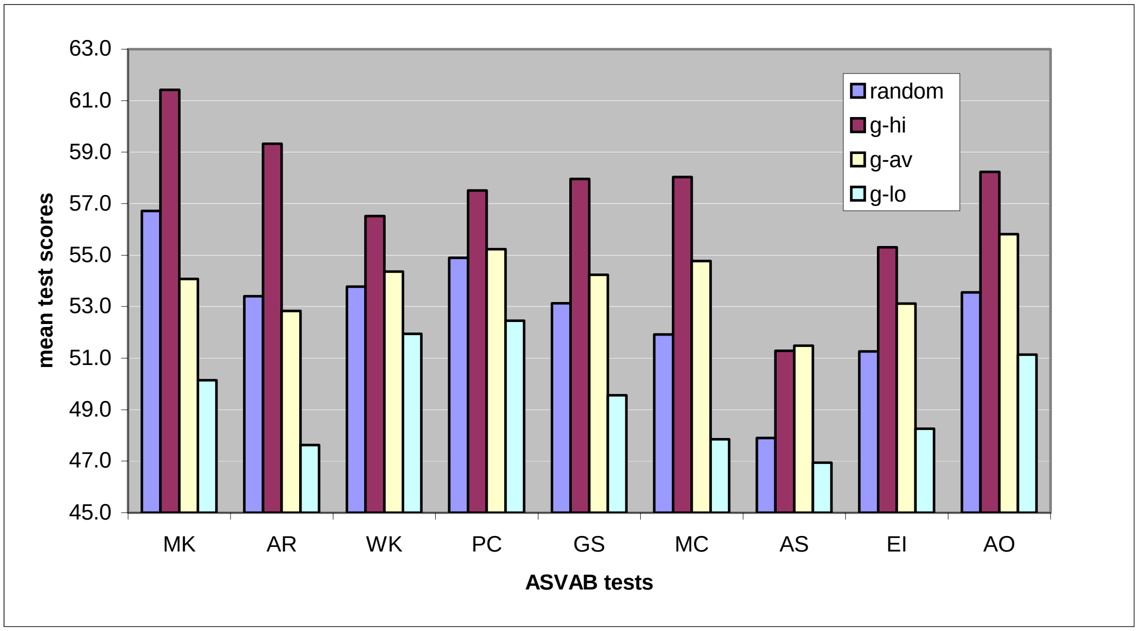

Figure 3: ASVAB subtests mean scores in samples with different levels of "g"

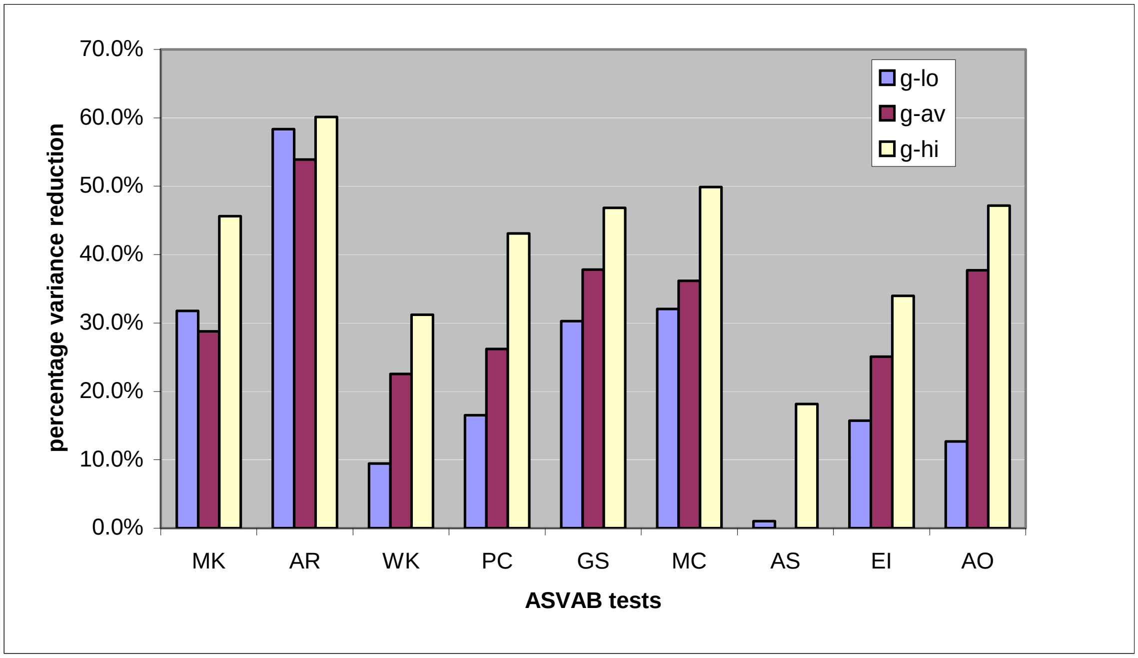

Figure 4: The effects of selection based on the latent variable"g" on the variance of ASVAB test scores