

Differential Prediction of Immigrant Versus Majority Group Training Performance Using Cognitive Ability and Personality Measures

Jan te Nijenhuis* and Henk van der Flier

There are very few studies of the differential prediction of work and training performance for immigrant groups in the Netherlands. Scores on the General Aptitude Test Battery, an intelligence test, and the Amsterdamse Biografische Vragenlijst (ABV), a frequently used Dutch personality test, and training data for 78 immigrant and 78 Dutch trainee truck drivers were collected. Lautenschlager and Mendoza's (1986) method of step-down hierarchical regression analysis was used to check for bias in the prediction of training results. Some predictor-criterion combinations showed differential prediction, especially with less cognitive and less objective criteria. The results with respect to the intelligence test appear to be in line with the findings of the US literature.

Differential prediction implies that tests show systematic errors in prediction that are related to group membership. Hunter, Schmidt and Hunter (1979), Jensen (1980), and Hunter (1983) review a large number of studies into differential prediction for mostly groups of Whites and Blacks in the USA, where standardized intelligence tests are used. The general conclusion is that in most studies no differential prediction is found. When differential prediction is found, it usually concerns differences in intercept, in which case the regression line of the majority group is almost always above the regression line of the minority group. In other words, in most cases it turns out that the intelligence tests are unbiased predictors and when bias is found, this turns out to be advantageous to the minority group. There are very few US studies on differential prediction using personality tests for selection purposes.

The US studies on differential prediction typically concern differences in test and criterion performance between native-born English-speaking minorities and Whites, which means that their conclusions cannot be extrapolated indiscriminately either to populations in developing countries or to immigrant populations in Western countries. In the last few decades, a growing number of immigrants have become part of the Dutch population. Most of these immigrants work in jobs at the lower end of the labour market, and a relatively large percentage are unemployed. Their mean test

results are lower than the mean test results of the Dutch. The use of tests to assess immigrants is being criticized in both the popular and scientific press in the Netherlands. These critics assume that tests are of limited use for assessing persons with a limited knowledge of the Dutch language and culture. A review of Dutch studies (te Nijenhuis and van der Flier 1999) shows only little test bias against immigrants; however, Dutch studies on differential prediction are rare.

The central question addressed in this study is whether there is differential prediction of the training results of truck drivers. Our research has relevance for the current debate on the assessment of non-native-born, non-native-language-speaking minorities in Europe. It is also relevant for the assessment of immigrants in the United States.

Method

Research Participants

In this study data on 78 immigrants and 78 majority group members who received training for a job as a truck driver in road transport and haulage for the Gemeenschappelijke Personeelsdienst Wegvervoer (GPdW) (Organization for Personnel Service in Road Transport) in Gouda were used. The immigrant sample constituted the complete population of immigrant trainees in 1994 and 1995; practically all were first-generation immigrants. The application process

*Address for correspondence:
Jan te Nijenhuis, Department
of Work and Organizational
Psychology, University of
Amsterdam, the Netherlands.
e-mail: AO_TeNijenhuis@mac
mail.psy.uva.nl

Table 1: Distribution of the immigrant group members and the majority group members with respect to native country; size of sub-group, age, and percentage of males

Group	N	Mean age	% males
Majority	78	27.35	96.2
Turks	14	27.21	100
North Africans	12	26.75	100
Surinamese + Antillians	8	29.14	100
Former Yugoslavs	27	31.41	100
Other	17	29.78	94.1
All immigrants	78	28.83	98.7

Note: The group of North Africans consist of persons mainly from Morocco, and also some from Algeria and Tunisia.

included a psychological examination which took place at the Work Conditions Service Unit of the Dutch Railways in ten centres throughout the Netherlands. A sub-sample was selected from the majority group in such a way that the distribution with respect to the regions in this sub-sample and the date of testing were as close as possible to that in the immigrant group. Table 1 shows the distributions of the groups in terms of demographic variables. The immigrant group from countries classified as Other consists of persons from: Asia (with the exception of Turkey, Israel, Japan, and the countries of the former Soviet Union); Africa (with the exception of South Africa and the North African countries; South America (with the exception of Surinam); Central America (with the exception of the Netherlands Antilles) and Romania. It can be noticed that the immigrant group is very culturally and linguistically heterogeneous. This, however, reflects the present situation with respect to immigrants within the Netherlands and many West European countries.

Tests and Assessments

Two tests were employed in this study: the Dutch version of the GATB 1002 (General Aptitude Test Battery) and the ABV (Amsterdam Biographical Questionnaire). The GATB is a test of general intelligence. All of its eight sub-tests are timed, and all but one are in multiple choice format. The sub-tests Computation, Three-Dimensional Space, Tool Matching, Form Matching, and Mark Making of the Dutch version are literal translations of the US version; the original stimuli were retained. The other sub-tests contain minor adaptations to the Dutch language and culture such as the introduction of Dutch and other European names in Name Comparison and Dutch money in Arithmetic Reason. As in the original GATB, the Dutch GATB contains three psychomotor sub-tests; of these three sub-tests, only Mark Making is used.

A review of test research in the Netherlands by Evers, van Vliet-Mulder and ter Laak (1992) showed that the Dutch version of the test has good predictive validity, content validity, and construct validity. Scores on the sub-tests of the majority group are therefore generally considered to provide a good indication of their capacities. Jensen (1985) showed that the *g* factor of the GATB is highly, but not perfectly, similar to the General Ability factor reflected in the total score of tests, which are seen as representing *g* well on the basis of their broad and well-balanced sampling of abilities. Te Nijenhuis and van der Flier (1997) investigated the comparability of GATB scores of immigrants and majority group members in the Netherlands and concluded that group differences with respect to construct validity were small and that the test can be put to good use for comparisons within culturally homogeneous groups of non-native-born, non-native-language minorities.

In the part of this study on differential prediction only the *g* scores will be taken into consideration. To obtain a measure of general intelligence the *g* score was computed using the regression method with the scores on the eight GATB sub-tests and weights from the Dutch GATB manual (see also te Nijenhuis and van der Flier 1997).

The ABV, one of the two personality questionnaires used most frequently in the Netherlands, is mainly based on the MPI (Maudsley Personality Inventory, Eysenck 1959); it consists of the following four scales: Neuroticism (N) as manifested by the presence of psycho-neurotic complaints (30 items), Neurosomaticism (NS) as manifested by the presence of functional somatic complaints (17 items), Extraversion (E) (21 items), and Social Conformity or Lie (L) (23 items). Contrary to the later Eysenck questionnaires, the E scales of the MPI and the ABV consist of the dimensions Sociability and Impulsivity. The review of test research in the Netherlands by Evers *et al.* (1992) showed that the ABV is one of the best personality tests in the

Netherlands: it has good predictive validity, content validity, and construct validity. Scale scores of the majority group members are therefore generally considered to provide a good indication of their personality. Te Nijenhuis, van der Flier and van Leeuwen (1997) investigated the dimensional comparability of ABV scores for immigrants and majority group members and concluded that the scales appear to measure the same constructs in the various groups and can be used for assessment within these groups. In this study on differential prediction only the scales for Neuroticism (ABVN) and Extraversion (ABVE) will be taken into consideration.

Criteria

The truck driver training course at the GPdW consists of different training modules, dealing with topics like traffic, technique, transportation administration, driving a truck and trailer, vehicle maintenance, dealing with various cargoes, professional social behaviour, and the transport of dangerous substances. Each of these modules ends with a differentiated assessment, which leads to a large number of criteria. A number of these criteria were clustered; all results of formal and objective examinations (based on instructor ratings) were clustered to Examination Marks. All variables from the module on professional behaviour (based on instructor ratings) were clustered to Professional Attitude. The cluster Practice Skills is a combination of parts of the modules on practical technique and practical treatment of goods (both instructor ratings and paper and pencil tests). The cluster scores were computed by summing the scores on the individual criteria and dividing them by the available observations. The individual criteria within a cluster turned out to correlate more strongly with each other, than the individual clusters with each other. The clusters inter-correlated positively.

Statistical Analyses

Background Variables. At the start of the training period some general data are available, concerning the period of unemployment of the candidate before the start of the training course, the number of certificates relevant for the various course modules which the candidate had already obtained, and the candidates' educational level. Using a Chi-square test, it was checked whether the immigrant and majority groups were comparable on these variables.

Means. The mean scores of the majority group members and immigrants on the predictor and the criteria are reported. The standardized score

difference was computed by dividing the mean difference between the two groups by the SD of the majority group.

Differential Prediction. To check for bias in the prediction of training results, use was made of Lautenschlager and Mendoza's method of step-down hierarchical regression analysis (1986). This method tests the hypothesis that one regression line can be used for both groups. If the hypothesis is rejected, this means that the prediction is significantly improved by using a separate regression line for each group. When using this method, it is first checked whether the predictor gives a significant prediction. Lautenschlager and Mendoza (1986) state that if this is not the case it makes little sense to carry on with the procedure, because of the lack of statistical power. If the test yields a significant prediction, the second step is to check for overall bias, that is, a combination of both slope bias and intercept bias. If there is overall bias, in steps 3 and 4 it is checked whether this is significant slope bias, significant intercept bias, or both.

Data were only reported when either step 1 or step 2 or both step 1 and step 2 led to effects at a significance level of 5%. This is a deviation from the Lautenschlager and Mendoza model, which stops when no significant prediction is found in step 1. Because of this, some analyses have an explorative character.

Results

Background Variables

The educational level appears to be somewhat higher in the immigrant group: $\chi^2(2, N=130) = 6.27, p = .04$. However, the quality of education in Third World countries and in Western countries is not generally comparable, so not too much importance should be attached to this difference in educational level. There is little difference between the period of unemployment for the two groups; the relation between background and duration of unemployment was not significant: $\chi(4, N=136) = 3.59, p = .46$. The number of candidates who had already obtained one or more driving licenses before the start of the training course was about the same in the two groups. No significant relation between background and obtained diplomas was found: $\chi(4, N=156) = .78, p = .98$. The conclusion is that there are hardly any differences between the background variables for the majority group and the immigrants.

Predictors

Table 2 shows that the immigrants have a lower mean score on the GATB than the majority

Table 2: Mean scores and SDs on the GATB sub-tests and the *g* score by group, the standardized score difference between the groups (effect size), and the significance of the effect size

Sub-test	Majority		Immigrants		effect size	p
	M	SD	M	SD		
Three-Dimensional Space	21.19	6.40	17.25	6.53	-0.62	.00
Computation	19.66	4.58	18.10	4.59	-0.34	.04
Figure Matching	28.55	5.71	26.11	5.38	-0.43	.01
Tool Matching	28.53	5.64	27.11	5.76	-0.25	.14
Name Comparison	54.60	10.93	45.17	11.21	-0.86	.00
Mark Making	67.77	7.66	70.99	10.63	0.42	.03
Arithmetic Reason	12.38	3.49	9.51	2.78	-0.82	.00
Vocabulary	22.04	6.40	11.59	6.29	-1.63	.00
<i>g</i> score	.40	.83	-.42	.82	-0.99	.00

Table 3: Mean scores and SDs on the ABV by group, the standardized mean score difference between the groups (effect size), and the significance of the effect size

Scale	Majority		Immigrants		effect size	p
	M	SD	M	SD		
ABVE	67.30	10.52	60.15	10.38	-0.68	.00
ABVN	25.48	10.08	35.70	13.95	1.01	.00

group; the mean difference is even greater than 1.5 SD on the sub-test Vocabulary. The mean difference on the *g* score is almost one SD. Table 3 shows that immigrants have substantially lower mean scores on extraversion and substantially higher mean scores on neuroticism; this is a negative score pattern.

Criteria

Table 4 shows that the mean differences between the two groups on the training criteria is about one quarter of a SD. Although differences of this size are traditionally not considered to be large, a group of employees that also performs on average 0.25 SD better after the training course will result in considerably higher operating results.

As shown in Table 5, some of the correlations in the immigrant group are significantly different from those in the majority group. This concerns

the correlation of *g* with the assessment of Professional Attitude and the correlations of ABVN (Neuroticism) and ABVE (Extraversion) with Practice Skills. Also, the correlation between ABVN and ABVE is significantly different in the immigrant group. Furthermore, it can be concluded that the correlations between the criterion variables are rather low for both the majority group and the immigrant group.

To find out whether the training criteria were predicted differently for the immigrants and the majority group members a step-down hierarchical regression procedure (Lautenschlager and Mendoza 1986) was carried out. The first step in this procedure is testing whether the predictors significantly predict the criterion scores and the second step is testing whether the prediction can be improved by assuming different regression equations for the two groups. As to the first step, Table 6 shows that the GATB *g* factor score significantly predicts

Table 4: Mean scores and SDs on the standardized aggregated criteria by group, the standardized mean score difference (effect size), and the significance of the effect size

Variable	Majority		Immigrants		effect size	p
	M	SD	M	SD		
Examination Marks	6.91	.51	6.77	.68	-0.27	.10
Professional Attitude	6.23	1.11	5.79	1.15	-0.40	.09
Practice Skills	6.14	0.72	6.00	1.10	-0.19	.49

Table 5: Intercorrelation matrix with 95% confidence intervals for predictor variables and aggregate criteria for majority group (below diagonal) and immigrants (above diagonal)

	1	2	3	4	5	6
1. <i>g</i>	—	-.094 (-.330/.150)	-.078 (-.315/.165)	.222 (-.025/.440)	-.224* (-.525/.125)	.226 (-.110/.515)
2. ABVN	.001 (-.230/.230)	—	-.083* (-.310/.155)	-.135 (-.365/.110)	.196 (-.150/.500)	.251* (-.080/.530)
3. ABVE	.216 (-.015/.425)	-.466 (-.625/-.270)	—	-.016 (-.255/.225)	.239 (-.110/.535)	-.319* (-.585/.005)
4. Examination Marks	.224 (-.010/-.435)	-.054 (-.280/.175)	.180 (-.050/.390)	—	.291 (-.050/.575)	.199 (-.135/.490)
5. Professional Attitude	.245 (-.050/.505)	-.073 (-.360/.225)	.161 (-.140/.435)	.228 (-.070/.490)	—	.246 (-.100/.540)
6. Practice Skills	.009 (-.290/.305)	-.369 (-.595/-.085)	.255 (-.040/.510)	.103 (-.200/.390)	.177 (-.130/.455)	—

Note: * Correlation in the immigrant group significantly different from correlation in the majority group ($p < .05$).

Examination Marks and that the ABVE scale significantly predicts Professional Attitude. As to the second step, Table 6 shows that the regression of Professional Attitude on *g* and the regression of Practice Skills on ABVN and ABVE differ for the two groups.

In Table 7 it is checked whether the overall bias shown in Table 6 should be interpreted as

slope bias, intercept bias, or both. As can be seen in all three cases the conclusion is that differences in slope are responsible for the biasing effect. The regression equations for the separate groups are given in Table 8. Table 8 also shows that the GATB *g* score is positively related to Professional Attitude in the majority group but negatively related in the group of

Table 6: First two steps in hierarchical regression procedure; prediction and increase in explained variance if different regression equations are assumed for immigrants and majority group members

	R	R ²	F	p
Examination marks on				
<i>g</i>	.277	.077	11.18	.00
<i>g</i> , group, <i>g</i> x group	.281	.079	.14	.87
ABVN	.091	.008	1.16	.28
ABVN, group, ABVN x group	.125	.016	.51	.60
ABVE	.101	.010	1.44	.23
ABVE, group, ABVE x group	.179	.032	1.54	.22
Professional Attitude on				
<i>g</i>	.142	.020	1.59	.21
<i>g</i> , group, <i>g</i> x group	.324	.105	3.55	.03
ABVN	.044	.002	.15	.70
ABVN, group, ABVN x group	.261	.068	2.66	.08
ABVE	.256	.065	5.38	.02
ABVE, group, ABVE x group	.295	.087	.88	.42
Practice Skills on				
<i>g</i>	.149	.022	1.77	.19
<i>g</i> , group, <i>g</i> x group	.207	.043	.82	.44
ABVN	.022	.000	.04	.85
ABVN, group, ABVN x group	.316	.100	4.32	.02
ABVE	.012	.000	.01	.92
ABVE, group, ABVE x group	.315	.099	4.30	.02

Table 7: Third and fourth step in the hierarchical regression procedure; increase in explained variance if slope or intercept bias is assumed

	Slope bias			Intercept bias		
	ΔR^2	F	p	ΔR^2	F	p
Professional Attitude on <i>g</i>	.042	4.38	.04	.034	2.86	.09
Practice Skills on ABVN	.087	7.54	.01	.006	.50	.48
Practice Skills on ABVE	.085	7.34	.01	.027	2.35	.13

immigrants. However, neither of the two within sub-group relationships is significant. ABVN is negatively related to Practice Skills in the majority group but has a positive relationship with Practice Skills in the group of immigrants. Only the relationship in the majority group is significant. ABVE is positively related to Practice Skills in the majority group but negatively in the group of immigrants. Only the negative relationship in the minority group is significant.

Discussion

The central question addressed in this study is whether there is differential prediction of the training results of immigrants. This appeared to be the case for some of the predictor-criterion combinations. General intelligence as a predictor of a cognitive (objective) performance criterion (Examination Results) revealed comparable relationships for immigrants and majority group members. The same result was found for the ABV Extraversion scale with respect to Professional Attitude. Differences in slope were found for the GATB *g* score with respect to Professional Attitude and for the ABV Neuroticism scale and the ABV Extraversion scale with respect to the assessment of Practical Skills.

While the relationships in the majority group sample were more or less in line with the expectations, immigrants with higher GATB *g* scores received lower ratings with respect to Professional Attitude and immigrants scoring higher on the Neuroticism scale and lower on

the Extraversion scale received better assessments for Practical Skills. Since the dimensional comparability of both the GATB sub-tests and the ABV scales for immigrants and majority group members in the Netherlands has been shown in earlier research (te Nijenhuis and van der Flier 1997; te Nijenhuis *et al.* 1997), this result might point to a difference in the assessment of immigrants during the training (criterion bias) in the sense that precise (worried) and less extraverted immigrants are seen as more serious trainees/employees.

The correlations between the predictors and the criteria were not very high. However, meta-analyses show that in individual studies correlations lower than .30 are often found. The literature gives causes for these low correlations implying that the empirical correlations from individual studies should be regarded as an underestimate of the real relationships. Among these causes are the non-perfect reliabilities of the predictors, the often rather low reliabilities of the criterion measures and restriction of range in the group of selected candidates. Correction for these factors would probably at least double the value of the correlations. Apart from these, one reason that correlations were not significant may have been that the samples were rather small.

The general conclusion from this study is that immigrants have lower or less favourable scores on the predictors than the majority group members and that a score difference is also found for the scores on the training criteria. Examination Results are related to General Intelligence in the same way for both groups.

Table 8: Regression equations for majority group members and immigrants

Majority group	Immigrants
Prof. Attitude = 6.09 + .34 (<i>g</i>) ($R^2 = .061$, $F = 2.79$, $p = .10$)	Prof. Attitude = 5.63 - .36 (<i>g</i>) ($R^2 = .054$, $F = 1.84$, $p = .18$)
Pract. Skills = 6.86 - .03 (ABVN) ($R^2 = .136$, $F = 6.79$, $p = .01$)	Pract. Skills = 5.24 + .02 (ABVN) ($R^2 = .063$, $F = 2.35$, $p = .13$)
Pract. Skills = 5.05 + .02 (ABVE) ($R^2 = .065$, $F = 3.00$, $p = .09$)	Pract. Skills = -7.75 - .03 (ABVE) $R^2 = .102$, $F = 3.97$, $p = .05$)

Differential prediction (differences in slope) plays a role for some of the predictor-criterion combinations, especially with the less cognitive and less objective criteria. The results with respect to the g scores for immigrants appear to be in line with the findings in the US literature.

Acknowledgement

The Training and Recruitment Foundation of Dutch Railways provided financial support for this research.

References

- Evers, A., van Vliet-Mulder, J. C. and ter Laak, J. (1992) *Documentatie van tests en testresearch in Nederland* (Documentation of tests and test research in the Netherlands). Assen, The Netherlands, Van Gorcum.
- Eysenck, H.J. (1959) *Manual for the Maudsley Personality Inventory*. London, University of London Press.
- Hunter, J. E. (1983) *Fairness of the General Aptitude Test Battery: Ability differences and their Impact on Minority Hiring Rates* (USES Test Research Rep. No. 46). Washington, DC, US Employment Service, US Department of Labor.
- Hunter, J. E. and Hunter, R. F. (1984) Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, **96**, 72–98.
- Hunter, J.E., Schmidt, F.L. and Hunter, R. (1979) Differential validity of employment tests by race: a comprehensive review and analysis. *Psychological Bulletin*, **87**, 721–735.
- Jensen, A.R. (1980) *Bias in Mental Testing*. London, Methuen.
- Jensen, A.R. (1985) The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences*, **8** 193–219.
- Lautenschlager, G.J. and Mendoza, J.L. (1986) A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement*, **10**, 133–139.
- te Nijenhuis, J. and van der Flier, H. (1997) Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology*, **82**, 675–687.
- te Nijenhuis, J. and van der Flier, H. (1999) A review of bias research in the Netherlands and its consequences. *European Journal of Psychological Assessment*, **15**, 165–175.
- te Nijenhuis, J., van der Flier, H. and van Leeuwen, L. (1997) Comparability of personality test scores for immigrants and majority group members: Some Dutch findings. *Personality and Individual Differences*, **23**, 849–859.