

## VU Research Portal

### The use of a test for neuroticism, extraversion and rigidity in Dutch immigrant groups.

te Nijenhuis, J.; van der Flier, H.; Leeuwen, van

**published in**

Applied Psychology

2003

**DOI (link to publisher)**

[10.1111/1464-0597.00155](https://doi.org/10.1111/1464-0597.00155)

**document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

**citation for published version (APA)**

te Nijenhuis, J., van der Flier, H., & Leeuwen, V. (2003). The use of a test for neuroticism, extraversion and rigidity in Dutch immigrant groups. *Applied Psychology*, 52(4), 630-647. <https://doi.org/10.1111/1464-0597.00155>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## The Use of a Test for Neuroticism, Extraversion, and Rigidity for Dutch Immigrant Job-applicants

Jan te Nijenhuis\*

*Universiteit van Amsterdam, The Netherlands*

Henk van der Flier

*Vrije Universiteit, The Netherlands*

Liesbeth van Leeuwen

*Meurs Personnel Advice, Woerden, The Netherlands*

Peut-on extrapoler les dimensions Névrotisme, Extraversion et Rigidité aux groupes d'immigrés aux Pays-Bas ? Et est-il pertinent d'utiliser les scores de l'IRT dans l'évaluation des candidats à un emploi relevant de ces populations (L'IRT est le test de Rigidité de l'Institut de Psychologie clinique et industrielle de l'Université d'Utrecht aux Pays-Bas) ? On a exploité les données issues de tests d'autochtones ( $N = 796$ ) et d'immigrants de la première génération ( $N = 1,302$ ) postulant en vue de postes ouvriers dans les Chemins de Fer néerlandais. Il est apparu que les échelles mesuraient les mêmes concepts dans les différents groupes et que la majorité des items avaient la même signification. La typologie peut par suite être transposée aux populations migrantes et le questionnaire peut être valablement utilisé pour effectuer des comparaisons dans des groupes culturellement homogènes de migrants candidats à un emploi nés à l'étranger et dont le néerlandais n'est pas la langue maternelle. Des travaux complémentaires sont indispensables pour s'assurer qu'un même positionnement sur les échelles de personnalité pour des individus issus de différents groupes détermine bien un même comportement professionnel.

The questions addressed are whether a taxonomy with the basic dimensions Neuroticism and Extraversion and the dimension Rigidity can be generalised to immigrant groups in the Netherlands, and whether scores on the IRT (ICIP [Institute for Clinical and Industrial Psychology, University of Utrecht, the Netherlands] Rigidity Test) can be used for assessment of job applicants within these groups. Use was made of test data on first-generation immigrants

---

\* Address for correspondence: Jan te Nijenhuis, Leiden University, Social and Organizational Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands. Email: JanteNijenhuis@planet.nl  
The Training and Recruitment Foundation of Dutch Railways provided the main part of the financial support for this research.

( $N = 1,302$ ) and majority group members ( $N = 796$ ) applying for blue-collar jobs at Dutch Railways. The scales appear to measure the same constructs in the various groups and the majority of the items do not show differential item functioning. Therefore the taxonomy can be generalised to immigrant populations and the questionnaire can be put to good use for comparisons within culturally homogeneous groups of non-native born, non-native language immigrant job applicants. More research is needed to determine whether the same position on the personality scales for persons from different groups leads to the same job-related social behavior.

## INTRODUCTION

Every human is essentially unique, but this uniqueness can only become manifest when it is expressed in generally recognisable characteristics (Feij, 1985). A class of these characteristics is formed by personality traits and one of the goals of personality theories is the development of a taxonomy of these traits. If personality traits are arbitrarily shaped by culture, then very different traits and trait structures might be found in different cultures (McCrae & Costa, 1997). An important test of such a taxonomy or structure is therefore its generalisability to other groups. Many empirical studies of group differences, using etic (using foreign measures), emic (using indigenous measures), or lexical approaches (using all terms to describe personality traits in the lexicon of a specific language) have shown the applicability of the two basic personality factors Extraversion (E) and Neuroticism (N) (Barrett, Petrides, Eysenck, & Eysenck, 1998; Benet-Martínez & John, 1998; Cheung & Leung, 1998; Church, Katigbak, & Reyes, 1998; De Raad, Perugini, Hrebícková, & Szarota, 1998; Hahn, Lee, & Ashton, 1999; McCrae & Costa, 1997; McCrae, Costa, Del Pilar, Rolland, & Parker, 1998; Paunonen & Ashton, 1998; Somer & Goldberg, 1999). Although groups from various cultures have been studied, research on the four main immigrant groups in the Netherlands (Turks, Moroccans, Surinamese, and Netherlands Antillians) is rare. Therefore, the first question addressed in this paper is whether a taxonomy with the basic dimensions Neuroticism and Extraversion and with the dimension Rigidity can be generalised to four immigrant groups in the Netherlands.

## IMMIGRANTS ON THE DUTCH LABOR MARKET

In recent decades growing numbers of immigrants have become part of the Dutch population, now making up about 6 per cent of it. The status of most immigrants in society is low: most immigrants work in jobs at the lower end of the labor market and a relatively large percentage is unemployed. However, some of the immigrant groups, such as Chinese, are more successful than others. This paper concerns only those groups that are in a disadvantageous position, such as immigrants from Surinam, the Netherlands Antilles,

Morocco, and Turkey. Since 1979 the official Dutch policy on immigrants has been to integrate them into society, to alleviate their social and economic problems, and to stop or prevent discrimination against them. The integration of immigrants into a new society takes place predominantly through participation in the labor market. It is therefore alarming that in 1998 unemployment among immigrant groups was about 16 per cent, whereas in the Dutch population it was 6 per cent (Central Bureau for Statistics, 1999).

Dutch policy makers have different opinions about the most effective approach to the problem of immigrant unemployment. The Ministry of Social Affairs and Employment favors a policy that will stimulate employers to employ immigrants. The employers' organisations prefer initiatives from industry itself. The trade unions support contract compliance, meaning that companies only receive government subsidies or commissions if the percentage of immigrants in their workforce corresponds to the percentage of immigrants in their region. Finally, immigrant organisations think that a quota system is inevitable, which implies that a fixed minimum proportion of immigrants is guaranteed a job.

In 1998, the Dutch government passed a bill, the so-called Law for the Advancement of Work Participation of Immigrants, that compels companies with more than 35 employees to register their personnel's ethnicity. It also obliges companies to draw up a plan in which it is shown where in the organisation potential hindrances for the hiring and the career development of employees from immigrant groups can be found and how these hindrances will be removed. Companies must also strive for an equal representation of immigrants. Up to now, however, this law has not led to a clear improvement of the status of immigrants on the Dutch labor market.

The most frequently mentioned possible causes for the high percentage of unemployment in immigrant groups are: low education levels, the low level of knowledge of Dutch culture, the low level of command of the Dutch language, discrimination, and unfair treatment when trying to enter an organisation. Of these, immigrants' low level of education is generally seen as being the main reason for their high unemployment. Seventy per cent of the Turkish and Moroccan working population has had no more than primary education; for Surinamese and Antillians this figure is 30 per cent. The hypothesised unfair treatment immigrants receive when trying to enter an organisation may be influenced by the use of non-optimal selection models and instruments.

## EVALUATION OF CROSS-CULTURAL VALIDITY OF TESTS

A lasting multicultural society requires instruments for assessment that are valid for all different groups. It is therefore imperative to continue to evaluate the cross-cultural validity of standard tests (Dana, 1998). A conservative

strategy is to use tests for assessment only if their validity has been convincingly demonstrated in empirical research. Test validation for minorities in the United States focuses on American Indians (Allen, 1998), African Americans (Lindsey, 1998), Asian Americans (Okazaki, 1998), and Hispanic Americans (Cuéllar, 1998). Test validation for minorities in the Netherlands focuses on the aforementioned immigrant groups. The use of personality tests to assess immigrants is criticised in both the popular and the scientific press in the Netherlands. These critics assume that these tests are of limited use for assessing persons with a limited knowledge of the Dutch language and culture.

The assessment of the use of a test for comparisons between groups focuses on two different questions. The first is whether in the two groups the same scale score reflects the same position on a latent trait; the second question is whether members of different groups with the same scale score have the same probability of showing specific criterion behavior in the future. These two questions pertain to the traditional distinction between construct validity and predictive validity. We will deal here only with the first question.

One way to assess construct validity is to look at the occurrence of quantitative DIF (differential item functioning). An item is said to show quantitative DIF when a group, given a position on a latent trait, obtains a systematically higher or lower score on that item than another group. The occurrence of some quantitative DIF does not always have to be disturbing, because in applied and research settings one is working with scale scores that can be obtained by different combinations of item scores; a higher score on one item can be compensated by a lower score on another item.

Another question addressed in this paper is whether the IRT ICIP Rigidity test can be used for assessment within the four largest immigrants groups in the Netherlands. The IRT measures the classical dimensions Neuroticism and Extraversion (Eysenck, 1959), besides the less common dimension of Rigidity. The first two traits are represented in most personality inventories. The criterion-related validity of these two dimensions is only modest, but because there is no correlation with cognitive abilities, the combination of personality tests and cognitive tests results in substantial incremental criterion-related validity (see Schmidt & Hunter, 1998; Tett, Jackson, & Rothstein, 1991).

## METHOD

### Research Participants

This project used test data on first-generation immigrants and majority group members who applied for blue-collar jobs at the Dutch Railways and regional bus companies in the Netherlands from 1988 until 1992. Majority group is defined as native-born Dutch, excluding native-born second- and

third-generation immigrants. The application process included a psychological examination which took place at the Work Conditions Service Unit of the Dutch Railways in 10 centers throughout the Netherlands. The immigrant sample constituted the complete population of first-generation immigrant job applicants.

A sample was selected from the majority group applicants in such a way that the distribution with respect to the jobs and regions in this sample was as close to that in the immigrant group as possible. The majority group consisted of  $N = 796$  research participants (87.7% males, mean age 28.4). The immigrant group consisted of persons from Turkey ( $n = 271$ , 96.4% males, mean age 24.0), North Africa (persons from Morocco, Algeria, Tunisia, Libya, and Egypt) ( $n = 165$ , 97.4% males, mean age 27.4), the Netherlands Antilles ( $n = 126$ , 83.5% males, mean age 31.2), and Surinam ( $n = 524$ , 82.5% males, mean age 30.3). The immigrant group from countries classified as *other* consists of persons from countries other than those mentioned above ( $n = 216$ , 87.7% males, mean age 28.4). In view of the overly large heterogeneity of this group, its test scores are only reported in analyses of all the immigrants, treated as a single group. The mean number of years that the immigrants had been residing in the Netherlands at the time of their application was 11.2 years ( $SD = 6.9$  years).

## Employment Selection Tests

*IRT.* The IRT (ICIP [Institute for Clinical and Industrial Psychology, University of Utrecht, the Netherlands] Rigidity Test) (Tellegen, 1968) measures Rigidity, Extraversion, and Neuroticism. It consists of the following seven scales: Emotional Perseveration (EP) as manifested by difficulties in dismissing emotional experiences from one's mind (11 items), Dogmatism (D) as manifested by holding on to ideals, intentions, and principles (6 items), Achievement Orientation (AO) or achieving in work situations (6 items), Order (O) or striving for order, precision, and punctuality (11 items), Social Adaptation (SA) or adaptive orientation towards the outside world (7 items), Variation Need (VN) or exploratory orientation towards the outside world (7 items), and Test Attitude (TA) or self-defensive versus self-critical test attitude (18 items). EP measures predominantly neuroticism and to a smaller degree introversion. AO, SA, and VN measure extraversion. D measures predominantly rigidity and to a smaller degree neuroticism. O measures rigidity.

The TA scale (also known as the L or Lie scale) had originally been included in questionnaires to control for social desirability as an answering tendency during the completion of the questionnaire, but research into the construct validity of the scale showed that it has the character of a stable personality trait rather than that of a situation-dependent answering tendency which disturbs the validity of the other scales. This personality factor

may be described as *Social Conformity*. Consistent, stable individual differences on the L scale have been found that correlate with other measures, such as adjustment (Eysenck & Eysenck, 1975; McCrae & Costa, 1983). The mean scores on the L scale of the EPQ in West European countries are lower than those in Southern European and Third World countries (Barrett et al., 1998). Nyborg, Eysenck, and Kroll (1982) suppose that in societies where many kinds of behavior are acceptable, people will obtain lower L scores and that in restrictive societies people will obtain higher L scores. The IRT consists of 66 items that can be answered on a 5-point scale.

Tellegen (1968) tried to construct a third basic dimension, Rigidity, beside the basic dimensions of Neuroticism and Extraversion. He described Rigidity as "a restriction in behavioral repertoire" and within this dimension he wanted to make a more refined classification into "rigids" and "dogmatics". Dogmatics were supposed to react more defensively to their environment, whereas rigids were supposed to react more preventively by trying to stabilise their environment. Tellegen saw the Order scale as the operationalisation of the Rigidity construct. However, this theoretical distinction is not supported by the empirical research.

Scales making up the dimension of Rigidity are used in many popular Dutch personality questionnaires and in American questionnaires such as the California Psychological Inventory (Gough & Bradley, 1996) and the Minnesota Multiphasic Personality Inventory (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). However, the status of the dimension Rigidity in taxonomies is not entirely clear. According to Eysenck and Eysenck (1962) Rigidity, or Dogmatism, is a dysthymic trait, that is, a combination of Neuroticism and Introversion. However, many empirical studies show that Neuroticism is the dimension of personality related to individual differences in Dogmatism, whereas Extraversion is irrelevant to Dogmatism (Francis, 1998). Several factor analyses of Dutch personality questionnaires, not discussed by Francis, but reviewed by Te Nijenhuis (1997), indicate that a separate dimension of Rigidity is best called for. However, the patterns of secondary loadings reported by Francis are also found in these studies. So, all in all, there appears to be support for a separate factor of Rigidity, albeit that some of the scales making up this factor have secondary loadings on the factor Neuroticism.

The influential review of test research in the Netherlands by Evers, Van Vliet-Mulder, and Ter Laak (1992) showed that although some of the IRT scales have a rather low reliability, the validity of most scales is sufficient for practical use within the Dutch majority group. It is, however, not known whether this test can be used for the assessment of immigrant job applicants.

*ABV*. The Amsterdam Biographical Questionnaire (*Amsterdamse Biografische Vragenlijst*, ABV) (Wilde, 1970), 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). The L scale strongly resembles the IRT's TA scale.

A 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. Research on the validity of the ABV for the assessment of immigrants is reported in Te Nijenhuis, Van der Flier, and Van Leeuwen (1997) suggesting that the ABV can be used quite well for the assessment of immigrants. The ABV is generally seen as an excellent measure of the two basic dimensions Neuroticism and Extraversion. It is added to the data on the IRT to aid the interpretation of the outcomes of factor analyses.

## STATISTICAL ANALYSES AND RESULTS

### Means and Reliabilities

The deviation of the mean scale scores of the immigrants from the mean scale scores of the majority group members was calculated in terms of the standard deviation of the majority group. Mean scale scores of the majority group and deviations of the mean scale scores of immigrant groups from these mean scale scores are presented in Table 1. The immigrants showed higher mean scores on the scales corresponding with the Neuroticism and Rigidity dimension and lower mean scores on the scales corresponding with the Extraversion dimension. The differences were larger for the Turks and North Africans than for the Surinamese and Antillians.

Cronbach's alpha was used to assess the reliability. Most of the values of alpha are comparable for immigrants and the majority group; however, a quarter of the alpha values are clearly lower in the immigrant group. The reliability of most IRT scales is moderate (.60–.79) to rather low (.40–.59) in the majority group, especially for D and SA subscales; this calls into question the value of these subscales for assessment. However, most scales are quite short.

### Dimensional Comparability

*PCA and Pecon.* According to Barrett (1986) there is no single best way to compare factor solutions in different groups. He suggests using different procedures and measures in the same study so as to reduce the influence of



TABLE 1  
Means and Standard Deviations of the Scales of the IRT for the Majority Group;  
Deviations from the Mean of the Majority Group Expressed in Standard  
Deviations of the Majority Group, and Standard Deviations  
on the Scales of the IRT by Immigrant Group

Scale	Groups											
	Majority		Turks		North Africans		Antillians		Surinamese		All Immigrants	
	M	SD	Dev.	SD	Dev.	SD	Dev.	SD	Dev.	SD	Dev.	SD
EP	22.99	5.09	-.47	5.82	-.41	6.22	-.04	5.68	-.21	6.17	-.28	6.03
D	15.06	3.54	-1.03	3.62	-1.15	3.80	-.88	32.86	-.73	4.05	-.84	3.99
AO	24.07	2.71	.62	3.32	.62	3.11	-.08	3.01	.02	3.07	.22	3.18
O	50.14	6.49	-.54	6.28	-.54	5.83	-.54	6.98	-.57	6.63	-.52	6.45
SA	27.09	3.24	.53	3.42	.47	3.66	.28	3.43	.32	3.43	.37	3.50
VN	25.22	3.67	.39	3.43	.07	3.58	-.04	3.47	.04	3.75	.11	3.67
TA	57.30	8.88	-.60	8.09	-.54	7.86	-.36	8.82	-.53	8.43	-.48	8.32

Note: The minimum number of subjects was 794 for the majority group, 249 for the Turks, 144 for the North Africans, 122 for the Antillians, 513 for the Surinamese, and 1,228 for all immigrants.

EP = Emotional Perseveration, D = Dogmatism, AO = Achievement Orientation, O = Order, SA = Social Adaptation, VN = Variation Need, and TA = Test Attitude. Dev. = Deviation from the mean of the majority group.

the weaknesses of each. The dimensional comparability was therefore examined by means of two combinations of a procedure with a measure. Principal Component Analysis (PCA) solutions with varimax rotation of the majority group and the various immigrant groups were compared with the congruence coefficient (Burt, 1948; Tucker, 1951). A value of the congruence coefficient greater than .85 is generally considered to be high. The Pecon procedure (Ten Berge, 1986) was developed for component comparisons across populations, using rotation to perfect congruence. It is assumed that for any given component in a first population a parallel component in any second population can be defined which has, up to a constant of proportionality, the same component scores coefficients. When using the Pecon procedure, one begins by deriving a theoretically sound factor structure for one group. Then the weight matrix of this first group is applied to the correlation matrix of the second group. In this way, the second group's factor structure has by definition the same meaning as the original group's factor structure. The question is then whether the same factors explain as much variance in the second group as in the first group. In this study, the weight matrix of the PCA with varimax rotation of the majority group was used as a starting point for Pecon.

*Combination of Data.* To reduce the number of comparisons in the main analyses and to diminish the risk of accidental deviations or deviations of little practical significance, it was checked whether the data from the Turks and the North Africans on the one hand, and the data from the Surinamese and the Antillians on the other, could be combined. These combinations seemed obvious, considering the similarity between the Surinamese and Antillians with respect to their proficiency in Dutch. Through education and exposure to the media in their native countries, the Surinamese and Antillians came into contact with the Dutch language and culture, which is not true of the Turks and North Africans. The hypothesis that the correlation matrices of two groups are equal was tested with the aid of LISREL. In addition to the chi-square values the values of the Goodness-of-Fit Index (GFI) (Jöreskog & Sörbom, 1988) were calculated. The GFI indicates to what extent the correlation matrix of one of two groups deviates from a common matrix that is estimated by the LISREL program. The correlation matrices of the scale scores for the Turks and North Africans and those for the Antillians and Surinamese did not differ significantly at the .05 level and the fit indices were very high; for the comparison of the Turks with the North Africans,  $\chi^2(21, N = 442) = 22.48, p = .373, GFI = .991$  for the Turks, and  $GFI = .972$  for the North Africans; for the comparison of the Antillians with the Surinamese,  $\chi^2(21, N = 661) = 22.89, p = .350, GFI = .954$  for the Antillians, and  $GFI = .997$  for the Surinamese. It can be concluded that the correlation matrices in the two comparisons are essentially identical. The a priori division of the groups in the Method section was consistent with the outcomes of the LISREL analyses.

Several of the IRT scales measure more than one dimension; other scales measure lower-order dimensions in the personality taxonomy. In contrast, the ABV clearly measures the two basic dimensions Extraversion and Neuroticism; factor analyses of the scales of the ABV result in a very clear, simple structure (Te Nijenhuis et al., 1997). Therefore, we decided to use the ABV scales as markers for the two basic dimensions of E and N. So, in order to create a theoretically sound factor structure and one that is easy to interpret the scales of the IRT and the ABV were factor analysed together, leaving out the Test Attitude scale of the IRT because of its overlap with the ABV's Lie scale.

*Factor Solutions.* For the majority group, the Turks and North Africans, and the Antillians and Surinamese a three-factor solution was chosen because the eigenvalues dropped off sharply, and because three factors could be expected on theoretical grounds. Table 2 shows that this is a clearly interpretable factor solution, notwithstanding the secondary loadings. The first factor is interpreted as Neuroticism; the scales of Neuroticism and Emotional Perseveration load on it. The second factor is interpreted as Extraversion; the scales of Extraversion, Social Adaptation, and Variation

TABLE 2  
 Varimax Rotated Factor Matrices for the Majority Group, the Turks and North Africans, and the Antillians and Surinamese for the Scales of the ABV and the IRT

Scales	Neuroticism			Extraversion			Rigidity		
	Maj.	T + N	A + S	Maj.	T + N	A + S	Maj.	T + N	A + S
ABV									
N	.83	.81	.84	-.14	.00	-.09	-.09	-.11	-.07
NS	.71	.70	.73	.00	.07	.03	-.06	-.10	-.02
E	-.26	-.42	-.25	.66	.62	.64	.02	.06	-.05
L	-.46	-.34	-.36	-.16	-.46	-.28	.58	.58	.63
IRT									
EP	.66	.81	.71	-.34	-.18	-.27	-.07	-.07	-.12
D	.44	.34	.40	-.06	-.20	-.07	.69	.76	.66
AO	-.25	-.53	-.31	.59	.52	.55	.33	.10	.40
O	-.23	-.34	-.16	.13	.20	.08	.79	.67	.84
SA	-.13	-.22	-.11	.74	.67	.71	-.04	.27	.13
VN	.19	-.09	.13	.64	.68	.67	-.12	-.01	-.18

Note: Maj. = majority group; T + N = Turks and North Africans; A + S = Antillians and Surinamese.

Need load on it. The third factor is interpreted as Rigidity; the scales Lie, Dogmatism, and Order load on it.

A comparison of factor solutions of immigrants and the majority group shows that all scales have their highest loading on the same factors. There are group differences only in the secondary loadings. The differences in loadings are in all cases smaller than 10 per cent of the scale variance for Antillians and Surinamese. The differences in loadings are in some cases larger than 10 per cent of the scale variance for the Turks and North Africans. The Emotional Perseveration and Achievement Orientation scales load higher on Neuroticism. The scale Social Adaptation loads lower on Extraversion. The scale Dogmatism loads higher on Rigidity and the scale Order loads lower on Rigidity.

The congruence coefficients for the comparison of the solutions after varimax rotation were high; for the comparison of Turks and North Africans, for Neuroticism  $\phi = .966$ , for Extraversion  $\phi = .940$ , for Rigidity  $\phi = .985$ ; for the comparison of Antillians and Surinamese, for Neuroticism  $\phi = .994$ , for Extraversion  $\phi = .992$ , for Rigidity  $\phi = .934$ .

The Pecon procedure was used for the comparison of the percentage of explained variance of the varimax rotated solution of the majority group and the solutions after rotation to perfect congruence of the immigrant groups. The weight matrix of the three-factor varimax solution of the majority group applied to the correlation matrix of the Turkish and North African group and the Antillian and Surinamese group yielded the following percentages of

explained variance: for the majority group 57.76 per cent; for the Turks and North Africans 59.52 per cent; and for the Antillians and Surinamese 58.13 per cent. The Rotation to Perfect Congruence showed that the factors in the three groups, that have by definition the same interpretation, explained virtually the same amount of variance in the immigrant groups. Taken as a whole, the deviations in the factor structures of the immigrant groups are small.

## Differential Item Functioning

*Qualitative and Quantitative DIF.* Research into differential item functioning (DIF) starts with a definition of what constitutes differentially functioning and nondifferentially functioning items. Then, on the basis of statistical procedures that are operationalisations of the definition of DIF, the question of which items show DIF can be addressed. The term DIF is used because it pertains only to statistical deviance; the items that show DIF deviate only in a statistical sense from the other items. Finally, based on the statistical results and other information, hypotheses are formulated about the qualities of the tested persons or the items (or both) that might be responsible for the statistical deviance.

An item is said to show *qualitative* DIF when its loading changes in a factor analysis. An item is said to show *quantitative* DIF when in different groups, persons with the same trait level (or scale score) obtain systematically higher or lower scores on the item, or, phrased differently, when, after correcting for differences in trait level between groups, the item score is related to group membership. Stricker's (1982) partial correlation index was chosen as a check for quantitative DIF. Unlike other DIF techniques, it is also applicable to nondichotomous items. The partial correlation index corrects the correlation between item score and subgroup for the scale score, minus the item score. In Stricker's formula  $r(iS.T_{\infty})$   $i$  is the item score,  $S$  is the subgroup, and  $T_{\infty}$  is the total score minus the item score, corrected for attenuation. Because the samples were large, the partial correlation index was already significant ( $p < .001$ , two-tailed) at a value around .10. In this study an item will be considered as showing quantitative DIF when the absolute value of the partial correlation index is larger than .20; a value below .20 means that less than 4 per cent of the variance in the item is explained by DIF, which can be considered a small effect.

*An Iterative Procedure.* A serious weakness of conditional DIF indices like the partial correlation index is that if the scale contains a relatively large number of DIF items the scale score may be biased. This can mean that some DIF items are not classified as such, and that some non-DIF items are erroneously classified as showing DIF. Therefore an iterative procedure was used for the computation of the partial correlation index (Van der Flier,

Mellenbergh, Adèr, & Wijn, 1984). Briefly this method involves removing DIF items from the scale (or, more accurately, not counting them in determining the scale score) so that the number of items removed increases by one with each successive iteration. After the first iteration, the item that showed the most DIF was not included for the computation of the scale score, and the partial correlation index was computed again. Next, the two items that showed the most DIF were identified and not included for the computation of the scale score, the partial correlation index was computed again, and so forth. The procedure was ended when all items with an absolute value of the partial correlation index lower than .20 had been excluded from the computation of the scale score, or when less than half of the items were left over. This was done to prevent there being too few items left to reliably estimate the position on the trait.

*Explaining Statistical Deviance.* After the items that showed DIF had been identified, the next task was to explain the statistical deviance. Post-hoc inspection of the DIF items was conducted, with the emphasis on identifying striking features that could be related to well-known differences between majority and immigrant groups. The effect of the biased items on the mean scores of the immigrants was estimated. Therefore, the effect of the removal of items that showed DIF on the scale scores was computed. This was done by first computing the deciles for the majority group. After this, the distribution of the scores of the complete immigrant group over the deciles was computed. The items that showed quantitative DIF for the complete immigrant group were removed and the procedure was repeated to find out whether the distributions became more alike.

*Outcomes.* A non-univocal factor solution is found at the level of the individual items. A three-factor solution yields two bipolar factors in the majority group. Table 3 shows a simplified report; the number of items of each scale that load on every factor is given. The factor solution of the Surinamese resembles that of the majority group but the resemblance of the Antillians with the majority group is less good. For the Turks and North Africans, the first factor still reasonably resembles that of the majority group, but the second and third factor do not show a strong resemblance.

The number of items that showed quantitative DIF was the largest for the Turks (12), followed by the North Africans (7), the Surinamese (7), and the Antillians (2). Four items showed quantitative DIF for the entire immigrant group, about 6 per cent of the 66 items. In this study, quantitative DIF is to be interpreted as agreeing or disagreeing to a greater extent than the majority group members having the same scale score.

Post-hoc inspection of items that showed quantitative DIF resulted in explanations in terms of the importance of social relationships, a traditional

TABLE 3  
 Varimax Rotated Factor Matrices at the Item Level by Group for the IRT, Number of Items in a Scale, and Number of Items of a Scale that Load on a Factor

Scales	Items	Groups														
		Majority			Turks			North Africans			Antillians			Surinamese		
		f1	f2	f3	f1	f2	f3	f1	f2	f3	f1	f2	f3	f1	f2	f3
EP	11	11	-	-	9	2	-	11	-	-	8	-	3	10	1	
D	6	1	5	-	1	5	-	1	2	3	1	5	-	-	5	1
AO	6	5	1	-	3	3	-	4	1	1	2	1	3	3	3	-
O	11	2	9	-	3	5	3	4	3	4	2	9	-	2	9	-
SA	7	5	-	2	1	5	1	3	1	3	-	3	4	1	5	1
VN	7	2	-	5	3	4	-	1	2	4	2	1	4	2	-	5
TA	18	2	10	6	6	6	6	3	11	2	6	11	3	5	9	4

Note: Items = number of items in a scale, f = Number of items that load on a factor.

way of living, a relatively unfavorable socioeconomic position, and a low level of knowledge of the Dutch language. An item on the Emotional Perseveration scale showing quantitative DIF was: "If I have had an argument, this usually puts me in a bad mood for a few days". Turks and North Africans responded *agree* relatively more often, which may be explained by the fact that Turks and North Africans interpret this as an argument with somebody within their own group and the fact that good contacts within this so-called "ingroup" are seen as important. An item on the Dogmatism scale showing quantitative DIF was: "People are predestined for a specific role in life". North Africans responded *agree* relatively more often, which may be explained by the fact that in a traditional society one does what one is expected to do, there being a strong emphasis on fatalism. An item on the Achievement Orientation scale showing quantitative DIF was: "Actually I'm quite indifferent to my work and I only do it because it needs to be done". Turks responded *agree* more often, which may be explained by the fact that they are probably less likely to have challenging jobs. An item on the Social Adaptation scale showing quantitative DIF was: "One should not always stay sober, but one should sometimes let oneself go". Immigrants responded *disagree* more often; it is possible that in the immigrant group the word "sober" is associated with the use of alcohol and "let oneself go" is associated with lack of self-control, both of which may be regarded negatively. It may also be the case that both these Dutch expressions were imperfectly understood. An item on the Variation Need scale showing quantitative DIF was: "I prefer to take my vacation each year at the same place". Turks and North Africans responded *agree* more often, which may

be explained by the fact that many Turks and Moroccans visit their family in their country of origin every year. It was not possible to find a post-hoc explanation for a number of other items that showed quantitative DIF; moreover, the expectation that certain items would show quantitative DIF was not borne out.

The effect of the quantitative DIF on the immigrants' mean scores was estimated. After removing the DIF items, the distribution of the scores of the immigrants came at most one decile closer to the distribution of the majority group for all scales. The quantitative DIF therefore influenced the decile into which immigrants were classified. Experienced selection psychologists at the Work Conditions Service Unit of the Dutch Railways, however, regard the influence of a one-decile difference on the chance of being selected for a position as being small.

## DISCUSSION

It appears from the results that a taxonomy including the basic dimensions Neuroticism and Extraversion and the dimension Rigidity can be generalised to immigrant populations in the Netherlands. The results suggest that the IRT has good cross-cultural validity and can be used for assessment within immigrant groups.

All scales of the IRT show mean differences between immigrants and the majority group. These differences are unfavorable to immigrants: they score higher on scales related to neuroticism and rigidity and lower on scales related to extraversion. This profile clearly deviates from the regular job requirements of emotional stability, sociability, and flexibility and implies a lower suitability for many positions.

The high TA scores of the immigrants are comparable to the high TA scores found in many non-Western countries. These scores may reflect the high level of social conformity and the lack of flexibility in the value systems of a more traditional society.

The reliability of most IRT scales is moderate to rather low in the majority group. However, most scales are quite short. Most alphas have comparable values in all groups and about a quarter are clearly lower for the immigrants.

If there are changes in factor loadings, the scale scores reflect differences between persons they were not supposed to measure. The factor analyses of the scale scores of the IRT and ABV showed that the factor solution of the immigrant groups strongly resembled the factor solution of the majority group. Only small changes in factor loadings were found and the values of the congruence coefficients were high; the percentages of explained variance in the majority and immigrants groups are quite similar. The results of the various analyses turned out to be consistent with one another and clearly support the hypothesis of dimensional comparability.

The items showed the most qualitative and quantitative DIF for the Turks and North Africans and the least qualitative and quantitative DIF for the Antillians and Surinamese. The items that showed quantitative DIF turned out to partially compensate for each other. This means that leaving out the items showing quantitative DIF did not result in a substantial improvement of the score profile of the immigrants.

Post-hoc explanations based on well-known differences between the immigrant and the majority group, such as the importance of social relationships, the traditional way of living, the unfavorable socioeconomic position, and a low level of knowledge of the Dutch language could be found for several items that showed quantitative DIF. However, on the whole it turned out to be difficult to predict the incidence of DIF beforehand.

It can be concluded that a taxonomy with the basic dimensions neuroticism and extraversion and the dimension rigidity can be generalised to immigrant populations in the Netherlands, and that the IRT can be used for assessment of immigrant groups that originate from Turkey, Northern Africa, the Netherlands Antilles, and Surinam. The scales appear to measure the same constructs in the various groups, the reliabilities are highly comparable in the majority of cases, and most of the items are equivalent. Validity data relating scale scores to training proficiency and job performance are of course required to allow strong conclusions.

The status of most immigrants in Dutch society is low and their level of unemployment is high. The most frequently mentioned possible causes for this are low education levels, the low level of knowledge of Dutch culture, the low level of command of the Dutch language, discrimination, and unfair treatment when trying to enter an organisation, specifically the use of non-optimal selection models and instruments. However, the data in this study give no proof that the low status of immigrants on the labor market is influenced by non-optimal personality tests.

Some comments about the use of the scales for comparison between groups may be required. The test has construct validity in the immigrant groups; comparability of dimensional structures and a partial compensation of the items showing quantitative DIF would seem to suggest that immigrants and majority group members with the same scale scores do not obtain widely different positions on the latent trait. More research is needed to determine whether the same position on the personality scales for persons from different groups leads to the same social behavior.

## Conclusion

The practical outcome of this study is that the personality test IRT can be used for comparisons within culturally homogeneous groups of non-native born, non-native language minority job applicants. On a theoretical level it



can be concluded that a taxonomy with the basic dimensions neuroticism and extraversion, and with the dimension rigidity can be generalised to immigrant populations in the Netherlands.

## REFERENCES

- Allen, J. (1998). Personality assessment with American Indians and Alaska Natives: Instrument considerations and service delivery style. *Journal of Personality Assessment, 70*, 17–42.
- Barrett, P. (1986). Factor comparison: An examination of three methods. *Personality and Individual Differences, 7*, 327–340.
- Barrett, P.T., Petrides, K.V., Eysenck, S.B.G., & Eysenck, H.J. (1998). The EPQ: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences, 25*, 805–819.
- Benet-Martínez, V., & John, O.P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait Multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729–750.
- Burt, C. (1948). The factorial study of temperamental traits. *British Journal of Psychology, 1*, 178–203.
- Butcher, J.N., Dahlstrom, W., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Central Bureau for Statistics (1996). *Statistisch jaarboek 1996* [Statistical yearbook 1996]. The Hague, The Netherlands: Central Bureau for Statistics.
- Cheung, F., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of Cross-Cultural Psychology, 29*, 233–248.
- Church, A.T., Katigbak, M.S., & Reyes, J.A. (1998). Further exploration of Filipino personality structure using the lexical approach: Do the Big-Five or Big-Seven dimensions emerge? *European Journal of Personality, 12*, 249–269.
- Cuellar, I. (1998). Cross-cultural clinical psychological assessment of Hispanic Americans. *Journal of Personality Assessment, 70*, 71–86.
- Dana, R.H. (1998). Cultural identity assessment of culturally diverse groups: 1997. *Journal of Personality Assessment, 70*, 1–16.
- De Raad, B., Perugini, M., Hrebícková, M., & Szarota, P. (1998). Lingua franca of personality: Taxonomies and structures based on the psycholexical approach. *Journal of Cross-Cultural Psychology, 29*, 212–232.
- Evers, A., Vliet-Mulder, J.C. van, & 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.
- Eysenck, H.J., & Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire* (Junior and Adult). London: Hodder & Stoughton.
- Eysenck, S.B.G., & Eysenck, H.J. (1962). Rigidity as a function of introversion and neuroticism: A study of unmarried mothers. *International Journal of Social Psychiatry, 7*, 180–184.

- Feij, J.A. (1985). Persoonlijkheidspsychologie [Personality psychology]. In J.F. Orlebeke, P.J.D. Drenth, R.H.C. Janssen, & C. Sanders (Eds.), *Compendium van de psychologie, deel 5*. Muiderberg, The Netherlands: Coutinho.
- Francis, L.J. (1998). Dogmatism and Eysenck's two-dimensional model of personality revisited. *Personality and Individual Differences, 24*, 571–573.
- Gough, H.G., & Bradley, P. (1996). *Manual for the California Psychological Inventory* (2nd edn.). Palo Alto, CA: Consulting Psychologists Press.
- Hahn, D.-W., Lee, K., & Ashton, M.C. (1999). A factor analysis of the most frequently used Korean personality trait adjectives. *European Journal of Personality, 13*, 261–282.
- Jöreskog, K.G., & Sörbom, D. (1988). *Lisrel-7* (computer manual). Mooresville, IN: Scientific Software.
- Lindsey, M.L. (1998). Culturally competent assessment of African American clients. *Journal of Personality Assessment, 70*, 43–53.
- McCrae, R.R., & Costa, P.T., Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882–888.
- McCrae, R.R., & Costa, P.T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509–516.
- McCrae, R.R., Costa, P.T., Jr., Del Pilar, G.H., Rolland, J.P., & Parker, W.D. (1998). Cross-cultural assessment of the five-factor model. *Journal of Cross-Cultural Psychology, 29*, 171–188.
- Nyborg, H., Eysenck, S.B.G., & Kroll, N. (1982). Cross-cultural comparison of personality in Danish and English children. *Scandinavian Journal of Psychology, 23*, 291–297.
- Okazaki, S. (1998). Psychological assessment of Asian Americans: Research agenda for cultural competency. *Journal of Personality Assessment, 70*, 54–70.
- Paunonen, S.V., & Ashton, M.C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Assessment, 29*, 150–170.
- 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.
- Somer, O., & Goldberg, L.R. (1999). The structure of Turkish trait-descriptive adjectives. *Journal of Personality and Social Psychology, 76*, 431–450.
- Stricker, L.J. (1982). Identifying test items that perform differentially in population subgroups: A partial correlation index. *Applied Psychological Measurement, 6*, 261–273.
- Tellegen, B. (1968). *Over rigiditeit* [On rigidity]. Zaltbommel, The Netherlands: Avanti.
- Ten Berge, J.M.F. (1986). Rotation to perfect congruence and the cross-validation of component weights across populations. *Multivariate Behavioral Research, 21*, 41–64.
- Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta analytic review. *Personnel Psychology, 44*, 703–742.
- Te Nijenhuis, J. (1997). Comparability of test scores for immigrants and majority group members in the Netherlands. Doctoral dissertation, Vrije Universiteit, Amsterdam.

- Te Nijenhuis, J., Van der Flier, H., & 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.
- Tucker, L.R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, DC: Dept of the Army.
- Van der Flier, H., Mellenbergh, G.J., Adèr, H.J., & Wijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement, 21*, 131–145.
- Wilde, G.J.S. (1970). *Neurotische labiliteit gemeten volgens de vragenlijstmethode* (2nd edn.) [Neurotic lability measured by the questionnaire method]. Amsterdam: Van Rossen.

