Running head: SIMILARITY VERSUS DISSIMILARITY DURING SELECTION

Assessor-Related Factors and Score Differences between Ethnically Diverse Dutch Police Applicants

Lonneke A. L. de Meijer and Marise Ph. Born

Erasmus University Rotterdam, The Netherlands

Hans van Loon

Police Academy of The Netherlands

Henk T. van der Molen

Erasmus University Rotterdam, The Netherlands

Paper for the 22nd Annual SIOP Conference, New York, NY

Correspondence: Lonneke A.L. de Meijer, Erasmus University Rotterdam, FSW Institute of Psychology, Woudestein T13-34, P.O. Box 1738, NL-3000 DR Rotterdam. E-mail may be sent to demeijer@fsw.eur.nl.

Abstract

The present study examined the effects of demographic and perceived similarity between assessors and applicants on assessors' evaluations of Dutch ethnic majority and minority applicants. Results showed that demographic similarity did not explain score differences between ethnic groups. Perceived similarity did explain score differences, but for Turkish applicants solely.

Press Paragraph

Measuring devices in which a perceiving party is present may be labeled subjective, as it is through subjective perception that the evaluation takes place. Several factors may influence inferences made by an assessor, among which the similarity between assessor and applicant. The present study explored the similarity between assessor and applicant ethnicity and ratings given during selection at the Dutch police on the assessment center (AC), the employment interview, and the employment recommendation. Results showed that perceived similarity was able to explain evaluations, but, only for the Turkish minority group. Demographic similarity did not have any explanatory power.

Assessor-Related Factors and Score Differences between Ethnically Diverse Dutch Police Applicants

In personnel selection, differences between ethnic majority and minority groups have been widely published upon in the domain of cognitive ability (e.g., Goldstein, Zedeck, & Goldstein, 2002; Herrnstein & Murray, 1994) and, to a somewhat lesser extent, in the domain of personality (e.g., Hough, 1998; Ones & Anderson, 2002). A characteristic that cognitive ability tests and personality questionnaires have in common, is that these can be labeled as *objective* measures, as there is no influence of a perceiving party rating the applicant (cf., Bass & Barrett, 1981). In contrast, measuring devices in which a perceiving party is present (e.g., an assessor, an interviewer), may be labeled as *subjective* (cf., Bass & Barrett, 1981). It is through the subjective perception by an assessor that the evaluation of an applicant takes place.

During interpersonal perception a host of factors may influence impressions and inferences made by a rater, among which affective processes, interpersonal factors, and motivation and skills of the rater. With regard to interpersonal factors, the similarity between the rater and the ratee may be expected to have an influence on the outcome of perceptual processes (Fiske & Taylor, 1991; Klimoski & Donahue, 2001). It is this similarity issue, which is the focus of the present study. Our study explores the relationship between ethnicity and scores on several selection instruments in which judgments by assessors are involved. The study goal is to map the relative extent to which ethnic similarity between assessors and applicants is able to explain existing score differences between ethnic groups on several subjective instruments. These are the

assessment center (AC), the employment interview, and the final employment recommendation that the assessor gives about the applicant to the client organization. The study has been executed in the context of the selection procedure of the Dutch police. Pursuing this goal, two approaches are involved.

The first approach concerns whether *demographic similarity* – in this case actual ethnic similarity – between assessors and applicants will influence the way assessors rate applicants. Empirical findings until now have shown mixed results concerning the effects of demographic similarity in personnel selection and on work related outcomes. Sacco, Scheu, Ryan, and Schmitt (2003) did not find support for the effect of demographic similarity on interview ratings. However, McFarland, Ryan, Sacco, and Kriska (2004) found several complex interaction effects of rater ethnicity, ratee ethnicity, and panel ethnic composition. Sacco et al. (2003) examined the demographic similarity effect on interview scores differentiating various ethnic groups, (i.e., White, Black, Hispanic, and Asian raters and ratees). Using multilevel analysis, which takes into consideration the nesting of applicants within raters, they found no evidence that ethnic similarity played a significant role in determining the interview ratings assigned to any of the applicant groups. McFarland et al. (2004), not using multilevel analysis but less sophisticated analysis-of-variance techniques, examined Black and White raters and ratees only. They showed that Black raters evaluated Black applicants more favorably than White applicants, but only when the panel composition was predominantly Black. In the more general context of work, Chattopadhyay, Tluchowska, and George (2004) demonstrated that demographic similarity affects a range of work-related outcomes, including organizational commitment and performance. They examined existing literature on

demographic similarity; however, they did not differentiate between research using multilevel analysis and research using analysis-of-variance techniques while data were nested. When analysis-of-variance techniques are used while multilevel analysis is more appropriate (i.e., when data are nested), the question rises whether significant results were truly found when there was, in fact, inadequate evidence for rejecting the null hypothesis (type-I error).

When not taking into account the analyses used, the latter study probably most clearly has confirmed the common demographic similarity hypothesis. This hypothesis quite straightforwardly states that similarity, in general, will lead to higher ratings. This expectation is derived from social identity theory (SIT; Tajfel, 1982; Turner, 1987), which contends that aspects of an individual's self image come from the social categories to which he/she perceives him/herself as belonging (e.g., ethnic group, gender). Social identity is seen as necessary to boost one's self esteem. To the extent that individuals' social identities and self-categorizations are built around their demographic characteristics, demographic dissimilarity may have a negative effect on the attitudes and behaviors towards others, whereas higher identification and similarity may lead to more positive attitudes and behaviors towards other people. In line with this common notion, it may be expected that demographic similarity between assessor and applicant will lead to higher ratings (Hypothesis 1a).

Alternatively, it should be recognized that assessors in selection procedures will have a strong motive to be accurate. The costs of being wrong in an applied setting such as personnel selection implies that these raters will tend to invest more effort in the judgment task than individuals in general, and therefore that they may be applying

strategies which are more detailed and complex. Well-trained assessors, who have learned to focus on a structured task for the duration of an AC or during interviewing, may thus be less influenced by such perceptions of (dis)similarity. Therefore, competing with Hypothesis 1a, it may be expected that a weak or even null-relationship will be found between demographic similarity and ratings given by assessors in an applied setting of employee selection (Hypothesis 1b).

The second approach stems from organizational research looking into the effects of *perceived intergroup similarity*. In this approach, the focus moves from similarity in objective demographic characteristics to perceptions of similarity, including less tangible attributes such as values, beliefs, and personality. Most investigations of perceived-similarity effects have focused on employee relationships and performance (e.g., Ensher & Murphy, 1997; Lankau, Riordan, & Thomas, 2005; Strauss, Barrick, & Connerley, 2001; Turban, Dougherty, & Lee, 2002). As with demographic similarity, some studies (e.g., Strauss et al., 2001; Turban et al., 2002) did not use multilevel analysis while it would be appropriate because of their nested data. Again, the question rises whether significant results were truly found in these studies. Nevertheless, studies using ANOVA or regression as an approach that is well suited to examining data that does not show a nested structure (e.g., Ensher & Murphy, 1997; Lankau et al, 2005) have generally supported that perceived similarity is positively related to relevant dependent variables (such as mentoring quality). We would like to extend the findings to the selection context. It may, therefore, be expected that the more assessors perceive an applicant's ethnic group as similar to themselves, the higher the applicant will be rated. Nevertheless, in the present study, we would like to go one step further and examine the differential

effect of perceived similarity between ethnic minority groups. That is, we argue that perceived similarity might have a different effect for one ethnic minority group relative to another, depending on the degree to which a certain ethnic group is integrated into society. When the members of a certain ethnic minority group are very isolated from the society in which they live and the general societal perception of this group is one of not being integrated, this group will be perceived as less similar relative to other – more integrated – groups. Perceptions of similarity toward a less integrated – more isolated – minority group may have a more outspoken positive effect on evaluations of applicants than perceptions of similarity toward a minority group that is more integrated and thus already more similar to the ethnic majority group.

In The Netherlands several studies have been conducted to examine the integration hierarchy in Dutch society of different ethnic groups (e.g., Hraba, Hagendoorn, & Hagendoorn, 1989; Verkuyten, Hagendoorn, & Masson, 1996; Weijters & Scheepers, 2003). In this ethnic hierarchy, European groups were placed on top, followed by former colonial and finally Islamic groups at the bottom. The largest ethnic minority groups in The Netherlands are from the Dutch Antilles, Morocco, Surinam, and Turkey. The Dutch Antilles and Surinam are former Dutch colonies and Morocco and Turkey are (mostly) Islamic. Assuming that assessors in the selection context share the general notion about the integration hierarchy in Dutch society, it is expected that the Antillean and Surinamese groups are viewed by assessors as most integrated in Dutch society and the Turkish and Moroccan groups as least integrated (Hypothesis 2).

As mentioned before, it may be expected that the perceived-similarity effect is not found to the same extent across ethnic groups, but that this effect is dependent on the

degree of integration in society. Combining what we know about perceived similarity and integration, it may be expected that the less assessors in general view an ethnic minority group as integrated, i.e., the more this group is viewed as isolated from society, the more effect an individual assessor's perceived similarity of this minority group will have on the scores given. Vice versa, it may be expected that the more an ethnic minority group is viewed as integrated into society, i.e., the more it is viewed as similar to the ethnic majority group, the less effect perceived similarity by the individual assessor of this – well-integrated – minority group will have on the scores given. Given the integration hierarchy in Dutch society, it, therefore, is expected that the more an assessor perceives Turkish or Moroccan minority groups as similar to himself or herself, the higher Turkish or Moroccan applicants will be rated (Hypothesis 3a). Furthermore, it is expected that perceived similarity has a weak or even a null-effect on ratings given to Dutch Antillean and Surinamese applicants (Hypothesis 3b).

Nevertheless, as stated before for the case of demographic similarity, we may again, alternatively, expect that well-trained assessors who are motivated to increase the accuracy of their judgments will show a weak or even null-relationship for all ethnic minority groups between perceived similarity and the evaluations given (Hypothesis 3c).

Finally, relating the findings on demographic and perceived similarity, several researchers found stronger effects for perceived than for demographic similarity in the domains of mentoring relationships (e.g., Ensher, Grant-Vallone, & Marelich, 2002) and performance appraisal (e.g., Strauss et al., 2001). Ferris and Judge (1991) suggest that one reason for stronger effects of perceptions of similarity is that people react on the bases of perceptions of reality, not on the basis of reality *per se*. In line with this

reasoning, it is expected that perceived similarity will have a stronger effect on ratings than demographic similarity (Hypothesis 4).

Method

Participants and Procedure

Data came from 11,432 applicants who applied for a position at the Police Academy of The Netherlands from September 2001 until July 2003. Of these, n = 1,406 (12%) were ethnic minority applicants. Minority applicants came from the largest ethnic minority groups in The Netherlands, namely from Dutch Antillean, Moroccan, Surinamese and Turkish ethnic groups.

Applicants who are interested in a job as police officer first apply to the local police force where they want to work after completion of their training. For the selection procedure, the local police forces routinely send all applicants to the national Police Center for Competence Assessment and Monitoring (CCM). Applicants go through two stages in the selection process. During the first stage a Dutch language-proficiency test is filled out. During the second phase a physical exercise, a cognitive ability test, a personality test, an assessment center (AC) assignment and an employment interview are executed. The psychologist who conducts the interview is also the one who writes and gives the final employment recommendation to the local police force. For this recommendation, the test results of the personality test, the AC ratings, and the employment interview ratings are integrated.

To investigate the effects of demographic and perceived similarity, ratings from the AC, the employment interview, and the final employment recommendation were used. In the remainder of this paper, two groups of raters are examined, namely the

assessors who conduct the AC, and the psychologists who conduct the interview and write the final recommendation.

Data from 82 assessors (80% female; 93% ethnic majority-group member, M_{age} = 30.61, SD = 6.22) and 75 psychologists (81% female; 97% ethnic majority-group member; M_{age} = 29.16, SD = 4.91), evaluating 11,432 applicants, were used to investigate the effect of *demographic similarity* (Hypotheses 1a and 1b) on score differences between ethnic groups. On average, each assessor evaluated 131 applicants and each psychologist 144 applicants.

Related to *perceived similarity* and *integration* (Hypotheses 2, 3a, 3b, and 3c), evaluations by 15 assessors (80% female; 93% ethnic majority-group member; M_{age} = 31.20, SD = 6.45) and 12 psychologists (92% female; 100% ethnic majority-group member; M_{age} = 28.55, SD = 4.30), evaluating 4,714 applicants, were used. On average, each assessor evaluated 171 applicants and each psychologist 244 applicants. With regard to perceived similarity, the assessors (n = 15) and the psychologists (n = 12) are sub-samples of the total group of assessors (n = 82) and psychologists (n = 75) examined for demographic-similarity effects.

All raters had a high educational level (higher professional education ["HBO"] or academic-oriented education ["WO"]). Table 1 shows the sample sizes of each applicant type-rater type combination.

Measures

Personality Test

To measure the Big Five factors Extraversion, Altruism, Conscientiousness, Emotional Stability, and Intellect, the Police Personality Questionnaire (PPV; Van

Leeuwen, 2000) was used. The applicants completed the PPV in the Dutch language. A report by Klinkenberg and Van Leeuwen (2003) indicated alpha reliabilities varying from .72 for Conscientiousness, to .78 for Intellect. Correlations between the scales were all lower than .60. A comparison with the NEO-PI-R showed observed construct validity coefficients between .17 and .58 (N = 160). A study by Lem and Van Doorn (2000) showed observed predictive validity coefficients between .15 and .43 (N = 61) for the prediction of supervisory evaluations of job performance.

Assessment Center (AC)

A role-play exercise was utilized, in which an assessor and an actor independently made ratings on a 7-point Likert-scale ranging from 1 (extremely weak) to 7 (excellent), on each of the following seven dimensions: Communication Skills, Social Skills, Empathy, Initiative, Stress Tolerance, Authority, and Decisiveness. Interrater reliabilities ranged from .82 to .88 (N = 198). Principal component analysis with varimax rotation yielded two factors, Agency and Communion (in accordance with Wiggins and Trapnell, 1996), which explained 77% of the variance. As a measure of Agency, the average rating across the dimensions Authority, Decisiveness, Initiative, Communication Skills, and Stress Tolerance was used ($\bar{r}$ = .59; α = .87). As a measure of Communion, the average rating of the dimensions Social Skills and Empathy was used ($\bar{r}$ = .77; α = .87). The reliability of the difference (r_{diff}) between scores on Agency and Communion was .78.

Employment Interview

The interview questions were focused on evaluating behavior on the following eight dimensions: Communication Skills, Social Skills, Flexibility, Stress Tolerance, Emotional Stability, Tolerance Towards Others, Integrity, and Self-Understanding. A

single psychologist conducted the interview. The interviews were semi-structured and behaviorally based, with one behaviorally anchored 7-point Likert scale ranging from 1 (extremely weak) to 7 (excellent) for each of the eight dimensions. The average rating across the eight dimensions was used as the dependent variable because the ratings were substantially correlated ($\bar{r}$ = .42; α = .85). Moreover, principal component analysis with varimax rotation yielded one interview factor that explained 50% of the variance.

Final Employment Recommendation

The final employment recommendation as to whether an applicant is fit for a job as police officer, was based on results from the PPV, the AC, and the employment interview. These scores were integrated into a final recommendation. The dimensions in the final recommendation were: Communication Skills, Social Skills, Empathy, Initiative, Flexibility, Stress Tolerance, Authority, Decisiveness, Tolerance Towards Others, Integrity, and Self-Understanding (for definitions, see Appendix 1). A 7-point Likert scale ranging from 1 (extremely weak) to 7 (excellent) was used to evaluate the behavior on each of the eleven dimensions. Principal component analysis with varimax rotation yielded three final-recommendation factors, Agency, Communion, and Socio-Cultural Awareness, which altogether explained 67% of the variance. As a measure of Agency, the average rating across the dimensions Authority, Decisiveness, Initiative, Communication Skills, Stress Tolerance, and Flexibility was used ($\bar{r}$ = .48; α = .85). As a measure of Communion, the dimensions Social Skills and Empathy were used ($\bar{r}$ = .66; α = .79), and for Socio-Cultural Awareness the dimensions, Tolerance Towards Others, Integrity, and Self-Understanding ($\bar{r}$ = .39; α = .65). The reliability of the difference (r_{diff}) between scores on Agency and Communion equals .51, r_{diff} between scores on Agency

and Socio-Cultural Awareness equals .58, and r_{diff} between scores on Communion and Socio-Cultural Awareness equals .57.

Perceived Similarity Questionnaire

Fifteen assessors and twelve psychologists filled out a questionnaire measuring perceived similarity, which was derived from a measure by McCroskey, Richmond, and Daly (1975). In this questionnaire, assessors and psychologists filled out to what extent they perceived the average member of a particular ethnic minority group (Dutch Antilleans, Surinamese, Moroccans, and Turks) to be similar to themselves on four aspects, namely attitudes, values, physical appearance, and background. A sample item is 'I am of the opinion that Turkish people have the same norms and values as I have' (Likert scale from 1 to 7). For each ethnic minority group, the scores were averaged across the four aspects, as the intercorrelation between the similarity perceptions was quite high ($\bar{r}$ = .60; α = .90).

Perceived Integration

An additional item on the perceived similarity questionnaire asked the assessors and psychologists to what extent they perceived the average member of a particular ethnic minority group (Dutch Antilleans, Surinamese, Moroccans, and Turks) to be integrated into Dutch society (on a 4-point scale).

Analyses

To investigate the effect of demographic similarity (Hypotheses 1a and 1b) and perceived similarity (Hypotheses 3a, 3b, and 3c) on the scores given on the AC, the employment interview, and the final employment recommendation, hierarchical linear modeling with MLwiN 1.10 (Center for Multilevel Modeling, 1997) was used. This

technique provides for a statistically accurate treatment of nested variables. Since evaluations of applicants (level 1 [L1]) involve data nested within raters (level 2 [L2]), such dependency needs to be dealt with correctly.

Hypothesis testing in MLwiN involves evaluating a series of models. We followed the procedure used by Sacco et al. (2003), which will be outlined here. We refer to level 1 (L1) or level 2 (L2) when discussing applicant and rater effects, respectively. The significant difference in deviance (-2 * log likelihood) between an initial model and a subsequent model is a prerequisite for finding significant results in this subsequent model. In the first step, which examines within- and between-group variance (equivalent to one-way ANOVA), a *null model* is tested.

$$\text{L1: } y_{ij} = \beta_{0j} + e_{ij} \quad (1)$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + \mu_{0j} \quad (2)$$

The L1 equation predicts ratings received by applicants on the AC, the interview, or the final recommendation (y_{ij}) based on the mean rating (i.e., intercept) within each of the *j* raters (β_{0j}) and the error for each of *i* applicants (e_{ij}). The L2 equation models each rater's mean rating based on the grand mean (i.e., intercept; γ_{00}) and each rater's deviation (error parameter μ_{0j}). In addition, the associated variance components of the terms μ_{0j} and e_{ij} can be used to calculate the intra-class correlation (ICC), which indexes the ratio of the between-rater variance in ratings to the total variance. Barcikowski (1981) showed that even a small ICC can inflate the alpha level (type-I error) substantially. This means that even in the case of a small ICC, i.e., when raters do not differ much among each other in the ratings given, the nested data structure should be taken into account and multilevel analysis should be used.

In the second step, the first independent variable (i.e., applicant's ethnicity [x_{1ij}]) is added to the L1 equation:

$$\text{L1: } y_{ij} = \beta_{0j} + \beta_{1j}\, x_{1ij} + e_{ij} \quad (3)$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + \mu_{0j} \quad (4)$$

$$\text{L2: } \beta_{1j} = \gamma_{10} + \mu_{1j} \quad (5)$$

This model is known as the *random coefficients model* because the regression coefficients β_{0j} and β_{1j} are modeled as random effects at L2 (see Equations 4 and 5). This means that, in the random coefficient model, groups of applicants (i.e., applicants rated by different assessors or psychologists) are allowed to deviate from the mean solution, not only in the intercept (γ_{00}) but also in the slope (γ_{10}). The significance of the L2 parameters (γ_{00} and γ_{10}) indicates whether ratings are significantly different from zero and whether applicant's ethnicity is related to ratings, respectively. The error parameters μ_{0j} and μ_{1j} are associated with the corresponding coefficients at L1, namely β_{0j} and β_{1j} respectively.

If the fit of the random coefficients increases significantly over and above the null model, implying that taking into account the applicant's ethnicity results in a better fit to the data, the third step involves examining whether a L2 variable (i.e., rater's ethnicity when investigating demographic similarity or rater's perceived similarity when investigating perceived similarity [x_{2j}]) predicts the variability in the intercepts of applicants' ethnicity at L1:

$$\text{L1: } y_{ij} = \beta_{0j} + \beta_{1j}\, x_{1ij} + e_{ij} \quad (6)$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + \gamma_{01}\, x_{2j} + \mu_{0j} \quad (7)$$

$$\text{L2: } \beta_{1j} = \gamma_{10} + \mu_{1j} \quad (8)$$

This *intercepts-as-outcomes model* tests for significant differences in mean ratings as a function of rater's ethnicity or of rater's perceived similarity (γ_{01}). If the fit of the intercepts-as-outcomes model is better than the random coefficients model, the fourth and final step involves estimating the following equations:

$$\text{L1: } y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij} \quad (9)$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + \gamma_{01} x_{2j} + \mu_{0j} \quad (10)$$

$$\text{L2: } \beta_{1j} = \gamma_{10} + \gamma_{11} (x_{1ij} * x_{2j}) + \mu_{1j} \quad (11)$$

This is known as the *slopes-as-outcomes model* because rater's ethnicity or rater's perceived similarity is used to predict variability in the intercepts (γ_{01}) and the slopes (γ_{11}) of applicants' ethnicity at L1. A significant γ_{11} coefficient would be evidence for a cross-level interaction, implying that ethnicity of the rater or perceived similarity of the rater moderates the relationship between the applicant's ethnicity and the ratings given.

Concerning the integration hierarchy of the four largest ethnic minority groups in The Netherlands as viewed by assessors and psychologists (Hypothesis 2), the mean rank of each minority group and Kendall's coefficient of concordance (Kendall's *W*) was calculated. Significant differences between the mean ranks of the four groups were tested with a chi-square test.

Results

Preliminary Results

The intra-class correlation coefficients (ICC) related to rater differences in scoring, varied between .04 and .17 (see Table 2). An ICC below .10 is viewed as a rule of thumb below which multilevel analyses is not necessary. Barcikowski (1981), nevertheless, showed that even small values of the ICC can cause a substantial increase in

the chance of type-I error to occur. Therefore, we decided to use multilevel analyses for all selection measures and both for demographic and perceived similarity, even though some ICC values were below .10.

Before testing the hypotheses, scores differences between the ethnic majority and minority group were looked into, as investigating similarity effects on score differences is useful only when these score differences actually exist. With regard to demographic similarity, the results in Table 3 showed significant score differences (γ_{10}, $p < .05$) to the advantage of the ethnic majority group on almost all measures. Exceptions were the final-recommendation factor Socio-Cultural Awareness, where the Dutch Antillean ($\gamma_{10} = -.10$, *ns*) and the Moroccan group ($\gamma_{10} = -.08$, *ns*) did not score differently from the ethnic majority group. With regard to perceived similarity, the results in Table 4 showed that on thirteen of 24 comparisons the scores were significantly different in the expected direction. However, on eleven comparisons, the scores between the ethnic majority group and ethnic minority groups were not significantly different, namely concerning: 1) the Moroccan ($\gamma_{10} = -.18$, *ns*) and Surinamese group ($\gamma_{10} = -.22$, *ns*) on the AC-factor Agency, 2) the Antillean ($\gamma_{10} = -.36$, *ns*), Moroccan ($\gamma_{10} = -.30$, *ns*), and Surinamese group ($\gamma_{10} = -.31$, *ns*) on the AC-factor Communion, 3) the Antillean group ($\gamma_{10} = -.20$, *ns*) on the interview, 4) the Moroccan group ($\gamma_{10} = -.13$, *ns*) on the final-recommendation factor Agency, 5) the Antillean ($\gamma_{10} = -.01$, *ns*) and Moroccan group ($\gamma_{10} = -.19$, *ns*) on the final-recommendation factor Communion, and 6) the Antillean ($\gamma_{10} = -.11$, *ns*) and Surinamese group ($\gamma_{10} = -.13$, *ns*) on the final-recommendation factor Socio-Cultural Awareness. In these cases, subsequent models (the intercepts-as-outcomes model and the slopes-as-outcomes model) were not tested.

Main Results

The *demographic similarity* hypothesis stated that actual ethnic similarity would lead to higher ratings (Hypothesis 1a). Alternatively, as assessors and psychologists during selection will have a strong motive to be accurate, Hypothesis 1b stated that a weak or even null-relationship would be found between demographic similarity and ratings given by assessors and psychologists in an applied setting such as employee selection.

With regard to all measures, hierarchical linear modeling results (see Table 3) did not show support for Hypothesis 1a, but for Hypothesis 1b. Neither concerning the AC (for Agency: $.69 < \Delta\chi^2$ ($\Delta df = 1$) < 1.40, *ns*; for Communion: $.00 < \Delta\chi^2$ ($\Delta df = 1$) $< .80$, *ns*), the interview ($.05 < \Delta\chi^2$ ($\Delta df = 1$) < 1.95, *ns*), nor concerning the final employment recommendation (for Agency: $.00 < \Delta\chi^2$ ($\Delta df = 1$) < 2.81, *ns*; for Communion: $.00 < \Delta\chi^2$ ($\Delta df = 1$) $< .79$, *ns*; and for Socio-Cultural Awareness: $.70 < \Delta\chi^2$ ($\Delta df = 1$) $< .75$, *ns*), did the slopes-as-outcomes model fit better than the intercepts-as-outcomes model. This implies that the results showed no effect of an interaction between applicants' ethnicity and assessors' or psychologists' ethnicity, which is in line with Hypothesis 1b – the no-effect hypothesis.

Concerning the integration hierarchy as viewed by the group of assessors and psychologists of the four largest ethnic minority groups in The Netherlands, Hypothesis 2 stated that Dutch Antillean and Surinamese groups would be placed on top, followed by Turkish and Moroccan groups at the bottom. A chi-square test revealed that the four ethnic minority groups indeed were perceived as not being equally integrated (χ^2 ($df = 3$) $= 36.00$, $p < .001$). Results in Table 5 supported Hypothesis 2 and showed the following

hierarchy (with substantial agreement among the assessors and psychologists; Kendall's $W = .60$): The Surinamese minority group (mean rank = 3.60) was perceived as most integrated into Dutch society, followed by the Antilleans (mean rank = 2.80). The Turks (mean rank = 2.40) and the Moroccans (mean rank = 1.20) were perceived as the least integrated minority groups.

With regard to *perceived similarity,* it was stated that the more an assessor or psychologist perceives Turkish or Moroccan minority groups as similar to himself or herself, the higher the Turkish or Moroccan applicants will be rated (Hypothesis 3a). Furthermore, it was expected that assessors' or psychologists' perceived similarity will have a weak or no effect on ratings given to the Dutch Antillean and Surinamese applicants (Hypothesis 3b). Alternatively, Hypothesis 3c stated that we might expect that well-trained assessors or psychologists who are motivated to be accurate in their judgments will show a weak or even null-relationship between perceived similarity and the evaluations given. Perceived similarity judgments were given for the four largest ethnic minority groups, namely Dutch Antilleans, Surinamese, Moroccans, and Turks.

With regard to the selection measures involved, hierarchical linear modeling results (see Table 4) showed partial support for Hypothesis 3a, namely concerning the Turkish minority group and concerning the final-recommendation factor Socio-Cultural Awareness solely. Hypotheses 3b and 3c – the no-effect hypotheses – were supported for the Antillean, the Surinamese, ánd the Moroccan minority group and on all other selection measures and dimensions.

For the evaluations on the final-recommendation factor Socio-Cultural Awareness concerning the Turkish minority group, the slopes-as-outcomes model fitted significantly

better than the intercepts-as-outcomes model ($\Delta\chi^2$ ($\Delta df = 1$) = 5.50, $p < .05$). Psychologists who perceived Turkish applicants in general to be more similar to themselves gave Turkish applicants higher scores on the final-recommendation factor Socio-Cultural Awareness ($\gamma_{11} = .32$, $p < .05$). No such effects were found for the AC (for Agency: $.08 < \Delta\chi^2$ ($\Delta df = 1$) < 1.40, *ns*; for Communion: $\Delta\chi^2$ ($\Delta df = 1$) = .05, *ns*), the interview ($.00 < \Delta\chi^2$ ($\Delta df = 1$) $< .37$, *ns*), and the final-recommendation factors Agency ($.00 < \Delta\chi^2$ ($\Delta df = 1$) $< .54$, *ns*) and Communion ($.04 < \Delta\chi^2$ ($\Delta df = 1$) $< .11$, *ns*). Thus, it particularly seems to be the Turkish group that is influenced by perceived similarity, although not on the interview or on the dimensions of Agency and Communion.

When comparing the effects found for demographic and perceived similarity, the present study shows stronger effects for perceived similarity for the Turkish group on Socio-Cultural Awareness. Therefore it seems to support Hypothesis 4 only partially and only in line with other study results in which effects for perceived similarity were found to be clearer than for demographic similarity in work-related domains such as mentoring and performance appraisal (e.g., Ensher et al., 2002; Strauss et al., 2001).

Discussion

Within the area of personnel selection, the diversification of the workforce has become an important goal in the industrialized world. One strategy to avoid the substantial adverse impact that generally is caused by cognitive ability tests (e.g., Murphy, 2002), is to use a series of face-valid non-cognitive ability selection tools. Such tools include the AC and the employment interview. These often involve a rater who will subjectively give an evaluation of the ability, behavior, or aptitude of the applicant. Although score differences between ethnic groups on these subjective measures are found

to be smaller than on the cognitive ability test, they exist and are still quite substantial (De Meijer, Born, Terlouw, & Van der Molen, 2006). Yet, it is not very well known to what extent subjectivity of ratings may contribute to systematic score differences between ethnic groups on such measures.

To investigate whether there are any systematic differences in assessors' evaluations of members of different ethnic groups, we looked into effects of similarity between raters and applicants in terms of ethnicity. In doing so, we used two approaches respectively investigating the effects of demographic similarity and perceived similarity between raters and applicants on score differences on the AC, the employment interview, and the final recommendation between the ethnic majority group and the four largest ethnic minority groups in The Netherlands (i.e., Dutch Antilleans, Moroccans, Surinamese, and Turks). Our data came from a field study in the context of personnel selection at the Dutch police (N = 11,432). Multilevel analysis was used to deal with the nested structure of our data. One earlier study, using this more appropriate method of analysis (Sacco et al., 2003) examining, however, only demographic similarity, yielded no effects on scores given. The question rises whether earlier research, which has analyzed demographic- as well as perceived-similarity effects at the individual level without taking into account the nested nature of the data (McFarland et al., 2004; Strauss et al., 2001; Turban et al., 2002), might have unjustly concluded that significant effects existed when there was, in fact, inadequate evidence for rejecting the null hypothesis (type-I error). This type-I error is likely to occur when data are structured in multiple levels, as in our study. Therefore, we followed a multilevel-analysis procedure used by

Sacco et al. (2003) to investigate demographic-similarity effects and also expected its use to investigate perceived similarity.

Results showed no effects of demographic similarity. Perceived similarity, however, did show an effect for those applicants that were viewed as less integrated into Dutch society. More specifically, the effect of perceived similarity pertained to one ethnic minority group only, namely the Turkish group and only on the final-recommendation factor Socio-Cultural Awareness. We will now try to explain these findings.

Several studies in The Netherlands (e.g., Van Rijn, Zorlu, Bijl, & Bakker, 2004) have indicated the isolated position for this specific minority group. Yet, in the present study, the assessors and psychologists placed the Moroccan minority group at the bottom of the integration hierarchy and not the Turkish group. Why then does perceived similarity not also play a role in explaining score differences for the Moroccan group, as it does for the Turkish minority group? With regard to the integration hierarchy, Hagendoorn (1995) found that the ethnic majority group ranks ethnic groups on the basis of cultural differences. Furthermore, Pinto (2004) showed that Moroccans are more culturally different from the ethnic majority group than Turks are. He argues that this is due to the fact that Moroccans are viewed as more traditional, more religious, and more aggressive than other ethnic minority groups. These are all quite negative viewpoints toward the Moroccan minority group that seem to exist in Dutch society at large. Furthermore, these arguments show why it is likely that Moroccans are viewed as more culturally different and, hence, as less integrated into Dutch society than other minority groups, as was found in our study. So, with regard to perceived similarity, it is somewhat surprising that no effect was found for the Moroccan group, but only for the Turkish

group? The Turkish group, as the Moroccan group, has a strong sense of their own Islamic culture and history (e.g., Nijsten, 1998). More than in the Moroccan group, however, this strong sense of an own culture and history in the Turkish group manifests itself in loyalty, cohesion, and solidarity within the Turkish group and in avoidance of contact with other ethnic groups (Verkuyten et al., 1996). Loyalty, cohesion, and solidarity will generally be seen as quite positive aspects of a group (in contrast to traditional, religious, and aggressive aspects of the Moroccan group). Because of these positive aspects, the Turkish group might be viewed as better functioning and better integrated in Dutch society than the Moroccan group. However, at the same time, the Turkish group might also be viewed as a highly cohesive group, as a group that avoids contact with other ethnic groups, and as an isolated group. It might actually be this type of integration, namely the avoidance of contact with or being isolated from society at large, which is of importance, here. In terms of isolation from Dutch society, the Turkish group is likely to score lower than the Moroccan group. Perceptions of similarity might therefore play a larger role for the Turkish minority group than for other minority groups. The fact that perceived similarity affects scores on the final-recommendation factor Socio-Cultural Awareness for Turkish applicants, but not scores on the interview and the two other specific AC- and final-recommendation factors (Agency and Communion), also may point towards issues of isolation from society, which are at stake for this group. As described before, the factor Socio-Cultural Awareness involves: 1) taking differences between people into consideration in one's own behavior, 2) willingness to acknowledge the norms and values in society, and 3) being able to reflect on one's own behavior and willingness to integrate these reflections in future behavior. When the general perspective

of the Turkish minority group in Dutch society is one of not being open and tolerant toward other ethnic groups, perceptions of similarity toward Turkish applicants might have a more outspoken effect on the dimension that measures this openness and tolerance. The fact that Turkish applicants who are perceived as more dissimilar by psychologists receive a lower score on the factor Socio-Cultural Awareness and not on other – less related – factors might, therefore, be seen as evidence for the validity of assessors' perceptions. The finding that for the less isolated groups, namely the Dutch Antillean, Moroccan, and Surinamese minority groups, perceived similarity did not play a role on scores given by psychologists on Socio-Cultural Awareness supports this interpretation.

With regard to demographic similarity, we do need to remark on the limitation that only a very small sample of ethnic minority assessors (n = 5) and psychologists (n = 2) were included in the study. This small sample may have suppressed any potential effects of demographic similarity between ethnic minority raters and applicants on the ratings given. Although no interaction-effect of rater ethnicity and applicant ethnicity on the scores given was found for ethnic majority raters, it might be too premature to conclude that the same null-effect exists for ethnic minority raters. Future research should include larger samples of ethnic minority raters to investigate this issue in detail.

In conclusion, in the present study no evidence was found for bias that differentially affected evaluations of ethnically diverse applicants during personnel selection. Although an effect of perceived similarity for the Turkish minority group was found on the dimension Socio-Cultural Awareness, this effect rather should be interpreted as evidence for the validity of raters' perceptions than evidence for biased raters. A first

explanation for the lack of effects of demographic and perceived similarity on given scores is that, during personnel selection, raters have a strong motive to be accurate. The costs of being wrong imply that raters will invest more effort in the judgment task than individuals in general. A second explanation is that well-trained raters have learned to focus on a structured task and, therefore, will be less influenced by aspects of (dis)similarity. Future research examining the effects of demographic and perceived similarity in other fields, such as work outcomes, is necessary.

References

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics, 6*(3), 267-285.

Bass, B. M., & Barrett, G. V. (1981). *People, work, and organizations* (2nd ed.). Boston: Allyn & Bacon.

Chattopadhyay, P., Tluchowska, M., & George, E. (2004). Indentifying the ingroup: A closer look at the influence of demographic dissimilarity on employee social identity. *Academy of Management Review, 29*(2), 180-202.

Center for Multilevel Modeling (1997). MLwiN 1.10 [Computer software]. Institute of Education, London.

De Meijer, L. A. L., Born, M. Ph., Terlouw, G., & Van der Molen, H. T. (2006). Applicant and method factors related to ethnic score differences in personnel selection: A study at the Dutch police. *Human Performance, 19*(3), 219-251.

Ensher, E. A., Grant-Vallone, E. J., & Marelich, W. D. (2002). Effects of perceived attitudinal and demographic similarity on protégés' support and satisfaction gained from their mentoring relationships. *Journal of Applied Social Psychology, 32*(7), 1407-1430.

Ensher, E. A., & Murphy, S. E. (1997). Effects of race, gender, perceived similarity, and contact on mentor relationships. *Journal of Vocational Behavior, 50*(3), 460-481.

Ferris, G. R., & Judge, T. A. (1991). Personnel/human resources management: A political influence perspective. *Journal of Management, 17,* 447-488.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Goldstein, H. W., Zedeck, S., & Goldstein, I. L. (2002). *G*: is this your final answer? *Human Performance, 15* (1/2), 123-142.

Hagendoorn, L. (1995). Intergroup biases in multiple group systems: The perception of ethnic hierarchies. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (pp. 199-211). London: John Wiley & Son.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve.* New York: Free Press Paperbacks.

Hough, L. M. (1998). Personality at work: Issues and evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131-159). Hillsdale, NJ: Erlbaum.

Hraba, J., Hagendoorn, L., & Hagendoorn, R. (1989). The ethnic hierarchy in The Netherlands: Social distance and social representation. *British Journal of Social Psychology, 28,* 57-69.

Klimoski, R. J., & Donahue, L. M. (2001). Person perception in organizations: An overview of the field. In M. London (Ed.), *How people evaluate others in organizations* (pp. 5-43). Mahwah, NJ. Lawrence Erlbaum.

Klinkenberg, E. L., & Van Leeuwen, A. E. (2003). *Voortgangsverslag ontwikkeling M5Q-IWSP* [Progress report development M5Q-IWSP]. Culemborg, The Netherlands: Meurs Personeelsadvies.

Lankau, M. J., Riordan, Ch. M., & Thomas, Ch. H. (2005). The effects of similarity and liking in formal relationships between mentors and protégés. *Journal of Vocational Behavior, 67,* 252-265.

Lem, J. & Van Doorn, E. (2000*). Voortgangsrapportage. Onderzoek "kenmerkende voorspellers" politie* [Progress report. Study "noticeable predictors" police]. Culemborg, The Netherlands: Meurs Personeelsadvies.

McCroskey, J. C., Richmond, V. P., & Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research, 1*, 323-332.

McFarland, L. A., Ryan, A. M., Sacco, J. M., & Kriska, S. D. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management, 30,* 435-452.

Murphy, K. R. (2002). Can conflicting perspectives on the role of *g* in personnel selection be resolved? *Human Performance, 15*(1/2), 173-186.

Nijsten, C. (1998). *Opvoeden in Turkse gezinnen in Nederland* [Raising in Turkish families in The Netherlands]. Assen, The Netherlands: Van Gorcum.

Ones, D. S., & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology, 75*(3), 255-276.

Pinto, D. (2004). *Beeldvorming en integratie. Is integratie het antwoord?* [Representation and integration. Is integration the answer?]. Houten, The Netherlands: Bohn Stafleu Van Loghum.

Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and gender similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology, 88*(5), 852-865.

Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizaitonal Psychology, 74*(5), 637-657.

Tajfel, H. (1982). *Social identity and intergroup relations*. London: Cambridge University Press.

Turban, D. B., Dougherty, T. W., & Lee, F. K. (2002). Gender, race, and perceived similarity effects in developmental relationships: The moderating role of relationship duration. *Journal of Vocational Behavior, 61*(2), 240-262.

Turner, J. C. (1987). *Rediscovering the social group: A self-categorization theory.* Oxford, Great Britain: Blackwell.

Van Leeuwen, A. E. (2000). *Constructie van de M5Q voor IWSP* [Construction of the M5Q for IWSP]. Culemborg, The Netherlands: Meurs Personeelsadvies.

Van Rijn, A. S., Zorlu, A., Bijl, R. V., & Bakker, B. F. M. (2004). *De ontwikkeling van een integratiekaart* [The development of an integration map]. The Hague, The Netherlands: WODC & CBS.

Verkuyten, M., Hagendoorn, L., & Masson, K. (1996). The ethnic hierarchy among majority and minority youth in The Netherlands. *Journal of Applied Social Psychology, 26*(12), 1104-1118.

Weijters, G., & Scheepers, P. (2003). Verschillen in sociale integratie tussen etnische groepen: beschrijving en verklaring [Differences between ethnic groups in social integration: Description and explanation]. *Mensch en Maatschappij, 78*(2), 144-157.

Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality* (pp. 88-162). New York: The Guilford Press.

Appendix 1

Dimensions, Dimension Descriptions, and Selection Tool Used

Dimension	Description	Instrument
Communication Skills	The ability to transmit information, ideas, and opinions, both verbally and non-verbally.	AC, Interview
Social Skills	The desire to have and begin social contacts, and to keep up these contacts.	AC, Interview
Empathy	The ability to put oneself in the thoughts, feelings, and reactions of others.	AC
Initiative	Taking or starting action of one's own accord, without incitement from outside, instead of waiting.	AC
Flexibility	Changing tasks fast and easily, being able to adapt to changing circumstances, and desiring changes and variation.	Interview
Emotional Stability	Being able to cope with emotional far-reaching situations.	Interview
Stress Tolerance	Being able to cope with high work- and time-pressure in daily work situations.	AC, Interview
Authority	Being able to influence others, both verbally and non-verbally, and being accepted as an authority by other people.	AC

Decisiveness	Being able and prepared to make decisions in dilemmas and with incomplete information, and taking responsibility for the consequences of these decisions.	AC
Tolerance Towards Others	Accepting and respecting differences between people, and taking these differences into consideration in one's own behavior.	Interview
Integrity	Being aware of the general acknowledged norms and values in society and showing willingness to act on these.	Interview
Self-Understanding	Being aware of one's own qualities and behavior, being able to reflect on qualities and behavior, and willing to integrate these reflections in future behavior.	Interview

Table 1

Sample Sizes of Each Applicant Type – Rater Type Combination

Applicant ethnicity	Demographic similarity				Perceived similarity	
	Assessors		Psychologists		Assessors	Psychologists
	Ethnic majority	Ethnic minority	Ethnic majority	Ethnic minority	Total	Total
Ethnic majority						
Applicant *n*	8,014	1,163	9,307	251	2,298	2,613
Rater *n*	70	5	71	2	15	12
Antillean						
Applicant *n*	75	14	90	1	24	25
Rater *n*	32	4	43	1	8	10
Moroccan						
Applicant *n*	154	15	173	1	33	45
Rater *n*	38	3	54	1	9	11
Surinamese						
Applicant *n*	183	33	221	8	52	46
Rater *n*	44	4	56	2	9	12
Turkish group						
Applicant *n*	379	55	427	11	102	117
Rater *n*	50	5	62	2	12	12

Table 2

Intra-Class Correlations (Proportions of Variance Due to Rater Differences)

	Demographic similarity	Perceived similarity
AC		
Agency	0.08	0.06
Communion	0.06	0.04
Employment Interview	0.17	0.16
Final Recommendation		
Agency	0.13	0.16
Communion	0.10	0.05
Socio-Cultural Awareness	0.17	0.11

Table 3

Demographic-Similarity Results

Applicant-group comparison		L1 parameter estimates		L2 parameter estimates	
		γ_{00} (*SE*)	γ_{10} (*SE*)	γ_{01} (*SE*)	γ_{11} (*SE*)
AC-Factor Agency					
1	Ethnic majority Dutch Antilleans	4.61** (0.13)	-0.43** (0.12)	-0.08* (0.03)	*ns*
2	Ethnic majority Moroccans	4.51** (0.09)	-0.34** (0.08)	-0.08* (0.03)	*ns*
3	Ethnic majority Surinamese	4.47** (0.09)	-0.30** (0.08)	-0.08* (0.03)	*ns*
4	Ethnic majority Turks	4.77** (0.07)	-0.59** (0.05)	-0.08* (0.03)	*ns*
AC-Factor Communion					
1	Ethnic majority Dutch Antilleans	4.36** (0.14)	-0.34* (0.13)	0.07 (0.04)	*ns*
2	Ethnic majority Moroccans	4.34** (0.11)	-0.32* (0.10)	0.07 (0.04)	*ns*
3	Ethnic majority Surinamese	4.24** (0.10)	-0.23* (0.09)	0.07 (0.04)	*ns*
4	Ethnic majority Turks	4.46** (0.08)	-0.44** (0.07)	0.07 (0.04)	*ns*
Employment Interview					
1	Ethnic majority Dutch Antilleans	4.49** (0.09)	-0.26** (0.07)	0.14* (0.04)	*ns*
2	Ethnic majority Moroccans	4.34** (0.07)	-0.12* (0.05)	0.14* (0.04)	*ns*
3	Ethnic majority Surinamese	4.43** (0.07)	-0.21** (0.05)	0.14** (0.04)	*ns*
4	Ethnic majority Turks	4.57** (0.06)	-0.33** (0.03)	0.13* (0.04)	*ns*
Final-Recommendation Factor Agency					
1	Ethnic majority Dutch Antilleans	4.54** (0.11)	-0.41** (0.09)	0.05 (0.06)	*ns*

2	Ethnic majority Moroccans	4.37** (0.09)	-0.24** (0.07)	0.05 (0.06)	*ns*
3	Ethnic majority Surinamese	4.44** (0.09)	-0.33** (0.06)	0.07 (0.06)	*ns*
4	Ethnic majority Turks	4.67** (0.07)	-0.54** (0.04)	0.04 (0.06)	*ns*
Final-Recommendation Factor Communion					
1	Ethnic majority Dutch Antilleans	4.41** (0.13)	-0.23* (0.11)	0.04 (0.07)	*ns*
2	Ethnic majority Moroccans	4.43** (0.11)	-0.25* (0.08)	0.04 (0.07)	*ns*
3	Ethnic majority Surinamese	4.40** (0.10)	-0.24** (0.07)	0.06 (0.07)	*ns*
4	Ethnic majority Turks	4.48** (0.09)	-0.31** (0.06)	0.04 (0.07)	*ns*
Final-Recommendation Factor Socio-Cultural Awareness					
1	Ethnic majority Dutch Antilleans	4.36** (0.07)	-0.10 (0.06)	*ns*	*ns*
2	Ethnic majority Moroccans	4.33** (0.07)	-0.08 (0.05)	*ns*	*ns*
3	Ethnic majority Surinamese	4.42** (0.06)	-0.18** (0.04)	0.01 (0.04)	*ns*
4	Ethnic majority Turks	4.45** (0.05)	-0.18** (0.03)	-0.01 (0.04)	*ns*

Note. A significant γ_{00} means that the intercept (grand mean) differs from zero. A negative γ_{10} means that ethnic minority applicants receive lower scores than majority applicants. A negative γ_{01} means that ethnic minority raters give lower scores than majority raters.

γ_{11} is the interaction of applicant and rater ethnicity, which is the focus regarding demographic similarity.

* $p < .05$ (two-tailed), ** $p < .001$ (two-tailed), *ns* means not significant.

Table 4

Perceived-Similarity Results

Applicant-group comparison		L1 parameter estimates		L2 parameter estimates	
		γ_{00} (*SE*)	γ_{10} (*SE*)	γ_{01} (*SE*)	γ_{11} (*SE*)
AC-Factor Agency					
1	Ethnic majority Dutch Antilleans	4.96** (0.23)	-0.49* (0.21)	-0.07* (0.02)	*ns*
2	Ethnic majority Moroccans	4.36** (0.19)	-0.18 (0.19)	*ns*	*ns*
3	Ethnic majority Surinamese	4.40** (0.15)	-0.22 (0.14)	*ns*	*ns*
4	Ethnic majority Turks	5.20** (0.15)	-0.61** (0.10)	-.11** (0.03)	*ns*
AC-Factor Communion					
1	Ethnic majority Dutch Antilleans	4.63** (0.26)	-0.35 (0.26)	*ns*	*ns*
2	Ethnic majority Moroccans	4.58** (0.22)	-0.30 (0.21)	*ns*	*ns*
3	Ethnic majority Surinamese	4.59** (0.19)	-0.31 (0.18)	*ns*	*ns*
4	Ethnic majority Turks	5.16** (0.18)	-0.64** (0.13)	-0.07* (0.03)	*ns*
Employment Interview					
1	Ethnic majority Dutch Antilleans	4.59** (0.14)	-0.20 (0.14)	*ns*	*ns*
2	Ethnic majority Moroccans	3.64** (0.14)	-0.24* (0.11)	0.36** (0.03)	*ns*
3	Ethnic majority Surinamese	3.68** (0.16)	-0.32* (0.12)	0.33** (0.03)	*ns*
4	Ethnic majority Turks	3.66** (0.12)	-0.40** (0.07)	0.39** (0.03)	*ns*
Final-Recommendation Factor Agency					
1	Ethnic majority Dutch Antilleans	4.12** (0.21)	-0.35* (0.17)	0.15** (0.04)	*ns*

2	Ethnic majority Moroccans	4.36** (0.14)	-0.13 (0.14)	*ns*	*ns*
3	Ethnic majority Surinamese	4.12** (0.20)	-0.36* (0.15)	0.15** (0.04)	*ns*
4	Ethnic majority Turks	4.26** (0.15)	-0.56** (0.09)	0.18** (0.04)	*ns*
Final-Recommendation Factor Communion					
1	Ethnic majority Dutch Antilleans	4.34** (0.21)	-0.01 (0.21)	*ns*	*ns*
2	Ethnic majority Moroccans	4.52** (0.16)	-0.19 (0.16)	*ns*	*ns*
3	Ethnic majority Surinamese	3.91** (0.23)	-0.35* (0.17)	0.24** (0.05)	*ns*
4	Ethnic majority Turks	4.00** (0.18)	-0.37* (0.12)	0.24** (0.05)	*ns*
Final-Recommendation Factor Socio-Cultural Awareness					
1	Ethnic majority Dutch Antilleans	4.42** (0.12)	-0.11 (0.12)	*ns*	*ns*
2	Ethnic majority Moroccans	3.77** (0.13)	-0.24* (0.11)	0.28** (0.03)	*ns*
3	Ethnic majority Surinamese	4.44** (0.11)	-0.13 (0.11)	*ns*	*ns*
4	Ethnic majority Turks	4.55** (0.42)	-1.09* (0.40)	0.29** (0.03)	0.32* (0.13)

Note. A significant γ_{00} means that the intercept (grand mean) differs from zero. A negative γ_{10} means that ethnic minority applicants receive lower scores than majority applicants. A negative γ_{01} means that raters who perceive a certain ethnic minority group as more similar to themselves give lower scores than raters who perceive a this ethnic minority group as less similar.

γ_{11} is the interaction of applicant ethnicity and perceived similarity toward the applicant's ethnic group, which is the focus regarding perceived similarity.

* $p < .05$ (two-tailed), ** $p < .001$ (two-tailed), *ns* means not significant.

Table 5

Mean Ranks in the Integration Hierarchy as Viewed by Raters

Ethnic minority group	Mean Rank
Surinamese	3.60
Dutch Antilleans	2.80
Turks	2.40
Moroccans	1.20

Note. The higher the mean rank, the more the ethnic minority group is viewed as being integrated into Dutch society.