

HUMAN PERFORMANCE, 19(3), 219–251
Copyright © 2006, Lawrence Erlbaum Associates, Inc.

Applicant and Method Factors Related to Ethnic Score Differences in Personnel Selection: A Study at the Dutch Police

Lonneke A. L. de Meijer and Marise Ph. Born
Erasmus University Rotterdam, The Netherlands

Gert Terlow
Police Academy of The Netherlands, Apeldoorn, The Netherlands

Henk T. van der Molen
Erasmus University Rotterdam, The Netherlands

The aim of this study was to examine applicant and method factors related to ethnic score differences on a cognitive ability test, a personality test, an assessment center (AC), an employment interview, and a final employment recommendation in the context of police officer selection ($N = 13,526$). Score differences between the majority group and the first-generation minority groups were comparable to research findings from the literature. However, score differences between the majority group and second-generation minority groups were much smaller. On the cognitive ability test and the personality test most variability was explained by Dutch language-proficiency. Confirming assumed-characteristics theory, more variability on the interview and the employment recommendation was explained by Dutch language-proficiency and education than on the AC. Unsupportive of complexity–extremity theory, there seemed to be a general tendency to give lower scores to the ethnic minority group.

The personnel selection literature has extensively investigated differences on psychological measures between ethnic minority and majority groups. This study focuses on ethnicity-related applicant demographics, such as language-proficiency and education, and their interplay with selection-method factors in their impact on test scores. Personnel selection measures can differ in the extent of assessor influ-

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

ence in the selection process. Those selection measures where there is no assessor influence (e.g., cognitive ability and personality tests) are labeled as *objective measures*. By contrast, measures characterized by the involvement of an assessor (such as the assessment center and the employment interview) are labeled *subjective measures* (Bass & Barrett, 1981).

This study focuses on objective as well as subjective selection measures, with special attention paid to the subjective measures. First, an overview will be given of the literature findings stemming from North America and Europe on score differences between ethnic groups on objective measures, which will then be followed by a discussion on subjective measures and also the final employment recommendation to hire or not. Second, two theoretical perspectives developed within social psychology will be described. These perspectives concern the impact of perceptions of groups on evaluators' ratings, defined according to the ethnicity-related demographic characteristics of these groups. These are assumed-characteristics theory (Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982a, 1982b) and complexity-extremity theory (Linville, 1982; Linville & Jones, 1980). Hypotheses will then be derived and tested in the context of the Dutch police officer selection procedure over the last couple of years.

OBJECTIVE MEASURES

Cognitive Ability Tests

General cognitive ability, or *g*, has been found to be a consistent predictor of job performance across a variety of occupations (Schmidt & Hunter, 1998, 2004). This is especially the case for more complex job levels (Hunter, 1986). At the same time, several researchers (e.g., Goldstein, Zedeck, & Goldstein, 2002; Murphy, 2002; Outz, 2002) have shown that the cognitive ability test represents the predictor most likely to have substantial adverse impact on employment opportunities for most ethnic minority groups. Ethnic score differences between $.50 SD$ and $1.50 SD$ on cognitive ability tests have often been found (e.g., Herrnstein & Murray, 1994). However, evidence has been found that ethnic differences in cognitive ability test scores are considerably larger than ethnic differences in measures of job performance (Hattrup, Rock, & Scalia, 1997; Waldman & Avolio, 1991).

Striving for a fully ethnicity-proof cognitive ability test has turned out to be a utopia. In such endeavors, a distinction has been made between tests more influenced and tests less influenced by cultural aspects. Cattell (1987) made a well-known distinction between "fluid" intelligence and "crystallized" intelligence. Fluid intelligence relates to basic reasoning, which is necessary for problem solving, is dependent on neuronal efficiency, and is very hereditary (Bors & Forrin, 1995; Horn & Noll, 1997; Jensen, 1993; Plomin, 1988). Crystallized intelligence

can be seen as a result of the action of fluid intelligence on a certain (cultural) environment, which is dependent on one's learning experience and on the perceived importance of certain abilities. Thus, crystallized intelligence can have different forms in different cultures. Most cognitive ability tests appeal to basic cultural knowledge and abilities (e.g., instructions and items written are in a certain language or the tests appeal to scholastic abilities [Van den Berg & Van Leest, 1999]). For Western ethnic majority group members with a comparable cultural and scholastic background, the appeal to basic cultural knowledge and abilities is not a problem. The required knowledge and abilities are "overlearned" and, therefore, these group members have the basic knowledge and ability that is required in almost every situation. However, for ethnic minority members, the appeal to scholastic abilities for cognitive tests may indeed be a problem. Research in The Netherlands by Bleichrodt and Van den Berg (1995) has shown it is not so much the period of residence in The Netherlands that impacts on cognitive ability test scores in general and crystallized intelligence in specific, but much more the age of immigration.

Personality Tests

The use of personality tests to assess ethnic minority group members has been criticized as well. Though personality tests are generally of adequate reliability and validity in different ethnic groups (Anderson & Ones, 2003), critics assume they are of limited use for assessing individuals in a certain country or area who have a limited knowledge of the spoken language and culture of that area (Te Nijenhuis, 1997). Yet, less research than on cognitive ability tests has been done to answer the question whether different ethnic groups exhibit different scores on personality tests, and what has been done has found mixed results. Hough (1998) in the United States, and Ones and Anderson (2002) in the United Kingdom reported ethnic group differences in the negligible to moderate range. In The Netherlands, ethnic score differences on personality tests have been investigated by Van Leest (1997) and Te Nijenhuis, Van der Flier, and Van Leeuwen (1997). Van Leest (1997) investigated ethnic score differences between majority group members and Turks. He found a mean difference of $-.29 SD$, ranging between $-1.34 SD$ for routine in methods (i.e., avoidance of uncertainty) and $1.04 SD$ for assertiveness, where positive values indicated the ethnic majority group scoring higher. Te Nijenhuis et al. (1997) found significantly higher mean scores for ethnic minorities on neuroticism, varying between $-.79 SD$ and $-.43 SD$, for neurosomatics, ranging between $-1.19 SD$ and $-.28 SD$, and for social conformity, ranging from $-.78 SD$ to $-.52 SD$, where negative d values indicated ethnic minorities scoring higher. Lower mean scores for ethnic minorities were found for Extraversion, differing between $.05 SD$ and $.50 SD$ (where positive values indicated the ethnic majority group scoring higher). The differences between the ethnic majority and minority groups were

larger for Turks and Moroccans than for Surinamese and Caribbean groups. As regular jobs in general require emotional stability, sociability, and flexibility, utilizing personality test findings would imply lower hiring chances for many positions for ethnic minority groups (Te Nijenhuis et al., 1997).

Differences between ethnic groups on personality test scores seem to be much smaller than differences on cognitive ability tests. However, the reported differences on personality tests seem unsystematic and therefore difficult to interpret. Reported ethnic group differences to the disadvantage of ethnic minorities are larger in The Netherlands than in the United States and the United Kingdom. Relatively little has been published that provides explanations for such differences in findings.

When employers want to maximize the skill level of their employees on the one hand and diversify their workforce on the other hand, they are saddled with a dilemma. Both goals cannot be achieved simultaneously, because of existing subgroup differences in the results on objective measures. One solution has been sought in the use of face-valid simulations as selection tools to evaluate both cognitive and noncognitive job-relevant knowledge, skills, and abilities (Schmitt & Mills, 2001). We now turn to several of these tools, which contain a subjective evaluative element by an assessor.

SUBJECTIVE MEASURES

The Assessment Center (AC)

ACs are mostly used for the selection of higher-level managerial jobs (Cascio, 1991; Heneman & Heneman, 1994). Past research has indicated that score differences between ethnic groups on the AC vary between $.02 SD$ and $.58 SD$ to the advantage of the ethnic majority group (e.g., Goldstein, Yusko, Braverman, Smith, & Chung, 1998; Goldstein, Yusko, & Nicolopoulos, 2001). Findings until now have been mostly restricted to highly complex jobs and to specific North American ethnic groups (Cascio, 1991; Heneman & Heneman, 1994).

The Employment Interview

The employment interview is probably the most commonly used selection tool (Huffcutt & Roth, 1998). Likewise, there has been a substantial amount of research examining ethnic score differences in the employment interview. Findings until now indicate that score differences between Blacks and Whites vary between $.14 SD$ and $.56 SD$ in favor of the ethnic majority group (Huffcutt & Roth, 1998; Motowidlo et al., 1992; Roth, Van Iddekinge, Huffcutt, Eidson, & Bobko, 2002). In The Netherlands, Van den Berg (2001) found a difference of $.23 SD$ between

ethnic minorities and the ethnic majority, in favor of the majority group. Van den Berg reported that an important part of the variability in evaluations could be explained by language-proficiency of the applicant.

The Final Employment Recommendation

Predictor information of several different selection measures needs to be combined to form a final employment recommendation of an applicant. Predictor information can be combined either mechanically (mechanical prediction) or judgmentally (clinical prediction). *Clinical prediction* refers to a procedure in which a judge puts data together using informal, subjective methods (Grove, Zald, Lebow, Snitz, & Nelson, 2000). *Mechanical prediction* refers to statistical prediction without the interference of a human evaluator (Grove et al., 2000). Though most final selection decisions are achieved through clinical prediction, it is found in numerous studies (for a review, see Grove et al., 2000) that mechanical prediction is either equal to or superior to clinical judgment. Why does a mechanical combination of data yield better results than a judgmental combination? One possible explanation is that decision makers are more likely to add considerable error if they are allowed to judgmentally combine both subjective data (e.g., ACs or interviews) along with objective data (e.g., scores on the cognitive ability test or personality test). Their perceptions of an applicant may influence their evaluations and ultimately their decisions to select or reject an applicant (Bass & Barrett, 1981). Because of the existing ethnic score differences on both objective and subjective measures, the combination of these measures into a final employment recommendation is likely to yield ethnic differences as well. However, to our knowledge no research specifically directed to this issue has been done until now.

In sum, a review of the literature indicates that ethnic score differences are found on all selection measures discussed and are mostly in favor of the ethnic majority group. Differences on the cognitive ability test seem to be the largest and most consistent throughout all studies (between $.50 SD$ and $1.50 SD$). Differences that were found on the AC (between $.02 SD$ and $.58 SD$) and on the employment interview (between $.14 SD$ and $.56 SD$) are in favor of the ethnic majority group as well, though these are smaller than on the cognitive ability test. Research on personality tests has also found differences between ethnic groups, but the results are mixed. Little attention has been given to why differences exist between ethnic groups on noncognitive measures as well as why differences exist between selection tools. Finally, to our knowledge, no research has been done on final employment recommendations in which scores on various measures are combined.

In this article ethnic score differences on the cognitive ability test, the personality test, the AC, the employment interview, and the employment recommendation are investigated. Furthermore, possible explanations for score differences between ethnic groups are searched. Before deriving hypotheses about ethnic score differ-

ences on objective and subjective measures, we first want to focus on two theoretical perspectives from social psychology pertaining to the subjective measures.

THEORETICAL PERSPECTIVES FROM SOCIAL PSYCHOLOGY

Two theoretical perspectives that address the influence of demographic information of individuals on evaluations by others are so-called assumed-characteristics theory (Locksley et al., 1980; Locksley et al., 1982a, 1982b) and complexity–extremity theory (Linville, 1982; Linville & Jones, 1980). Coleman, Jussim, and Kelley (1995); Jussim, Coleman, and Lerch (1987); and Jussim, Fleming, Coleman, and Kohberger (1996) have investigated assumed-characteristics theory and complexity–extremity theory in laboratory experiments. This study investigates the applicability of these two theories in a field setting. Both theories propose different processes to explain how ethnicity-related demographic information about individuals may influence evaluations by others. Assumed-characteristics theory, complexity–extremity theory, and their possible effects on applicant evaluations in the AC, the employment interview, and the employment recommendation are discussed next.

Assumed-characteristics theory suggests that, based on knowledge about certain demographics of a group, people make assumptions about other characteristics of this group. For example, on the basis of knowledge of a group's ethnicity, people *assume* that this group will have a certain socioeconomic status (SES), education, or personality. Assumed-characteristics theory suggests that members of an in-group will believe their own characteristics are more favorable than characteristics of members of an out-group (Coleman et al., 1995; Jussim et al., 1987, 1996). Having more relevant ethnicity-related demographic information about out-group members should nevertheless substantially decrease the unfavorable assumed characteristics and evaluations of out-group members should become more positive. In other words, this theory supposes that the new ethnicity-related demographic information of the applicant is more positive than the assumed characteristics on the basis of ethnicity. When in-group members have relevant ethnicity-related demographic information about out-group members (e.g., information about someone's education and language-proficiency) this information should diminish the negative group membership effects on the basis of ethnicity (Jussim, 1990, 1991, 1993; Locksley et al., 1980; Rokeach & Mezel, 1966; for a review see Swim, Borgida, Maruyama, & Myers, 1989). Assumed-characteristics theory would lead to the following expectations: Even if some factors, such as the knowledge of someone's ethnicity, lead to an evaluation in favor of individuals belonging to one group, other relevant ethnicity-related demographic information—information other than someone's ethnicity, for example, someone's education or

language-proficiency—should be more influential than group membership on the basis of ethnicity (Coleman et al., 1995; Jussim et al., 1987, 1996). Furthermore, if ethnicity-related demographic information, such as education and language-proficiency, is not available, factors such as ethnicity and the negative out-group membership effects of them will have a stronger influence on evaluations of individual assessors.

Complexity–extremity theory (Linville, 1982; Linville & Jones, 1980) starts with the assumption that people have more contact with in-group members than with out-group members. Because in-group members have more contact with other in-group members, they will develop more complex representations of in-group members than of out-group members. When observers use these more complex representations to evaluate an in-group member, they are likely to give accurate evaluations. When complex representations are developed, an observer has knowledge about both good and bad characteristics about the in-group member, which renders an extreme evaluation unlikely. Low complex or simple representations will be developed about out-group members. When these simple representations are used in evaluating an out-group member, extreme evaluations are more likely because the out-group member can more easily be seen as all good or all bad. Thus, a complex representation of someone will lead to less chance of extremity in evaluations, and a simple representation leads to a higher chance of extremity in evaluations (Coleman et al., 1995; Jussim et al., 1987, 1996). Complexity–extremity theory would lead to the following judgmental outcomes: (a) an out-group member whose demographics can be seen as positive (e.g., high education) will be evaluated extremely favorable—even more favorable than an in-group member with those same demographics (Coleman et al., 1995; Jussim et al., 1987, 1996); and (b) an out-group member whose demographics can be seen as negative (e.g., low education) will be evaluated extremely unfavorable—even more unfavorable than an in-group member with those same demographics (Coleman et al., 1995; Jussim et al., 1987, 1996).

We employ assumed-characteristics theory and complexity–extremity theory as potential explanatory frameworks for assessors' evaluations on the subjective measures (i.e., the dimensions that are assessed in the AC, the employment interview, and the employment recommendation).

Having now discussed the literature available on ethnic score differences on various selection tools and some possible explanations for these differences, hypotheses are formulated in the following section.

OVERVIEW OF HYPOTHESES

The hypotheses may be divided in three groups. First, consistent with findings from the literature, it is expected that differences between ethnic minorities and

ethnic majorities to the advantage of the majority group exist on all selection tools (i.e., the cognitive ability test, the personality test, the AC, the employment interview, and the employment recommendation; H1a). Furthermore, based on the literature it is expected that ethnic score differences will be largest on the cognitive ability test and lowest on the personality test (H1b). Research by Bleichrodt and Van den Berg (1995) has shown that first-generation ethnic minority group members who moved to The Netherlands from countries such as Turkey, Morocco, Surinam, or the Caribbean before the age of seven (before starting their primary education) score significantly higher on cognitive ability tests than first-generation ethnic minorities who moved to The Netherlands after the age of seven. The scores of first-generation ethnic minority group members who moved to The Netherlands before the age of seven were still lower, though, than scores of ethnic majority group members (Bleichrodt & Van den Berg, 1995). Second-generation minority group members, in contrast to first-generation minorities, are born in The Netherlands and, therefore, will have passed through the Dutch educational system. First-generation ethnic minority members are born outside The Netherlands. Large numbers of first-generation minorities did not receive their education in The Netherlands. Second-generation ethnic minorities do not only differ from first-generation ethnic minorities in terms of education. Second-generation ethnic minorities, because they are born in The Netherlands, are also confronted with the Dutch culture and its norm and values to a larger extent than first-generation ethnic minorities. However, most of the second-generation ethnic minority families still speak their native language at home and, to some extent, have their own customs. Therefore, second-generation minorities are not yet fully integrated into Dutch society and they still differ from the Dutch majority (Weijters & Scheepers, 2003). Extending the findings of Bleichrodt and Van den Berg (1995), it is expected that score differences between first- and second-generation ethnic minority groups to the advantage of the second-generation minority group exist on all selection tools (H1c). In line with findings from Bleichrodt and Van den Berg (1995) on differences in crystallized intelligence and fluid intelligence, it is expected that on subtests measuring crystallized intelligence (Cattell, 1987) the differences between first- and second-generation minority groups will be larger than on subtests measuring fluid intelligence (H1d).

A second and a third group of hypotheses is addressed next. A lot of research has been done to explain differences on cognitive ability tests between ethnic groups in North America. There has often been a tendency in existing research to treat ethnic minorities as a homogeneous group that merely contrasts with the ethnic majority group. That is, a dichotomous distinction is made between Whites and non-Whites or between the majority and the minority group. This approach ignores the many visible and cultural differences between ethnic groups that may affect scores on selection instruments. The main ethnic minority groups in North America are Blacks, Hispanics and Latinos, and Asians. These American eth-

nic-minority groups moved to North America generations ago, whereas in Europe ethnic-minority groups only moved to European countries from the 1960s onward. Therefore, first- and second-generation ethnic minority groups are at the center of attention in European research on ethnic group differences. Because of the difference between the length of residence of ethnic minorities in North America and in Europe, the language-proficiency of ethnic minority samples in North American research is probably better than the language-proficiency of ethnic minority samples in European research. Explanations for ethnic cognitive ability differences in North America are often searched in the context of SES and background characteristics, whereas in The Netherlands, where most research focuses on Caribbean, Moroccan, Surinamese, and Turkish minority groups, group differences are sought in Dutch language-proficiency and being a first- or second-generation minority (Bleichrodt & Van den Berg, 1995).

The second group of hypotheses is aimed at investigating to what extent ethnic score differences on the objective and subjective measures can be explained by the following applicant demographics: (a) Dutch language-proficiency, (b) education, and (c) ethnicity. It is hypothesized that as the objective tests (i.e., the cognitive ability test and the personality test) are written in Dutch, Dutch language-proficiency will explain more of the variability between ethnic groups than education and ethnicity (H2a).

H2b and H2c are derived from assumed-characteristics theory (Coleman et al., 1995; Jussim et al., 1987; Jussim et al., 1996). The hypotheses are aimed at investigating to what extent ethnic score differences on the employment interview and the employment recommendation on the one hand and the AC on the other hand can be explained by the following applicant ethnicity-related demographics: (a) Dutch language-proficiency, (b) education, and (c) ethnicity. In the employment interview and the final recommendation, the assessor has knowledge of the applicant's language-proficiency, education, and ethnicity. In the AC, no such knowledge is given to the assessors. The reason why assessors do have knowledge about demographic information of the applicant during the interview and the employment recommendation and assessors do not have this knowledge during the AC is that interviewers also write the final recommendation and all information about a certain applicant is at the interviewers' disposal.

From assumed-characteristics theory it is hypothesized that for the employment interview and the final recommendation applicant ethnicity-related demographics, namely Dutch language-proficiency and education, will explain more of the variability in assessors' evaluations than ethnicity itself (H2b). Furthermore, it is hypothesized that for the AC Dutch language-proficiency and education will not explain more of the variability in assessors' evaluations but as much as or less than ethnicity (H2c).

The third group of hypotheses is derived from complexity-extremity theory (Coleman et al., 1995; Jussim et al., 1987; Jussim et al., 1996). It is hypothesized

that ethnic majority assessors will evaluate ethnic minority applicants with an excellent Dutch language-proficiency and education higher on the interview and the employment recommendation than ethnic majority applicants with the same excellent Dutch-language-proficiency and education (H3a); and that ethnic majority assessors will evaluate ethnic minority applicants with a low Dutch language-proficiency and education lower on the interview and the employment recommendation than ethnic majority applicants with the same low Dutch-language-proficiency and education (H3b).

METHOD

Participants and Procedure

Data came from a first-generation minority group, a second-generation minority group, and a majority group, who applied for a position at the Police Academy of The Netherlands from September 2001 until July 2003. The largest first- and second-generation ethnic minority groups are from the Caribbean, from Morocco, Surinam, and Turkey. The dataset consisted of 11,432 applicants. Data of 11,409 applicants were used, of which 672 applicants were first-generation ethnic minorities and 734 applicants were second-generation ethnic minorities. Data of 23 applicants were incomplete. These cases were removed from the dataset. The professions for which accepted students were to be trained for were assistant police employee, police employee, or all-round police employee. Applicants who were interested in a job as police officer first applied to the local police force where they wanted to work after they would complete 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). During a requirement check at the CCM, the following minimal criteria are checked on the basis of an application form: minimal age (16 years), Dutch nationality (first or second), possession of a swimming diploma, no criminal record, possession of a school diploma (minimal level is preparatory vocational education level B [VBO-B]). Applicants in the selection process went through two stages. During the first stage a Dutch language-proficiency test was filled out. During the second phase a physical exercise, a cognitive ability test, a personality test, an AC assignment, and an employment interview were executed. The psychologist who conducts the interview is also the one who writes the final employment recommendation to the police force. For the employment recommendation, the test results of the personality test, the AC ratings, and the employment interview ratings are used. Next to the final recommendation, the final dossier to the local police forces exists of test scores of the physical exercise, the cognitive ability test, and the language-proficiency test (for an overview of the selection procedure see Figure 1).

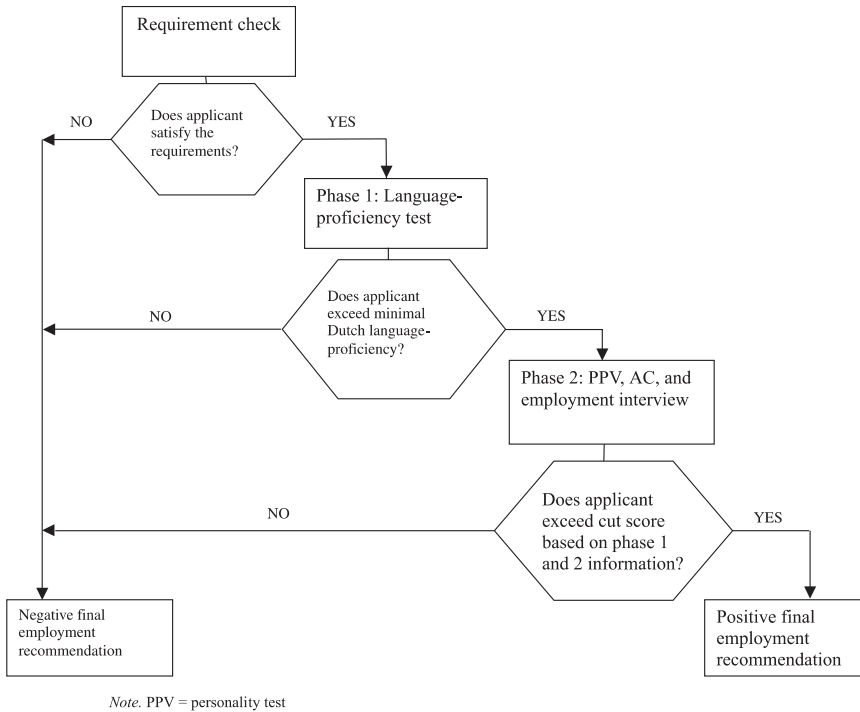


FIGURE 1 Selection process at the Center for Competence Assessment and Monitoring.
 Note. PPV = personality test.

Table 1 shows the distributions of the groups in terms of demographic variables. The ethnic minority group from countries classified as *other* ($N = 235$, 2% of total group) consists mostly of people from Eastern and Western Europe and Western non-European (e.g., United States, Canada, and Australia) countries (72% of *others*), but also from Asia (with the exception of Turkey, Japan, and Dutch India), Africa (with the exception of Morocco), South America (with the exception of Surinam), and Central America (with the exception of the Caribbean). In view of the heterogeneity of this group, its data are only used to test hypotheses 2 (a–c) and 3 (a–b). To test these hypotheses the ethnic minority groups are taken together.

Within all ethnic groups the largest number of applicants were male, especially within the first-generation minority group (mean % male = 75). Within the majority group 66% were male. The mean age of the applicants of the first-generation minority group ($M = 28.00$; $SD = 7.05$) was higher than the mean age of the second-generation minority group ($M = 21.85$; $SD = 4.57$; $t = 19.21$, $p < .05$) and of the majority group ($M = 23.92$; $SD = 7.11$; $t = 14.49$, $p < .05$). The largest percentage of applicants was within the majority group (88%). Six percent of the applicants were

TABLE 1
Distribution of Majority Group Members and First- and Second-Generation Minority Group Members
in Terms of Gender, and Age

	<i>n</i>	<i>%</i>	<i>%</i> <i>Male</i>	<i>M</i> <i>Age</i>	<i>SD</i>	<i>1st Generation Minority Group</i>					<i>2nd Generation Minority Group</i>				
						<i>n</i>	<i>%</i>	<i>%</i> <i>Male</i>	<i>M</i> <i>Age</i>	<i>SD</i>	<i>n</i>	<i>%</i>	<i>%</i> <i>Male</i>	<i>M</i> <i>Age</i>	<i>SD</i>
Position															
1st generation minority group	672	5.9	75	28.00	7.05										
2nd generation minority group	734	6.4	72	21.85	4.57										
Majority group	10,003	87.5	66	23.92	7.11										
Total	11,432	100	67	24.03	7.06										
Ethnicity															
Dutch Antilles						75	0.7	75	29.58	6.97	33	0.3	73	21.78	4.72
Morocco						90	0.8	74	27.57	6.74	110	1.0	71	21.16	3.28
Surinam						129	1.1	81	28.93	6.86	123	1.1	69	21.95	5.08
Turkey						162	1.4	74	27.54	6.85	334	2.9	75	21.23	3.25
Other						216	1.9	72	27.41	7.38	109	1.0	66	23.63	6.31

Note. Of 24 (0.2%) applicants it was not known if they were first-generation ethnic minority, second-generation ethnic minority, or majority group members and the ethnicity of 25 applicants (0.2%) could not be determined.

first-generation minority members and 6% were second-generation applicants. For 24 applicants it was not known if they were first-generation ethnic minority, second-generation ethnic minority, or majority group members.

All assessors (82 conducting the interview and the employment recommendation and 116 conducting the AC) in the selection process of the CCM had a background as vocational advisor or psychologist. Eighty-six percent of the assessors in the interview and the final recommendation were female and 78% of the assessors in the AC were female. Nearly all assessors were majority group members and all had a high educational level (higher professional education [HBO], or research-oriented education [WO]).

Measures

Cognitive ability. The Police Intelligence Test (PIT; Rijks Psychologische Dienst, 1975) is a cognitive ability test and consists of 107 items divided over 6 subtests: Analogies (verbal comprehension), Arranging Pictures (picture arrangement), Series of Numbers (numerical reasoning), Silent Reading (word fluency), Folding Figures (spatial ability), and Series of Figures (inductive reasoning). The time limit is 51 minutes. Applicants completed the PIT in Dutch. Prior research by Lem and Van Doorn (2000) indicated alpha reliabilities varying from .69 for Series of Numbers, to .87 for Folding Figures. The correlations between the subscales varied from .32 to .57. A study by Van der Maesen (1992) showed corrected predictive validity coefficients of .39 and .46 ($N = 162$).

Personality. 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 Dutch. A recent progress report by Klinkenberg and Van Leeuwen (2003) indicated alpha reliabilities varying from .72 for Conscientiousness, to .78 for Intellect. Correlations between the scales are all lower than .60. Comparison with 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 ($N = 61$).

The AC. A role-play exercise is utilized, in which an assessor and an actor independently make 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 & Trapnell, 1996), which together explained 77% of

the variance. As a measure of Agency, the average rating across the dimensions of Authority, Decisiveness, Initiative, Communication Skills, and Stress Tolerance was used ($\bar{r} = .59$; $\alpha = .87$). As a measure of Communion, the average rating of the dimensions Social Skills and Empathy was used ($\bar{r} = .77$; $\alpha = .87$). The reliability of the difference (r_{diff}) between scores on Agency and Communion was .78.

The employment interview. The interview questions are 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 interviewer conducts the interview. The interviews are 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 ($\bar{r} = .42$; $\alpha = .85$). Moreover, principal component analysis with varimax rotation yielded one interview factor that explained 50% of the variance.

The final employment recommendation. The final recommendation as to whether an applicant is fit for a job as police officer is based on results from the personality test (PPV), the AC, and the employment interview. These scores are integrated into an employment recommendation. The dimensions in the final recommendation are Communication Skills, Social Skills, Empathy, Initiative, Flexibility, Stress Tolerance, Authority, Decisiveness, Tolerance Towards Others, Integrity, and Self-Understanding. A 7-point Likert scale ranging from 1 (*extremely weak*) to 7 (*excellent*) is used to evaluate the behavior on the 11 dimensions. Principal component analysis with varimax rotation yielded three employment-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 ($\bar{r} = .48$; $\alpha = .85$). As a measure of Communion, the dimensions Social Skills and Empathy, were used ($\bar{r} = .66$; $\alpha = .79$) and for Socio-Cultural Awareness the dimensions ($\bar{r} = .39$; $\alpha = .65$), Tolerance Towards Others, Integrity, and Self-Understanding. The reliability of the difference (r_{diff}) between scores on Agency and Communion is .51, r_{diff} between scores on Agency and Socio-Cultural Awareness is .58, and r_{diff} between scores on Communion and Socio-Cultural Awareness is .57.

Analyses

First group of hypotheses. Results from preliminary analyses showed that all measures were found to be structural equivalent (for detailed information, please contact the first author). Levene's tests for equality of variances and *t* tests

for equality of means were conducted to index ethnic group differences on the various selection measures (H1a). Following Ones and Anderson (2002), standardized effect sizes (d values) between the means of the various groups of interest were computed to get an indication of the magnitude of the group differences on the various selection instruments irrespective of sample size (H1b–d). D values index the standardized mean differences between any two groups being compared (Cohen, 1988). Positive d values indicate higher mean scores for the majority group and negative d values indicate higher mean scores of a minority group (Caribbean, Moroccan, Surinamese, or Turkish group). Though effect sizes can theoretically range between positive and negative infinity, Cohen (1988) suggests that effect sizes of about .20 in magnitude are small, around .50 are medium, and above .80 are large. To conduct Levene's tests and t tests and to compute d values, observed differences on dimensions scores were used that were uncorrected for age, gender, and education. Corrected d values only differed marginally (about .01 SD) from uncorrected d values.

Second group of hypotheses. Structural equation modeling (SEM) with Amos 5.0 (Arbuckle, 2003) was used to investigate to what extent score differences on the cognitive ability test, the personality test (the objective measures), the AC, the employment interview, and the final recommendation (the subjective measures) between ethnic groups could be explained by a number of factors (H2a–c). These factors are (a) Dutch language-proficiency, (b) education, and (c) ethnicity. All factors are ordinal variables. An ordinal conception of ethnicity manifests itself in a hierarchy of ethnic groups in terms of social distance from the Dutch majority (Hraba, Hagendoorn, & Hagendoorn, 1989). Several studies have found consensus on the hierarchy of ethnic groups in The Netherlands (e.g., Hraba et al., 1989; Verkuyten, Hagendoorn, & Masson, 1996) where European groups were placed on top, followed by colonial and then Islamic groups at the bottom. More specifically, the following hierarchy is used (Hraba et al., 1989): (a) Dutch majority, (b) Western ethnic minority (which includes people from Western and Eastern Europe, and Western non-European countries), (c) Caribbean, (d) Surinam, (e) Morocco, and (f) Turkey.

Because the factors Dutch language-proficiency and ethnicity had a moderate intercorrelation ($r = .37$; education and ethnicity, and language-proficiency and education did not correlate), a general model was created which took the intercorrelation between Dutch language-proficiency and education into account (see Figure 2). For measuring Dutch language-proficiency, a Dutch language test (IBO; Bureau Interculturele Evaluatie, 2000) was used that had previously turned out to be very useful in the practice of educational institutes.

With regard to H2a, the effects of Dutch language-proficiency, education, and ethnicity were examined on cognitive ability. To examine this a specific model was created where g -loaded subtests—subtests that measure fluid intelligence—were

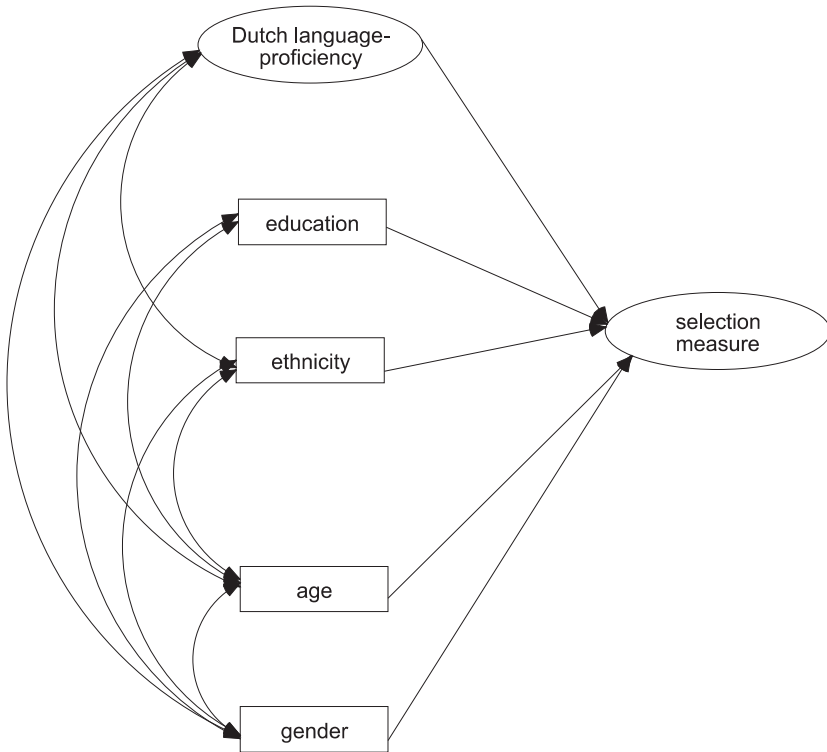


FIGURE 2 Path model to test the explanatory power of Dutch language-proficiency, education, and ethnicity. *Note.* Selection measures are the cognitive ability test, the personality test, the AC, the employment interview, and the final employment recommendation. Age and gender are control variables.

used as control variables. This was done because Dutch language proficiency was intercorrelated with the *g*-loaded cognitive ability subtests ($r = .67$).

Third group of hypotheses. To test complexity–extremity theory (H3a–b), Levene’s tests for equality of variances and *t* tests for equality of means were conducted. Also, standardized effect sizes (*d* values) were calculated to get an indication of the magnitude of the group differences. Positive *d* values indicate higher mean scores for the majority group. Because of the small first- and second-generation sample sizes operationalized as “high” and “low,” comparisons were only made between the ethnic majority group and the undifferentiated ethnic minority group of applicants with a high Dutch-language-proficiency and education and applicants with a low Dutch language-proficiency and education. The AC was not used for testing complexity–extremity theory because in the AC, as said earlier, no information was given to the assessors on Dutch lan-

guage-proficiency and education. Age, gender, and cognitive ability were used as control variables because the aim, here, was not directed at age, gender, and cognitive ability differences.

RESULTS

First Group of Hypotheses

The results relevant to the H1 (a–d) are presented in Table 2. Consistent with the findings from the literature, significant score differences between the ethnic majority group and ethnic minority groups to the advantage of the majority group, existed on all selection tools (H1a). The only exception was the personality test (PPV) dimension Conscientiousness, where minority groups systematically scored higher than the majority group.

In accordance with H1b, score differences between the ethnic majority group and the ethnic minority groups were largest on the cognitive ability test (PIT) and lowest on the personality test (PPV). Score differences on the PIT varied from d values of .06 SD ($t = .38, ns$) on Spatial Ability (PIT), to 1.30 SD ($t = 12.30, p < .001$) on Inductive Reasoning (PIT). Score differences on the PPV ranged between $-.49 SD$ ($t = -4.94, p < .001$) on Conscientiousness, and .65 SD ($t = 5.62, p < .001$) on Extraversion.

H1c predicted that differences between the first-generation minority groups and the second-generation minority groups to the advantage of the second-generation minority groups exist on all selection tools. To test this hypothesis, first- and second-generation ethnic minority groups were compared. Positive d values indicate the second-generation minority group scoring higher than the first-generation minority group. For the cognitive ability test (PIT), 96% of the comparisons supported the hypothesis. Less support was found on the personality test (PPV; 20%), the AC (25%), the employment interview (0%), and the employment recommendation (25%). Differences varied between $-.47 SD$ ($t = -2.31, p < .05$) on Conscientiousness (PPV) for the difference between the first- and second-generation Caribbean group, and .90 SD ($t = 4.81, p < .001$) on Verbal Comprehension (PIT), also for the difference in Caribbean group. Three remarkable findings are highlighted. Firstly, on the PPV dimension Conscientiousness, the difference between the first- and second-generation Caribbean and Surinamese groups was to the advantage of the first-generation Caribbean ($-.47 SD$; $t = -2.31, p < .05$) and Surinamese group ($-.27 SD$; $t = -2.17, p < .05$). Secondly, the Turkish group showed a different pattern. Scores on the interview ($-.24 SD$; $t = -2.36, p < .05$) and the employment recommendation (the dimension Agency [$-.21 SD$; $t = -2.01, p < .05$] and the dimension Socio-Cultural Awareness [$-.21 SD$; $t = -1.95, p < .05$]) showed differences to the advantage of the first-generation Turkish group. Lastly, as shown in Table 2 (majority vs. second-generation minority), the majority group

TABLE 2
Differences Between the Majority Group and First- and Second-Generation Minority Groups on a Cognitive Ability Test, a Personality Test, an AC, an Employment Interview, and a Final Employment Recommendation

Dimension	<i>d</i>				<i>t Test</i> <i>Levene's Test (F)</i>			
	<i>C</i>	<i>M</i>	<i>S</i>	<i>T</i>	<i>C</i>	<i>M</i>	<i>S</i>	<i>T</i>
Majority versus first-generation minority groups								
PIT								
Verbal comprehension	1.27	.98	1.00	1.18	11.06**	9.28**	11.41**	14.25**
					.08	1.60	1.25	3.62
Inductive reasoning	.90	1.30	1.00	.96	6.56**	12.30**	11.35**	10.66**
					6.48*	.41	1.51	8.07*
Numerical reasoning	.66	1.08	.79	.67	5.67**	13.61**	10.76**	8.49**
					3.38	13.88**	10.42*	3.23
Word fluency	1.03	1.15	.88	1.12	11.17**	10.76**	9.89**	17.12**
					6.16*	1.72	1.54	13.65**
Spatial ability	.69	1.06	.79	.87	5.91**	9.94**	8.95**	10.11**
					1.01	.45	2.33	3.90*
Picture arrangement	.90	.98	.98	1.11	6.15**	7.45**	8.79**	12.31**
					14.68**	15.07**	25.17**	6.68*
PPV								
Extraversion	.65	.13	.41	.19	5.62**	1.22	4.62**	2.39*
					3.51	2.93	1.59	.05
Altruism	.25	.01	.16	.00	1.92*	.08	1.65	-.05
					6.82*	6.73*	5.84*	10.95*
Conscientiousness	-.49	-.30	-.52	-.43	-4.94**	-2.88*	-4.72**	-5.47**
					8.66*	.03	2.09	2.67
Emotional stability	.04	.11	-.07	-.28	.39	1.02	-.79	-3.49**
					.01	.07	.14	1.18
Intellect	.35	.21	.30	.44	3.04*	1.66	3.13*	5.54**
					.53	14.65**	5.35*	1.69

(continued)

AC									
Agency	.56	.39	.43	.46	4.27**	3.38**	4.67**	4.09**	
					1.54	.00	.16	1.56	
Communion	.27	.29	.35	.30	2.31*	2.49*	3.73**	3.34**	
					5.44*	.03	.57	2.89	
Interview	.50	.23	.46	.34	3.87**	2.05*	4.99**	3.53**	
					.44	.33	.48	3.96*	
Final recommendation									
Agency	.59	.37	.55	.47	4.47**	3.15*	5.84**	5.17**	
					.34	.18	1.00	.20	
Communion	.35	.27	.39	.17	2.63*	2.35*	4.21**	1.74*	
					.06	1.20	.28	4.08*	
Sociocultural awareness	.27	.21	.33	.19	2.01*	1.77*	3.35**	2.06*	
					2.53	1.20	.81	3.58	
Majority versus second-generation minority groups									
PIT									
Verbal comprehension	.33	.48	.41	.82	1.89*	4.36**	3.91**	14.90**	
					.00	6.28*	6.53*	.87	
Inductive reasoning	.17	.62	.51	.68	.95	6.47**	4.84**	10.84**	
					.17	2.36	6.29*	11.41*	
Numerical reasoning	.43	.56	.58	.43	2.46*	6.69**	7.46**	8.50**	
					1.65	4.01*	5.75*	4.15*	
Word fluency	.28	.60	.48	.71	1.61	6.22**	5.32**	14.12**	
					.13	2.42	3.71	7.42*	
Spatial ability	.06	.63	.33	.55	.38	5.63**	3.64**	9.98**	
					1.86	4.39*	.09	.04	
Picture arrangement	.16	.47	.50	.74	.93	4.87**	4.79**	12.29**	
					.01	2.67	5.72*	10.77*	

(continued)

TABLE 2 (Continued)

Dimension	<i>d</i>				<i>t Test</i> Levene's Test (<i>F</i>)			
	<i>C</i>	<i>M</i>	<i>S</i>	<i>T</i>	<i>C</i>	<i>M</i>	<i>S</i>	<i>T</i>
	PPV							
Extraversion	-.03	-.05	.03	.08	-.16	-.51	.34	1.47
					5.88*	.00	.21	.15
Altruism	.14	-.04	.03	-.06	.79	-.40	.29	-.96
					.22	4.50*	7.68*	7.09*
Conscientiousness	-.11	-.23	-.15	-.45	-.62	-2.39*	-1.55	-8.79**
					.11	.00	3.95*	12.34**
Emotional stability	.05	-.17	-.09	-.20	.27	-1.80*	-1.04	-3.78**
					2.77	.35	.90	4.56*
Intellect	.06	.17	.02	.38	.34	1.59	.23	6.86**
					.79	4.62*	3.79	1.58
AC								
Agency	.24	.32	.21	.59	1.36	3.27**	2.16*	10.29**
					.14	.72	1.15	.00
Communion	.19	.24	.02	.36	1.09	2.38*	.22	6.28**
					1.01	.03	.02	.66
Interview	.24	.15	.24	.62	1.39	1.56	2.58*	10.72**
					2.08	1.19	.23	.46
Final recommendation								
Agency	.32	.32	.26	.68	1.80*	3.25**	2.78*	11.94**
					.99	.19	.92	.07
Communion	.11	.18	.09	.35	.61	1.85*	.94	6.15**
					.09	.05	.30	2.50
Sociocultural awareness	.36	.09	.29	.41	1.99*	.72	2.98*	7.06**
					.66	4.22*	.01	1.74

(continued)

First-generation versus second-generation minority groups

PIT

Verbal comprehension	.90	.44	.53	.29	4.81**	3.18*	4.35**	3.02*
					.02	.44	.96	.92
Inductive reasoning	.64	.60	.44	.25	3.21*	4.45**	3.61**	2.64*
					3.28	.22	.65	.13
Numerical reasoning	.26	.61	.25	.26	1.27	4.53**	2.01*	2.76*
					.03	2.61	.31	.10
Word fluency	.81	.60	.44	.46	4.22**	4.36**	3.56**	4.89**
					1.29	.02	.44	2.47
Spatial ability	.49	.37	.42	.31	2.42*	2.64*	3.39**	3.22**
					.30	.59	.61	2.16
Picture arrangement	.61	.45	.40	.34	3.09*	3.24**	3.29**	3.54**
					3.46	3.35	2.84	.06

PPV

Extraversion	.60	.15	.36	.10	2.71*	1.09	2.79*	1.06
					10.10*	1.29	.22	.01
Altruism	.08	.03	.11	.05	.36	.25	.88	.47
					1.05	.19	.10	1.35
Conscientiousness	-.47	-.09	-.27	.01	-2.31*	-.64	-2.17*	.11
					2.85	.03	6.46*	.39
Emotional stability	-.02	.27	.02	-.08	-.12	1.89*	.13	-.83
					1.96	.29	.82	.10
Intellect	.24	.03	.25	.05	1.18	.20	2.01*	.55
					.13	1.86	.04	2.75

(continued)

TABLE 2 (Continued)

Dimension	<i>d</i>				<i>t Test</i> Levene's Test (<i>F</i>)			
	<i>C</i>	<i>M</i>	<i>S</i>	<i>T</i>	<i>C</i>	<i>M</i>	<i>S</i>	<i>T</i>
AC								
Agency	.34	.09	.23	-.12	1.33	.41	1.70*	-1.25
Communion	.13	.06	.32	-.05	.18	.32	.22	.99
Interview	.29	.08	.21	-.24	.39	.31	2.40*	-.56
Final recommendation					.32	.05	.18	.94
Agency	.29	.06	.28	-.21	1.17	.50	1.59	-2.36*
Communion	.25	.09	.30	-.17	2.60	.07	.02	1.53
Sociocultural awareness	-.08	.11	.05	-.21	1.15	.26	2.03*	-2.01*
					.20	.00	.00	.05
					1.04	.60	2.30*	-1.58
					.01	1.14	.59	.67
					-.48	.72	.24	-1.95*
					.14	.22	.29	3.95*

Note. PIT = cognitive ability test; PPV = personality test; AC = assessment center; C = the Caribbean group; M = the Moroccan group; S = the Surinamese group; and T = the Turkish group. A positive *t* value means that the mean of the majority group is higher than the mean of the minority group. *d* = difference between majority and minority ethnic group means in standard deviation units, positive *d* values indicate the ethnic majority group scoring higher.

* $p < .05$, two-tailed for Levene's test and one-tailed for the *t* test. ** $p < .001$, two-tailed for Levene's test and one-tailed for the *t* test.

still scored higher than the second-generation minority groups. The personality test scores (PPV), again, showed different results.

The results relevant to H1d showed score differences on all subtests of the cognitive ability test between the first-generation minority groups and the second-generation minority groups. All differences were to the advantage of the second-generation minority group. To further look at the results, a distinction was made between subtests for crystallized intelligence and subtests for fluid intelligence. Subtests that measure crystallized intelligence are Verbal Comprehension, Numerical Reasoning, Word Fluency, and Picture Arrangement. Subtests that measure fluid intelligence are Inductive Reasoning and Spatial Ability. In line with findings from Bleichrodt and Van den Berg (1995), the score differences between both generations on subtests of fluid intelligence were somewhat smaller than the differences on subtests of crystallized intelligence. The differences of fluid intelligence varied from .25 *SD* ($t = 2.64, p < .05$) to .64 *SD* ($t = 3.21, p < .05$; mean difference is .44 *SD*). For crystallized intelligence, the differences varied from .25 *SD* ($t = 2.01, p < .05$) to .90 *SD* ($t = 4.81, p < .001$; mean difference is .48 *SD*).

Second Group of Hypotheses

The results relevant to H2 (a–c) are shown in Tables 3 and 4. From the fit indexes shown in Table 3, it can be concluded that the model fit (χ^2/df of 55.88 and 54.52; TLI of .88; CFI of .92; RMSEA of .07) was good. H2a stated that Dutch language-proficiency could explain more of the variability in ethnic score differences on the cognitive ability test (PIT) and the personality test (PPV) than education and ethnicity. Table 4 reports the unstandardized and standardized path coefficients. Support for this hypothesis was found on the PIT as well as the PPV. On the cognitive ability test (PIT), the explained variance by Dutch language-proficiency was 16% (unstandardized path coefficient of .40, $p < .001$) compared to 0.05% by education (unstandardized path coefficient of .07, $p < .001$) and 0.05% by ethnicity

TABLE 3
Fit Indexes of a Cognitive Ability Test, a Personality Test, an AC,
an Employment Interview, and a Final Employment Recommendation

<i>Dimension</i>	χ^2	<i>df</i>	χ^2/df	<i>TLI</i>	<i>CFI</i>	<i>RMSEA</i>
PIT	5922.79*	106	55.88	.88	.92	.07
PPV	5069.98*	93	54.52	.88	.92	.07
AC	2553.90*	54	47.29	.92	.95	.06
Interview	2449.10*	44	55.66	.90	.95	.07
Final recommendation	3319.56*	66	50.30	.91	.94	.07

Note. TLI = Tucker-Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; PIT = cognitive ability test; PPV = personality test; AC = assessment center.

* $p < .001$, one-tailed.

TABLE 4
Unstandardized and Standardized Path Coefficients of a Cognitive Ability Test, a Personality Test,
an AC, an Employment Interview, and a Final Employment Recommendation

<i>Dimension</i>	<i>Unstandardized Path Coefficients</i>			<i>Standardized Path Coefficients</i>			<i>R</i> ²
	<i>Dutch Language- Proficiency</i>	<i>Education</i> ^a	<i>Ethnicity</i> ^b	<i>Dutch Language- Proficiency</i>	<i>Education</i> ^a	<i>Ethnicity</i> ^b	
PIT	.40**	.07**	.10**	.40	.07	.07	.86
PPV	.09**	-.01	-.01	.08	.01	.01	.01
AC	.18**	.01*	.02*	.25	.02	.02	.09
Interview	.10**	.02**	.01*	.29	.06	.02	.12
Final recommendation	.14**	.02**	.02*	.35	.05	.03	.18

Note. PIT = cognitive ability test; PPV = personality test; AC = assessment center.

^a2 = preparatory vocational education medium level B (VBO-B); 3 = junior general secondary education (MAVO); 4 = senior secondary vocational education (MBO); 5 = high education means senior general secondary education (HAVO); 6 = university preparatory education (VWO); higher professional education (HBO); or research-oriented education (WO).

^b1 =Turkish, 2 = Moroccan, 3 = Surinamese, 4 = Caribbean, 5 = Western ethnic minorities, 6 = Dutch ethnic majorities.

* $p < .05$, one-tailed. ** $p < .001$, one-tailed.

(unstandardized path coefficient of .10, $p < .001$). For the personality test (PPV), support for H2a was found but the support was less overwhelming than for the cognitive ability test. Dutch language-proficiency explained more variance than education and ethnicity, accounting for 0.60% (unstandardized path coefficient of .09, $p < .001$) of the variability in test scores, whereas education (unstandardized path coefficient of $-.01$, *ns*) and ethnicity (unstandardized path coefficient of $-.01$, *ns*), together, accounted for only 0.02% of the variability.

H2b, derived from assumed-characteristics theory, predicted that Dutch language-proficiency and education together would explain more of the variability in score differences on the employment interview and the final recommendation than ethnicity. From the fit indexes shown in Table 3, it can be concluded that the model fit of the models for the interview and the employment recommendation (χ^2/df of 55.66 and 50.30; TLI of .90; CFI between .94 and .95; RMSEA of .07) was good. Support was found for H2b (see Table 4). For the interview, the explained variance of score differences by Dutch language-proficiency and education was 9% (unstandardized path coefficients of .10 for Dutch language-proficiency [$p < .001$] and .02 for education [$p < .001$]). Ethnicity explained 0.04% (unstandardized path coefficients of .01, $p < .05$) of the variability in test scores. For the employment recommendation, the explained variance of score differences by Dutch language-proficiency and education was 13% (unstandardized path coefficients of .14 for Dutch language-proficiency [$p < .001$] and .02 for education [$p < .001$]). Ethnicity explained 0.09% (unstandardized path coefficients of .02, $p < .05$) of the variability in test scores.

H2c predicted that Dutch language-proficiency and education would not explain more of the variability in assessors' evaluations on the AC than ethnicity. Support for this hypothesis was not found. On the AC, as on the interview and the employment recommendation, more variance was explained by Dutch language-proficiency and education (6%) than by ethnicity (0.04%). However, Dutch language-proficiency and education do seem to account for less explained variance on the AC (6%) than on the employment interview (9%) and the final employment recommendation (13%).

Third Group of Hypotheses

Tables 5 and 6 show the results relevant to H3 (a–b) which were derived from complexity–extremity theory. No support was found for H3a or H3b. H3a predicted that ethnic majority assessors would rate the ethnic minority group with an excellent Dutch language-proficiency and education higher than the ethnic majority group with the same language-proficiency and education. Score differences on the employment interview and the final recommendation between the ethnic majority group and the ethnic minority group did not exist or were to the advantage of the ethnic majority group. The ethnic minority group members with excellent Dutch

TABLE 5
High Dutch Language-Proficiency and Education: Means and Standard Deviations of the Majority
and Minority Group and Their Differences

<i>Dimension</i>	<i>Majority Group^a</i>		<i>Minority Group^b</i>		<i>d</i>	<i>Levene's Test</i>	<i>t Test</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>F</i>	<i>t</i>
Interview	4.48	.63	4.38	.60	.16	.06	1.38
Final recommendation							
Agency	4.41	.82	4.15	.83	.31	.02	2.66*
Communion	4.25	1.06	3.96	1.03	.27	.11	2.28*
Sociocultural awareness	4.41	.62	4.32	.59	.15	.27	1.35

Note. High education means senior general secondary education (HAVO), university preparatory education (VWO), higher professional education (HBO), or research-oriented education (WO). High Dutch language-proficiency means scoring in the top 4%. *t* = *t* test for equality of means; *F* = Levene's test equality of variances; *d* = difference between majority and minority ethnic group means in standard deviation units, positive *d* values indicate the ethnic majority group scoring higher

^a*N* = 400. ^b*N* = 92.

p* < .05, two-tailed for Levene's test and one-tailed for the *t* test. *p* < .001, two-tailed for Levene's test and one-tailed for the *t* test.

TABLE 6
 Low Dutch Language-Proficiency and Education: Means and Standard Deviations of the Majority
 and Minority Group and Their Differences

<i>Dimension</i>	<i>Majority Group^a</i>		<i>Minority Group^b</i>		<i>d</i>	<i>Levene's Test</i>	<i>t Test</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>F</i>	<i>t</i>
Interview	4.04	.51	4.00	.49	.08	.47	.95
Final recommendation							
Agency	3.62	.77	3.55	.71	.09	.11	.93
Communion	3.66	.94	3.62	.94	.04	.41	.44
Sociocultural awareness	4.06	.50	4.04	.47	.04	.68	.35

Note. Low education means preparatory vocational education medium level B (VBO-B), junior general secondary education (MAVO), or senior secondary vocational education (MBO). Low Dutch language-proficiency means scoring in the lowest 5%. *F* = Levene's test for equality of variances; *t* = *t* test for equality of means; *d* = difference between majority and minority ethnic group means in standard deviation units, positive *d* values indicate the ethnic majority group scoring higher.

^a*N* = 456. ^b*N* = 173.

p* < .05, two-tailed for Levene's test and one-tailed for the *t*-test. *p* < .001, two-tailed for Levene's test and one-tailed for the *t* test.

language-proficiency and education were rated significantly lower on the employment-recommendation factors of Agency ($t = 2.66, p < .05$) and Communion ($t = 2.28, p < .05$) than the ethnic majority group with the same Dutch language-proficiency and education.

H3b predicted that ethnic majority assessors would rate the ethnic minority group with low Dutch language-proficiency and education lower than the ethnic majority group with the same language-proficiency and education. The results showed no significant differences between the ethnic majority and minority group with low Dutch language-proficiency and education.

DISCUSSION

First Group of Hypotheses

Score differences found in the literature on the cognitive ability test, the AC, and the employment interview were replicated in this study by the score differences between the Dutch ethnic majority group and the first-generation minority groups. Noteworthy, striking score differences existed between the first- and second-generation minority groups. The differences on the personality test (PPV) were much less systematic with sometimes the majority group and sometimes the minority group scoring higher. Clear systematic differences were found on the dimension Conscientiousness with all ethnic minority groups, both first- and second-generation, scoring higher than the ethnic majority group.

Differences between the first-generation minority group and the second-generation minority group were the largest for the Caribbean group and the smallest for the Turkish group. Turkish minority applicants scored somewhat lower than the other ethnic minority groups on all selection measures. A recent publication by the Dutch National Bureau of Statistics (CBS, 2004) on marks of high school students in The Netherlands reported corresponding results: The Turkish students had poorer results than the Caribbean, Moroccan, and Surinamese students. Turkish people have a history of migrant labor. Most of the Turkish people are Muslim and have a strong sense of their own culture and history (e.g., Nijsten, 1998), whereas Caribbeans are from Dutch colonies. This might be one possible explanation why the differences between the majority group and Turkish minorities remain large, while the differences between the majority group and the second-generation Caribbean group is much smaller than the differences between the majority group and the first-generation Caribbean group. The Turkish group might be a more separate group because of their strong sense of culture, even after several generations, than the Caribbeans who might integrate more easily into Dutch society because of the connection of the Dutch Antilles with The Netherlands. The decrease in Moroccan and Surinamese first- and second-generation minority score differences was in be-

tween the decrease from Caribbean and Turkish groups. The studies finding consensus on a hierarchy of social distance to the Dutch majority group (e.g., Hraba et al., 1989; Verkuyten et al., 1996) confirm our findings, as these show an ethnic hierarchy where the Caribbean minority group is placed on top of the minority groups and the Turkish minority group at the bottom.

The results relevant to H1d showed score differences between first- and second-generation minority groups on all subtests of the cognitive ability test. Score differences were somewhat larger on subtests for crystallized intelligence. These findings are comparable to findings from Bleichrodt and Van den Berg (1995).

Second Group of Hypotheses

Dutch language-proficiency was able to explain more of the variability in ethnic score differences than education and ethnicity on the cognitive ability test and on the personality test. However, the results on the personality test were less profound than the results on the cognitive ability test. Though Dutch language-proficiency did explain more of the variance between test scores on the personality test than education and ethnicity, in general the entire model did only explain a very small amount of the variance for personality (R^2 was small). Therefore, it can be concluded that other variables than Dutch language-proficiency, education, and ethnicity are possibly related to ethnic score differences on the personality test. Certain applicant factors may be related to score differences between ethnic groups. One type of applicant factor related to ethnic groups, which Ryan (2001) investigated for cognitive ability tests, is test motivation and test-taking attitudes. These factors, which were not included in this study, may also influence the scores on personality tests.

Possible explanatory factors for score differences between ethnic groups on subjective measures have had little attention in past research. In this study, explanations were derived from two theories from social psychology, namely assumed-characteristics theory (Locksley et al., 1980; Locksley et al., 1982a, 1982b) and complexity-extremity theory (Linville, 1982; Linville & Jones, 1980). In this study, these theories were taken out of the lab for the first time. The results from assumed-characteristics theory have demonstrated that knowledge of relevant demographic information diminishes group membership effects. More variance in score differences was explained by Dutch language-proficiency and education on the employment interview and the final recommendation, during which this background information was known, than on the AC, where such knowledge was not given to the assessors. An explanation for the finding that Dutch language-proficiency and education did not explain as much as or less variance than ethnicity on the AC may be that assessors did have some knowledge of the applicants' ethnicity-related demographics just by looking at their behavior and hearing them speak. Research by Jussim et al. (1987, 1996) investigated only one group of assessors,

which had knowledge of applicants' demographics. They showed somewhat larger percentages of explained variance (4% for ethnicity, 21% for personal appearance [appearing upper vs. lower class], and 19% for dialect style [(non-) standard English speaking]) than the results from this study. This was probably due to the highly controlled setting of their lab experiment, explaining why lower percentages of explained variance were found in the present less controlled, but more ecological valid field study.

Third Group of Hypotheses

For complexity–extremity theory, the results were unsupportive. The ethnic minority and majority groups with a low Dutch language-proficiency and education showed no differences in scores. The ethnic minority group with excellent Dutch language-proficiency and education was not rated significantly higher but, on the contrary, lower on Agency and Communion than the ethnic majority group with the same Dutch language-proficiency and education. Thus, a general tendency seems to exist to rate the ethnic minority group a bit lower than the ethnic majority group on the employment interview and the final employment recommendation. Though systematic and positive *d* values were found, indicating the majority group scoring higher, the effect sizes were very small. Clearly, complexity–extremity processes have not been of influence on assessors' behavior. Tajfel's Social Identity Theory (Tajfel, 1978), which argues that the motivation to maintain a positive social identity and high self-esteem leads to a bias in favor of the in-group, might provide a better explanation for the assessors' evaluations. Maybe other processes, such as demographic similarity between applicants and assessors or perceived similarity of applicants by assessors, are responsible for the score differences found in this study. These issues should have more attention in future research.

SUMMARY

To summarize this study, three major points are highlighted. First, as expected, score differences between the first- and second-generation minority groups existed to the advantage of the second-generation minority group. The second-generation minority group did still score lower than the ethnic majority. First- and second-generation minority differences to the advantage of the second-generation minority group existed on both the objective and the subjective measures. They were largest for the Caribbean group and smallest for the Turkish group. Second, among the ethnicity-related demographic variables Dutch language-proficiency, education, and ethnicity, Dutch language-proficiency and education explained most of the variability in score differences on the employment interview and the final recommendation. This is in line with assumed-characteristics theory. Third, the results were unsupportive for complexity–extremity theory. Other possible explana-

tory factors for the score differences between ethnic groups on subjective measures, such as demographic and perceived similarity between applicants and assessors, should receive more attention in future research.

ACKNOWLEDGMENT

This article has benefited from the feedback provided by Iain J. Coyne.

REFERENCES

- Anderson, N., & Ones, D. S. (2003). The construct validity of three entry-level personality inventories used in the UK: Cautionary findings from a multiple inventory investigation. *European Journal of Personality, 17*, 39–66.
- Arbuckle, J. L. (2003). Amos 5.0 [Computer software]. Chicago: Smallwaters.
- Bass, B. M., & Barrett, G. V. (1981). *People, work, and organizations* (2nd ed.). Boston: Allyn & Bacon.
- Bleichrodt, N., & Van den Berg, R. H. (1995). *Multiculturele capaciteitentest middelbaar niveau (MCT-m). Handleiding* [Multicultural capacity test medium level. Manual]. Amsterdam: Stichting NOA.
- Bors, D. A., & Forrin, B. (1995). Age, speed of information processing, recall, and fluid intelligence. *Intelligence, 20*, 229–248.
- Bureau InterCulturele Evaluatie. (2000). *Intaketoets beroepsopleidingen op cd-rom*. [Intaketest for vocational education on cd-rom]. Lienden, The Netherlands: Bureau ICE.
- Cascio, W. F. (1991). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Amsterdam: North Holland.
- Central Bureau of Statistics (2004). *Mens en maatschappij—onderwijs: Aandeel niet-westerse allochtonen in examenklassen voortgezet onderwijs, naar schoolgemeente of woongemeente*. [Men and society—education: Number of non-western ethnic minority high school students in final classes per school or area]. Voorburg/Heerlen: The Netherlands: CBS.
- Cohen, J. (1988). *Standard power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Coleman, L. M., Jussim, L., & Kelley, S. H. (1995). A study of stereotyping: Testing three models with a sample of blacks. *Journal of Black Psychology, 21*, 332–356.
- Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology, 51*, 357–374.
- Goldstein, H. W., Yusko, K. P., & Nicolopoulos, V. (2001). Exploring black-white subgroup differences of managerial competencies. *Personnel Psychology, 54*, 783–807.
- Goldstein, H. W., Zedeck, S., & Goldstein, I. L. (2002). *G*: Is this your final answer? *Human Performance, 15*, 123–142.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, Ch. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.
- Hattrup, K., Rock, J., & Scalia, Ch. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology, 82*, 656–664.
- Heneman, H. G., & Heneman, R. L. (1994). *Staffing organizations*. Middleton, WI: Mendota House.

- Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gc-Gf theory. In D. P. Flanagan & J. L. Genshaft (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. New York: Guilford.
- 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: Lawrence Erlbaum Associates, Inc.
- 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.
- Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, 83, 179–189.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340–362.
- Jensen, A. R. (1993). Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science*, 2, 53–56.
- Jussim, L. (1990). Social reality and social problems: The role of expectancies. *Journal of Social Issues*, 46(2), 9–34.
- Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, 98, 54–73.
- Jussim, L. (1993). Accuracy in interpersonal expectations: A reflection-construction analysis of current and classic research. *Journal of Personality*, 61, 637–668.
- Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology*, 52, 536–546.
- Jussim, L., Fleming, C. J., Coleman, L., & Kohberger, C. (1996). The nature of stereotypes II: A multiple-process model of evaluations. *Journal of Applied Social Psychology*, 26, 283–312.
- Klinkenberg, E. L., & Van Leeuwen, A. E. (2003). *Voortgangverslag ontwikkeling M5Q-IWSP* [Progress report development M5Q-IWSP]. Culemborg, The Netherlands: Meurs Personeelsadvies.
- Lem, J., & Van Doorn, E. (2000). *Voortgangsrapportage. Onderzoek "kenmerkende voorspellers" politie* [Progress report. Study "noticeable predictors" police]. Culemborg, The Netherlands: Meurs Personeelsadvies.
- Linville, P. W. (1982). The complexity–extremity effect and age-based stereotyping. *Journal of Personality and Social Psychology*, 42, 193–211.
- Linville, P. W., & Jones, E. E. (1980). Polarized appraisals of out-group members. *Journal of Personality and Social Psychology*, 38, 689–703.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, 39, 821–831.
- Locksley, A., Hepburn, C., & Ortiz, V. (1982a). On the effects of social stereotypes on judgments of individuals: A comment on Grant and Holmes's "The integration of implicit personality theory schemas and stereotypic images." *Social Psychology Quarterly*, 45, 270–273.
- Locksley, A., Hepburn, C., & Ortiz, V. (1982b). Social stereotypes and judgments of individuals: An instance of base-rate fallacy. *Journal of Experimental Social Psychology*, 18, 23–42.
- Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., et al. (1992). Studies of the structural behavioral interview. *Journal of Applied Psychology*, 77, 571–587.
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance*, 15, 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, 255–276.

- Outz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15*, 161–171.
- Plomin, R. (1988). The nature and nurture of cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. IV). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Rijks Psychologische Dienst (1975). *Politie Intelligentie Test* [Police Intelligence Test]. Den Haag, The Netherlands: RPD.
- Rokeach, M., & Mezel, L. (1966). Race and shared belief as factors in social choice. *Science, 151*, 167–172.
- Roth, P. L., Van Iddekinge, C. H., Huffcutt, A. I., Eidson, C. E., & Bobko, P. (2002). Correction for range restriction in structured interview ethnic group differences: The values may be larger than researchers thought. *Journal of Applied Psychology, 87*, 369–376.
- Ryan, A. M. (2001). Explaining the black-white test score gap: The role of test perceptions. *Human Performance, 14*, 45–75.
- 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.
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162–173.
- Schmitt, N., & Mills, A. E. (2001). Traditional tests and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology, 86*, 451–458.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105*, 409–429.
- Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. New York: Academic.
- Te Nijenhuis, J. (1997). *Comparability of test scores for immigrants and majority group members in the Netherlands*. Unpublished 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.
- Van den Berg, R. H. (2001). *Psychologisch onderzoek in een multiculturele samenleving* [Psychological research in a multicultural society]. Unpublished doctoral dissertation, Vrije Universiteit, Amsterdam.
- Van den Berg, R. H., & Van Leest, P. (1999). Praktisch testgebruik bij allochtonen: wanneer zinlijk, nuttig of waardeloos? [Practical use of tests with ethnic minorities: when useful, useless, or worthless?]. *Psycholoog, juni*, 256–260.
- Van der Maesen, P. E. A. M. (1992). *Het rendement van personeelsselectie* [The efficiency of personnel selection]. Unpublished doctoral dissertation, Rijksuniversiteit Groningen, The Netherlands.
- Van Leest, P. (1997). *Persoonlijkheidsmeting bij allochtonen* [Ethnic minority personality measurement]. Unpublished doctoral dissertation, Vrije Universiteit, Amsterdam.
- Van Leeuwen, A. E. (2000). *Constructie van de M5Q voor IWSP* [Construction of the M5Q for IWSP]. Culemborg, The Netherlands: Meurs Personeelsadvies.
- 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.
- Waldman, D. A., & Avolio, B. J. (1991). Race effects in performance evaluations controlling for ability, education, and experience. *Journal of Applied Psychology, 76*, 897–901.
- 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: Guilford.