Chapter 7
The Reference-Group Effect in Personality Research: Data from a Turkish-Dutch Minority Group
Abstract

The Reference-Group Effect in Personality Research: Data from a Turkish-Dutch Minority Group*

Small or absent ethnicity effects on personality scales may indicate that ethnic groups do not differ on average on personality traits. However, another explanation for this (lack of) finding may lie with the so-called reference-group effect (RGE), which occurs when responses to self-report items are based not on respondents' absolute level of a construct but rather on their level relative to a comparison group. Until now, no studies have been conducted to test the RGE in personality judgments with different ethnic reference groups. Therefore, our goal was to examine to what extent Turkish-Dutch minorities are influenced by perceptions of comparison others when filling out a personality test. The results show that when the Turkish-Dutch compared themselves with people from their own Turkish-Dutch minority group (in-group comparison), there were no score differences between the Dutch majority and the Turkish-Dutch minority. Yet, when Turkish-Dutch minorities thought about how they behave in comparison to the Dutch majority group (out-group comparison), they saw themselves as less honest and humble. Furthermore, when the Turkish-Dutch used an out-group comparison other, they saw themselves as more emotional as well as less agreeable and less open for new experiences than when they used an in-group comparison other. The findings do suggest that Turkish-Dutch members are influenced by perceptions of comparison others when filling out a personality test.

* This chapter is submitted for publication as:
De Vries, A., Born, M. Ph., & De Vries, R. E. (submitted). The Reference-Group Effect in Personality Research: Data from a Turkish-Dutch Minority Group.
An advantage of using personality tests is that there is hardly any evidence that they have an adverse impact (e.g., De Meijer, Born, Terlouw, & Van der Molen, 2006). Adverse impact is defined by a substantially different rate of selection in hiring, promotion, or another employment decision which works to the disadvantage of members of a certain race, sex, or ethnic group. There are two main reasons for different selection ratios: 1) different selection ratios may result from true differences between two groups on important characteristics and/or 2) different selection ratios may result from the use of unfair and biased tests. Interestingly, while still retaining predictive validity for performance, personality tests have been found to bias selection ratios to a lesser extent than mental ability tests. For example, ethnic score differences on personality measures seem to be significantly smaller than on cognitive predictors (Hough, Oswald, & Ployhart, 2001; Ployhart & Holtz, 2008; Roth, Bobko, McFarland, & Buster, 2008).

Smaller or absent ethnicity effects on personality scales may indicate that ethnic groups do not differ on average on personality traits. However, another explanation for this (lack of) finding may lie with the so-called reference-group effect (RGE; Peng, Nisbett, & Wong, 1997; Heine, Lehman, Peng, & Greenholtz, 2002), which refers to “the tendency for people to respond to subjective self-report items by comparing themselves with implicit standards from their culture” (Heine, Buchtel, & Norenzayan, 2008, p. 309). In the cross-cultural literature, results seem to indicate that true differences between ethnic groups, particularly on values such as individualism and collectivism, may be masked when people from different ethnic groups use different referents in their self-reported values (e.g., Heine et al., 2002). One of the important questions that remains unanswered when assessing the potential benefits of using personality tests is: To what extent are members of ethnic groups influenced by perceptions of comparison others when filling out a personality test? Credé, Bashshur, and Niehorster (2010) were the first to examine the RGE in self-report measures of personality, using different age, gender, and relative (such as family, friends, and people in general) reference groups. Yet, until now, no studies have been conducted to test the RGE in personality judgments with different ethnic reference groups. Moreover, it is not known what, if any, effects are to be expected, especially among ethnic groups who live in the same country.

The reference-group effect in personality research

The RGE is described as the influence of comparison others on the responses to individual self-report items and occurs when responses to self-report items are based not on respondents' absolute level of a construct but rather on their level relative to a comparison group (Credé et al., 2010). In practice, the RGE reflects the phenomenon that respondents within a group are more likely to use people from their own group as comparison others than to use people from another group. This notion of the RGE is grounded in the social comparison theory, which claims that similar others are preferred as referents of social comparison rather than dissimilar others (Festinger, 1954). However,
in the case of cross-cultural personality comparisons, when all people compare themselves with similar others, i.e., people from the own cultural group, true personality differences between cultural or ethnic groups may be masked.

According to Heine et al. (2002), true personality differences may be masked due to the use of ordinal or (quasi-)interval scales, such as the Likert scale, which are typically used for cross-cultural and cross-ethnic personality comparisons. For example, people with a Chinese background will compare themselves with the norm of the Chinese cultural group, which by definition is always located around the middle of the scale. However, people with an American background will compare themselves with the norm of their cultural group, which again by definition is located around the middle of the scale (Peng et al., 1997). As a result, true personality differences between cultural or ethnic groups will be masked, due to a cognitive shift in standards when evaluating the self (cf. shifting standards research by Biernat, Manis, & Kobrynowicz, 1997; Biernat, Manis, & Nelson, 1991).

Cross-cultural comparisons can even lead to opposite results than might be expected based on true differences (Heine et al., 2002). In the case of height for example, when using a 5-point Likert scale (with anchors ranging from very tall to very short), Can1 from Turkey with a height of 70 inches (178 centimeters) may evaluate himself as tall compared with other males from the Turkish culture (mean male height 68 inches; 174 centimeters). In The Netherlands, the norm for the average male height is higher (72 inches; 184 centimeters); so Sem2 from the Dutch culture with a height of 71 inches (180 centimeter), who is in fact taller than Can, may evaluate himself as short compared with other males from the Dutch culture. If subsequently these evaluations of Can and Sem are compared directly, one unjustly may conclude that Sem is shorter than Can. Therefore, when cultures have different norms for characteristics, results of cultural comparisons may become confounded.

**Personality and perceptions of national character**

Until now, no field studies have been conducted to test the RGE in personality judgments using different ethnic reference groups. Consequently, it is not known what, if any, effects are to be expected, especially among ethnic groups who live in the same country. In the present study, we will examine whether reference-group effects play a role among ethnic minority members, given that they may compare themselves with other ethnic minorities from the same background and/or with ethnic majorities. The current study focuses on the Turkish-Dutch minority group living in The Netherlands. Before comparing the Dutch majority and Turkish-Dutch minority group, we looked at their national personality profiles. The most widely used framework for such personality research is the Big Five

---

1 Most popular given male name in Turkey (2009)
2 Most popular given male name in The Netherlands (2010)
model (or Five-Factor Model). This model distinguishes the following five higher order personality factors: Extraversion, Agreeableness, Neuroticism, Conscientiousness, and Openness to Experience (e.g., Goldberg, 1990; Costa & McCrae, 1992). Researchers have suggested that the Big Five factors are replicable across cultures (McCrae & Costa, 1997; McCrae, Costa, Del Pilar, Rolland, & Parker, 1998) and that comparing mean levels on the Big Five personality factors should reveal differences in personalities between cultures (Saucier & Goldberg, 2001; Schmitt, Allik, McCrae & Benet-Martinez, 2007). However, intercultural studies on personality have shown no relationship between national personality profiles based on self- and observer personality reports and national stereotypes about personality (McCrae & Terracciano, 2006; McCrae, Terracciano, Realo, & Allik, 2007; Terracciano et al., 2005). According to critics, this lack of findings is probably due to the RGE (Ashton, 2007; Heine et al., 2008; McGrath & Goldberg, 2006).

In addition, the present study will also rely on recent reanalyses of existing psycholexical studies which have shown that instead of five, six cross-cultural replicable personality factors emerge (Ashton et al., 2004; Ashton & Lee, 2007; Lee & Ashton, 2008). The six dimensions are known by the acronym ‘HEXACO’ for Honesty-humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to experience. The HEXACO model had been examined in more than ten different countries (including Turkey) and the six personality dimensions seem to be generalizable across these countries (e.g., Ashton et al., 2004; Lee & Ashton, 2008, 2009; Szarota, Ashton, & Lee, 2007; Wasti, Lee, Ashton, & Somer, 2008).

Heine et al. (2008) argued that perceived national character (PNC) may be a better marker of cross-cultural personality differences than self-reported personality. PNC is generally measured by asking respondents to describe the typical member of a culture. From the data of Terracciano et al. (2005) on PNC, it may be deduced that there is a large PNC-difference on Conscientiousness between The Netherlands and Turkey. To compute the PNC of The Netherlands, we used the mean of surrounding countries, that is, Germany, Belgium, and England. In contrast to the mean of the surrounding countries of The Netherlands, Turkey scored significantly lower on Conscientiousness ($t = 8.17, p < .01$). This may be in line with prior studies suggesting that between nations the most relevant distinction is by latitude, with Northerners distinguished from Southerners (e.g., Pennebaker, Rimé, & Blankenship, 1996). The former are generally held to be more conscientious than the latter (McCrae et al., 2007). With regard to the other Big Five personality dimensions, two other PNC-differences in personality were found between The Netherlands (e.g., the mean of the surrounding countries) and Turkey. The latter scored significantly higher on Neuroticism ($t = -4.57, p < .01$) and significantly lower on Openness to Experience ($t = 2.06, p < .05$) than the former.

The question whether there is a PNC-difference on the additional sixth dimension, namely Honesty–Humility, cannot be directly answered using previous data on PNC, given that those data was collected using Big Five measurements instead of HEXACO
measures. Prior research however has shown that two facets of Agreeableness (Straightforwardness and Modesty) from the NEO-PI-R (based on Five-Factor Model) were strongly related to Honesty-Humility in the HEXACO model (Ashton & Lee, 2005). Interestingly, data from McCrae et al. (2007) suggest that especially on the facet scale Straightforwardness, there is a PNC-difference between The Netherlands and Turkey ($t = 2.44$, $p < .05$), with the latter scoring lower on Straightforwardness than the former. There was no difference on the Modesty facet scale.

Assuming that there are PNC-differences on Conscientiousness, Neuroticism, Openness to Experience, and possibly on Honesty-Humility between The Netherlands and Turkey, we will investigate whether there is an indication for the RGE among the Turkish minority group in The Netherlands.\(^3\) Based on the RGE, it may be expected that (1) when an ethnic group compares itself with similar others, such as people from the own ethnic group (in-group comparison), true personality differences between ethnic groups may be masked and (2) when an ethnic group compares itself with another ethnic group (out-group comparison), personality scores will move towards its own PNC.

In the current study, Turkish-Dutch participants completed a personality test under three conditions: a no reference-group condition, a Turkish-Dutch reference-group condition, and a Dutch majority reference-group condition. The Dutch majority participants only filled out the personality test in the no reference-group condition. First, we do not expect significant personality differences between the Turkish-Dutch minority and Dutch majority in the no reference-group conditions as Turkish-Dutch may prefer similar others as referents of comparisons (see Festinger, 1954); thus Turkish-Dutch may compare themselves with people from the own ethnic group (in-group comparison) which may consequently result in the masking of expected personality differences between Turkish-Dutch minorities and Dutch majorities. Second, especially when Turkish-Dutch are instructed to compare themselves with people from the own ethnic group (in-group comparison), we do not expect significant personality differences between Turkish-Dutch minorities and Dutch majority members. Third, in contrast, we do expect to find significant personality differences when Turkish-Dutch are instructed to compare themselves with the Dutch majority (out-group comparison) and therefore we hypothesize the following:

**Hypothesis 1:** Turkish-Dutch minorities who compare themselves with a Dutch majority out-group will show lower scores on Conscientiousness, Openness to Experience, and Honesty-Humility, and higher scores on Neuroticism than Dutch majority members.

---

\(^3\) As behavioral genetics studies have shown that personality traits are strongly heritable (e.g., Bouchard, 1994), we presume that the PNC of Turkey also applies for Turkish-Dutch minorities.
In order to further examine RGEs, we will not only compare the personality scores between the Turkish-Dutch minority and the Dutch majority, but also the personality scores within the Turkish-Dutch sample across the three different conditions. As mentioned above, we expect that Turkish-Dutch are likely to compare themselves with people from the own ethnic group (in-group comparison) in the no-reference condition and therefore we do not expect to find score differences between the no-reference condition and the Turkish-Dutch reference-group condition. However, with regard to the Dutch majority reference-group condition, we expect the following:

*Hypothesis 2*: Turkish-Dutch minorities in the Dutch majority reference-group condition will show lower scores on Conscientiousness, Openness to Experience, and Honesty-Humility, and higher scores on Neuroticism than they do in the no reference-group condition and in the Turkish-Dutch reference-group condition.

**Integration, ethnic identity, and national identity**

Many studies about ethnic majorities and minorities focus on theoretical frameworks concentrating on acculturation, ethnic identity, and national identity (Arends-Tóth & Van de Vijver, 2004; Phinney, 1990; Phinney & Devich-Navarro, 1997). Acculturation theories focus on how an individual relates to the culture of origin and to the dominant or host culture. The well-known acculturation model of Berry (1997) distinguishes between four patterns of acculturation: assimilation, separation, marginalization, and integration. Of these, integration has been found to be the most often used strategy among minorities and is present when both cultural maintenance and involvement with the larger society are sought (Berry, 1997). Furthermore, ethnic identity can be seen as the aspect of acculturation that focuses on the extent to which one identifies itself with the own ethnic group (Phinney, 1990). However, the study of Phinney and Devich-Navarro (1997) revealed that ethnic minorities can experience themselves as being part of two cultures simultaneously, i.e., their own ethnic culture and the national culture of the larger society in which they live. Bringing these theoretical frameworks and the RGE together, we will examine whether the level of integration, ethnic (i.e., Turkish-Dutch) identity, and national (i.e., Dutch) identity explains possible unexpected differences in the Turkish-Dutch sample between the no reference-group condition on the one hand and the Dutch majority reference-group condition and the Turkish-Dutch reference-group condition on the other. In order to explain this, we again bring out Can. For instance, when Can from Turkey is living in The Netherlands and is asked to fill out a personality test without special instructions (no reference-group condition), we expect that he will compare himself with similar others, that is, with people from his own Turkish-Dutch ethnic group when completing the test. Subsequently, when Can is asked to complete the same personality test with the explicit instruction to compare himself with his own ethnic group and we will compare these personality scores with those in the no reference-group condition, we do...
not expect to find score differences, given that Can may used the same comparison group under both circumstances. However, when we do find score differences, it may be possible that Can has compared himself less with his own ethnic group, but more with the Dutch majority group in the no reference-group condition. In a similar vein, when Can is asked to complete the personality test with the explicit instruction to compare himself with the Dutch majority group and we will compare these personality scores with those in the no reference-group condition, we expect to find score differences, given that we expect that Can used a different comparison group in the no reference-group condition (that is, the Turkish-Dutch reference-group). However, when we do not find score differences, it may be possible that Can has compared himself less with his own ethnic group, but more with the Dutch majority group in the no reference-group condition. Therefore, 1) score differences between the no-reference group and the Turkish-Dutch reference group may occur when Can has a weak ethnic identity and/or a strong national identity (i.e., he compares himself with the Dutch majority group in the no-reference condition) and 2) a lack of score differences between the no reference-group condition and the Dutch reference group may occur when Can has a weak ethnic identity and/or a strong national identity (i.e., he compares himself with the Dutch majority group in the no-reference condition). Therefore, we will examine whether the level of integration, ethnic identity, and national identity explains possible unexpected differences in the Turkish-Dutch sample.

Method

Participants and Procedure
In order to collect data, we sent an e-mail to Dutch majority psychology students from the VU University Amsterdam and to Turkish-Dutch student members of a Turkish Academic Network in The Netherlands (TANNET). In line with Statistics Netherlands (Centraal Bureau voor de Statistiek; CBS), someone living in The Netherlands who is born in Turkey or has at least one parent who is born in Turkey belongs to the Turkish-Dutch minority group. All participants were asked to voluntarily fill out a personality inventory and were provided with an Internet link to access the personality questionnaire. The Turkish-Dutch minority participants were asked to complete extra items measuring the level of integration, ethnic identity, and national identity. After one week, only the Turkish-Dutch participants received a link to another version of the personality questionnaire (see the next section for more information about the different versions) and one week thereafter, they received a link to a third version. As a reward for participation, we offered all respondents individualized personality reports and raffled several gift coupons. The final sample consisted of 74 Dutch majority students (26.4% male, $M_{age} = 29.6$ years, $SD = 10.2$, 89% university degree) and 95 Turkish-Dutch students (25.3% male, $M_{age} = 25.2$ years, $SD = 5.1$, 45% higher-level occupational training degree and 51% university degree). Of the Turkish-Dutch participants, 82% were born in The Netherlands and the other 18% had been living
in The Netherlands for more than 11 years ($M_{\text{number of years}} = 23.6$). We checked whether the Dutch majority and Turkish-Dutch minority participants differed in terms of gender, age, and educational degree. The t-tests results showed that the Dutch majority members were older ($t = 3.32, p < .01$) and had higher educational degrees ($t = 4.12, p < .01$). There were no differences in gender.

**Design**

With regard to the Turkish-Dutch sample, the present study employed a within-subject design. First, Turkish-Dutch participants were asked to complete the personality test and received standard instructions which asked to indicate the extent to which each statement applied to them (condition A: no reference-group). Second, one week later, respondents received special instructions and were asked to compare themselves with people from the Dutch majority group when filling out the same personality questionnaire. Subsequently, a reminder was inserted on every screen to emphasize that participants should think about how they behave in comparison to the Dutch majority group when responding to each statement (condition B: Dutch majority reference-group). Third, after another week, we asked participants to evaluate themselves compared with people from their own Turkish-Dutch minority group when completing the personality test (condition C: Turkish-Dutch reference-group). To check for potential order effects, for one half of the Turkish-Dutch sample the sequence of condition B and C was reversed. Paired-sample t-tests indicated that the different order of the question sets did not result in personality score differences (results can be obtained from the first author). The Dutch majority participants only filled out the personality test with standard instructions (condition A: no reference-group).

As a manipulation check, we asked the Turkish-Dutch participants to whom they had compared themselves when filling out the personality questionnaire in condition B and C (i.e., with people from the Dutch majority group or with people from their own Turkish minority group). Furthermore, participants were requested to indicate whether they were able to compare themselves with the selected reference-group. As a result, 27% of the participants in condition B and 5% in condition C were removed from the dataset. The high percentage participants that was removed from condition B could be due to the fact that we asked participants to evaluate themselves compared to the Dutch ethnic majority group. It seems plausible that respondents were confused because of the word ‘ethnic’ when referring to the Dutch majority group. Finally, 41 Turkish-Dutch respondents participated in condition A as well as in condition B and 52 participated in conditions A and C. The final sample that completed every version consisted of 34 Turkish-Dutch participants (17.6% male, $M_{\text{age}} = 24.8$ years, $SD = 4.3$).

**Measures**

**Personality.** In order to measure personality, we used the short version of the HEXACO Personality Inventory-Revised (HEXACO-PI-R). This personality test assesses the six major
dimensions of personality: Honesty–Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience (Ashton & Lee, 2007; Lee & Ashton, 2006, 2008). Important to note is that, compared to the Big Five model, one of the notable changes in the HEXACO model is the different conceptualization of Emotionality and Agreeableness. In the HEXACO model, Agreeableness contains traits related to irritability and temper, which are filed under Emotional (In)stability in the Big Five model. Similarly, traits associated with sentimentality and sensitivity, which in the Big Five model are allocated to Agreeableness, are components of Emotionality in the HEXACO model. The HEXACO-PI-R contains 200 self-descriptive statements in its full-length version. However, for the purposes of this study, the Dutch half-length version with 100 items was used (HEXACO-100; De Vries, Ashton, & Lee, 2009; Lee & Ashton, 2004). The subordinate scales (i.e., facet scales) of this 100-item version are very short (4 items) and are not intended to have high levels of internal-consistency reliability. Therefore, we only looked at the six broad dimensions. Responses were assessed with a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Prior research indicated that the psychometric properties of the Dutch HEXACO-100 factor scales were satisfactory, with alpha reliabilities of the factor scales exceeding .75 (De Vries et al., 2009).

Integration. Respondents were asked to indicate their degree of integration, i.e., the extent to which they wanted to maintain their identity with the home culture, but also wanted to take on some characteristics of the new culture. To this end, we used five integration items of the acculturation questionnaire of Vedder and Van de Vijver (2004). Example items are: “I prefer social activities that involve both Dutch and Turkish-Dutch members” and “It is important to me to be fluent in both the Dutch language and in the Turkish language”. Ratings were made on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). All original items were in English. For the purpose of this study, items were translated in Dutch and back translated by two of the authors. Any disagreement on the final form of the items was resolved through discussion. In the present study, the alpha reliability of the 5 items was .44. A preliminary analysis showed that this low reliability partly was due to one item that measured whether someone was just as willing to marry a Dutch as a Turkish person. For this reason, we excluded this item after which the reliability became .60.

Ethnic identity. Ethnic identity was assessed using the Multigroup Ethnic Identity Measure (Phinney, 1992; Roberts et al., 1999). A total of 12 questions measured two related aspects of ethnic identity. The first aspect included affirmation, belonging, and commitment (7 items). An example item is: “I have a strong sense of belonging to my own ethnic group”. The second aspect measured ethnic identity search (5 items). An example item is: “I am active in organizations or social groups that include mostly members of my own ethnic group”. The translation of the original English items into Dutch was conducted
in the same way as with the integration items. Participants had to indicate how much they agreed or disagreed with each statement on a Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). According to Phinney (1992), it is preferable to use the mean of the 12 items for an overall score. In the present study, the alpha reliability of these items was .88. In this study, the intercorrelation between integration and ethnic identity was .13 (ns).

**National identity.** The American Identity Questionnaire (Phinney & Devich-Navarro, 1997) was used to measure national (i.e., Dutch) identity and consisted of seven statements regarding identification with the Dutch culture. Example items are: “I consider myself as being Dutch” and “I feel that I am part of mainstream Dutch culture”. Again, the translation of the original English items into Dutch was conducted in the same way as with the integration items. Ratings were made on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). In the present study, the alpha reliability of the seven items was .80. Furthermore, the intercorrelation of national identity with integration was .20 (p < .10) and with ethnic identity -.27 (p < .01).

**Results**

**Descriptive analyses**

Table 1 presents the alpha reliabilities, means, and standard deviations of the personality domain scales in the Dutch sample and in the Turkish-Dutch sample. The latter sample is divided into three conditions (no reference-group, Dutch majority reference-group, and Turkish-Dutch reference-group). Alpha reliabilities of the personality domain scales ranged from .68 for Agreeableness to .87 for Honesty-Humility. We used Hakstian and Whalen’s test (1976) to examine whether the alpha reliabilities of the personality domain scales were similar between samples and across conditions. First, between samples, in the no reference-group condition significantly lower alpha reliabilities for Agreeableness, Conscientiousness, and Openness to Experience were found in the Turkish-Dutch minority sample than in the Dutch majority group (respectively .73, .71, and .71 vs. .86, .82, and .83). Second, within the Turkish-Dutch sample, there was only one difference in alpha reliability across conditions: The alpha of the Honesty-Humility scale was significantly higher in the Turkish-Dutch reference-group condition than in the no reference-group condition (.87 vs. .77).

**Mean differences between the Dutch and Turkish-Dutch sample**

PNC findings suggest that our Turkish-Dutch participants score lower on Conscientiousness, Openness to Experience, and possibly Honesty-Humility, and higher on Neuroticism than the Dutch majority. An independent one-sample t-test revealed that there were no score differences between the Dutch majority and the Turkish-Dutch minority in the no reference-group conditions (see Table 1). This result is not in line with
PNC findings but was expected given that reference-group effects may confound ethnic comparisons. Furthermore, Table 1 also shows that there were no score differences between the Dutch majority and the Turkish-Dutch when the latter compared themselves with people from their own Turkish-Dutch minority group. This is also not in line with PNC findings but was again expected, given that when people compare themselves with similar others, that is, people from their own ethnic group (in-group comparison), true personality differences between ethnic groups may be masked. At last, we compared the personality scores between the Dutch majority and the Turkish-Dutch minority when the latter compared themselves with the Dutch majority group. Our first hypothesis stated that Turkish-Dutch minorities who compare themselves with a Dutch majority out-group will show lower scores on Conscientiousness, Openness to Experience, and Honesty-Humility, and higher scores on Neuroticism (cf. Emotionality) than Dutch majority members (Hypothesis 1). The results showed that there was one significant score difference, namely on Honesty-Humility, with the Turkish-Dutch scoring lower that the majority members ($t = -2.04, p < .05$). In other words, when Turkish-Dutch minorities thought about how they behave in comparison to the Dutch majority group when responding to each personality statement, they see themselves as less honest and humble than the Dutch majority. Thus, our first hypothesis was only confirmed with regard to Honesty-Humility.

**Mean differences across conditions within the Turkish-Dutch sample**

In order to further examine RGEs, we compared the means of the personality domain scales within the Turkish-Dutch sample across the three different conditions using paired-samples $t$-tests. First, we expected that in the no-reference condition Turkish-Dutch would compare themselves with people from the own ethnic group (in-group comparison) and therefore we expected that there were no score differences between the no-reference condition and the Turkish-Dutch reference-group condition. However, the results showed that compared to the no reference-group condition, the Turkish-Dutch minority reference-group condition showed significantly lower scores on the domain scales Extraversion ($t = 2.06, p < .05$) and Conscientiousness ($t = 3.45, p < .01$).

Second, we expected that Turkish-Dutch minorities in the Dutch majority reference-group condition showed lower scores on Conscientiousness, Openness to Experience, and Honesty-Humility, and higher scores on Neuroticism (cf. Emotionality) than in the no reference-group condition and in the Turkish-Dutch reference-group condition (Hypothesis 2). The Dutch majority reference-group condition indeed yielded significantly lower scores than the no reference-group condition with regard to Honesty-Humility ($t = 1.74, p < .05$) and Conscientiousness ($t = 3.21, p < .01$). Although not expected, we also found significantly lower scores on the factor scales Extraversion ($t = 2.66, p < .01$) and Agreeableness ($t = 2.45, p < .01$) in the Dutch majority reference-group condition compared to the no reference-group condition.
Furthermore, in the Dutch reference-group condition significantly lower scores on Openness to Experience ($t = 1.90, p < .05$) and Agreeableness ($t = 2.24, p < .05$), and significantly higher scores on Emotionality ($t = -2.33, p < .05$) were observed than in the Turkish-Dutch reference-group condition. The score differences with respect to Openness to Experience and Emotionality were as expected. These findings do suggest that an in-group comparison and an out-group comparison result in different personality profiles. In other words, when Turkish-Dutch minorities thought about how they behave in comparison to the Dutch majority group (out-group comparison) when responding to each personality statement, they see themselves as more emotional as well as less agreeable and less open for new experiences than when they compared themselves with people from their own Turkish-Dutch minority group (in-group comparison).

**Integration, ethnic identity, and national identity**

In addition, we examined whether the level of integration, ethnic (i.e., Turkish-Dutch) identity, and national (i.e., Dutch) identity explained possible unexpected differences in the Turkish-Dutch sample between the no reference-group condition on the one hand and the Dutch majority reference-group condition and the Turkish-Dutch reference-group condition on the other. In order to test this, we computed new variables by subtracting the mean personality scores in the no reference-group condition from (1) the Dutch majority reference-group condition and (2) the Turkish-Dutch reference-group condition. Subsequently, we correlated these new variables with the level of integration, ethnic identity, and national identity.

First, we expected that in the no reference-group condition the Turkish-Dutch minorities compared themselves with similar others, that is, people from their own Turkish-Dutch ethnic group. Therefore, we did not expect to find score differences between the no reference-group condition and the Turkish-Dutch reference-group condition, given that the same comparison group may be used. However, the abovementioned results showed that compared to the no reference-group condition, the Turkish-Dutch minority reference-group condition revealed significantly lower scores on the domain scales Extraversion and Conscientiousness. Therefore, we checked whether the level of integration, ethnic identity, and national identity may explain these two differences. For example, these score differences may occur when Turkish-Dutch minorities are to a large extent integrated into the Dutch culture, have a weak ethnic identity and/or a strong national identity (i.e., they compared themselves with the Dutch majority group in the no-reference condition). However, the results (which can be obtained from the first author) did not offer support for this assumption.
Table 1
Reliabilities ($\alpha$), means ($M$), and standard deviations (SD) of the personality domains across conditions in the Dutch and the Turkish-Dutch sample

<table>
<thead>
<tr>
<th></th>
<th>Dutch majority</th>
<th>No reference-group</th>
<th>Turkish-Dutch minority</th>
<th>Dutch majority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n=74</td>
<td>n=95</td>
<td>n=52</td>
</tr>
<tr>
<td>HEX: H</td>
<td>$\alpha_1$</td>
<td>.74</td>
<td>.77</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>$M_1$</td>
<td>3.64</td>
<td>3.62</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>$SD_1$</td>
<td>0.44</td>
<td>0.48</td>
<td>0.51</td>
</tr>
<tr>
<td>HEX: E</td>
<td>$\alpha_2$</td>
<td>.76</td>
<td>.78</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>$M_2$</td>
<td>3.19</td>
<td>3.06</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>$SD_2$</td>
<td>0.47</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td>HEX: X</td>
<td>$\alpha_3$</td>
<td>.84</td>
<td>.81</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>$M_3$</td>
<td>3.56</td>
<td>3.68</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>$SD_3$</td>
<td>0.48</td>
<td>0.46</td>
<td>0.43</td>
</tr>
<tr>
<td>HEX: A</td>
<td>$\alpha_4$</td>
<td>.86</td>
<td>.73</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>$M_4$</td>
<td>2.89</td>
<td>2.89</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td>$SD_4$</td>
<td>0.56</td>
<td>0.45</td>
<td>0.40</td>
</tr>
<tr>
<td>HEX: O</td>
<td></td>
<td>.82</td>
<td>.71</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>$M_5$</td>
<td>3.68</td>
<td>3.67</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>$SD_5$</td>
<td>0.51</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>HEX: C</td>
<td></td>
<td>.83</td>
<td>.71</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>$M_6$</td>
<td>3.38</td>
<td>3.43</td>
<td>3.42</td>
</tr>
<tr>
<td></td>
<td>$SD_6$</td>
<td>0.56</td>
<td>0.46</td>
<td>0.47</td>
</tr>
</tbody>
</table>

$t_{1.2}$, $t_{1.3}$, and $t_{1.4}$ values indicate lower scale scores for the no reference-group condition.

* $p < .05$

Notes: HEXACO-100 (HEX) scales are Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O); positive $t_{1.2}$, $t_{1.3}$, and $t_{1.4}$-values indicate lower scale scores for the no reference-group condition.
### Table 2
**Reliabilities (α), means (M), and standard deviations (SD) of the personality domains across conditions within the Turkish-Dutch sample (n=34)**

<table>
<thead>
<tr>
<th></th>
<th>No reference-group</th>
<th>Turkish-Dutch reference-group</th>
<th>Dutch majority reference-group</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α₁</td>
<td>M₁</td>
<td>SD₁</td>
<td>α₂</td>
<td>M₂</td>
<td>SD₂</td>
<td>α₃</td>
</tr>
<tr>
<td>HEX: H</td>
<td>.77</td>
<td>3.66</td>
<td>0.49</td>
<td>.83</td>
<td>3.58</td>
<td>0.56</td>
<td>.88</td>
</tr>
<tr>
<td>HEX: E</td>
<td>.83</td>
<td>3.11</td>
<td>0.59</td>
<td>.75</td>
<td>3.05</td>
<td>0.46</td>
<td>.80</td>
</tr>
<tr>
<td>HEX: X</td>
<td>.82</td>
<td>3.68</td>
<td>0.49</td>
<td>.78</td>
<td>3.56</td>
<td>0.45</td>
<td>.83</td>
</tr>
<tr>
<td>HEX: A</td>
<td>.74</td>
<td>2.90</td>
<td>0.46</td>
<td>.64</td>
<td>2.88</td>
<td>0.40</td>
<td>.72</td>
</tr>
<tr>
<td>HEX: C</td>
<td>.59</td>
<td>3.77</td>
<td>0.35</td>
<td>.67</td>
<td>3.63</td>
<td>0.38</td>
<td>.71</td>
</tr>
<tr>
<td>HEX: O</td>
<td>.70</td>
<td>3.41</td>
<td>0.48</td>
<td>.76</td>
<td>3.46</td>
<td>0.48</td>
<td>.74</td>
</tr>
</tbody>
</table>

Notes: HEXACO-100 (HEX) scales are Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O); positive $t_{1-2}$-values and $t_{1-3}$-values indicate higher scale scores for the no reference-group condition; positive $t_{2-3}$-values indicate higher scale scores for the Turkish-Dutch reference-group condition.

* $p < .05$; ** $p < .01$
Second, as mentioned before, we expected that in the no reference-group condition the Turkish-Dutch minorities compared themselves with people from their own Turkish-Dutch ethnic group. Therefore, we did expect to find score differences on the domain scales Conscientiousness, Neuroticism (cf. Emotionality), Openness to Experience, and Honesty-Humility between the no reference-group condition and the Dutch majority reference-group condition, as different comparison groups may be used. The abovementioned results showed that compared to the no reference-group condition, the Dutch majority reference-group condition revealed significantly lower scores on the domain scales Honesty-Humility, Extraversion, Agreeableness, and Conscientiousness. However, the lack of score differences between the no reference-group condition and the Dutch reference group on the domain scales Openness to Experience and Emotionality may occur when Turkish-Dutch minorities are to a large extent integrated into the Dutch culture, have a weak ethnic identity and/or a strong national identity (i.e., they compared themselves with the Dutch majority group in the no-reference condition). Again, the results did not offer support for this assumption.

Discussion
Several studies have shown no relationship between national personality profiles (based on self- and observer personality reports) and perceived national character (PNC; Terracciano et al., 2005). Although some scholars (e.g., McCrae et al., 2007) have argued that lack of findings are due to the fact that perceptions of national character may be based on unfounded stereotypes, others (e.g., Heine et al., 2008) claim that an absence of findings is the result of the RGE on the assessment of self-reported personality. Until now, no studies have been conducted to test the RGE in personality judgments with different ethnic reference groups. Therefore, our goal was to examine to what extent Turkish-Dutch members are influenced by perceptions of comparison others when filling out a personality test.

Reference-group effects
The RGE is described as the influence of comparison others on the responses to individual self-report items and occurs when responses to self-report items are based not on respondents’ absolute level of a construct but rather on their level relative to a comparison group (Credé et al., 2010). According to Heine et al. (2002), people from different cultures adopt different standards when evaluating themselves on subjective ordinal or (quasi-)interval scales. Comparing measures with such scales conceals cultural differences. Until now, the RGE has primarily been found to influence responses to measures of individualism–collectivism (e.g., Heine et al., 2002). However, Credé et al. (2010) were recently the first to examine the RGE in self-report measures of personality, using different age, gender, and relative reference groups. Yet, no studies have been conducted to test the RGE in personality judgments with different ethnic reference
groups. Moreover, it is not known what, if any, effects are to be expected, especially among ethnic groups who live in the same country.

Although Heine et al. (2002) argued that reference-group effects confound cross-national comparisons more than within-country comparisons, as minority members may have the same referents than majorities members (in our case Dutch majority referents), we on the other hand expected that in general Turkish-Dutch minorities compare themselves with similar others, that is, people from their own Turkish-Dutch ethnic group which may indeed result in the masking of expected personality differences between Turkish-Dutch minorities and Dutch majorities. In contrast but in line with Heine et al., we did expect to find significant personality differences between Turkish-Dutch minorities and Dutch majorities when Turkish-Dutch are instructed to compare themselves with the Dutch majority (out-group comparison), as both groups may used the same reference group. Therefore, we examined to what extent Turkish-Dutch members are influenced by perceptions of comparison others when filling out a personality test.

Prior PNC findings may suggest that Turkish-Dutch participants score lower on Conscientiousness, Openness to Experience, and possibly Honesty-Humility, and higher on Neuroticism than the Dutch majority. Our results showed there were no score differences between the Dutch majority and the Turkish-Dutch minority in the no reference-group conditions. This result is not in line with earlier PNC findings but may be consistent with the notion that reference-group effects confound ethnic comparisons. Furthermore, there also were no score differences between the Dutch majority and the Turkish-Dutch when the latter compared themselves with people from their own Turkish-Dutch minority group. This is also not in line with PNC findings but was expected, given that when people compare themselves with similar others, that is, people from their own ethnic group (in-group comparison), true personality differences between ethnic groups may be masked. Our expectation was that when an out-group comparison is made, personality scores will move towards the PNC of the Turkish-Dutch, i.e., Turkish-Dutch participants will score lower on Conscientiousness, Openness to Experience, and possibly Honesty-Humility, and higher on Neuroticism than the Dutch majority. The results showed that when Turkish-Dutch minorities thought about how they behave in comparison to the Dutch majority group when responding to each personality statement, they see themselves as less honest and humble than the Dutch majority.

We further examined RGEs within the Turkish-Dutch sample across different conditions. The results seem to suggest an in-group comparison and an out-group comparison indeed result in different personality profiles. For instance, when Turkish-Dutch minorities thought about how they behave in comparison to the Dutch majority group (out-group comparison) when responding to each personality statement, they see themselves as more emotional as well as less agreeable and less open for new experiences than when they compared themselves with people from their own Turkish-Dutch minority group.
group (in-group comparison). To conclude, Turkish-Dutch members seem to be influenced by perceptions of comparison others when filling out a personality test.

**Limitations and future research**

There are several limitations of the present study. First, one important limitation is the sample size of the Turkish-Dutch minorities. The final sample consisted of 95 Turkish-Dutch minority participants, but only 34 completed every version. The difficulty was that we focused on a homogenous ethnic group and employed a within-subject design. Second, we assumed that there are PNC-differences on Conscientiousness, Neuroticism, Openness to Experience, and possibly on Honesty-Humility between The Netherlands and Turkey. However, we presumed that the PNC of Turkey also applied for Turkish-Dutch minorities. Future research should examine whether this is the case.

It is important to note, although the current study did not pay special attention to this, that the RGE may be substantially different for ethnically diverse people from different countries than for ethnically diverse people who live in the same country, given that the latter may compare themselves with other ethnic minorities from the same background and/or with ethnic majorities. Although in research across countries, lack of mean personality differences may signal reference-group effects, there may be other processes at work when investigating ethnic minority members within one country. On the one hand, strong in-group networks may prevent the formation of out-group cross-ethnic contacts and thus the inclination to compare oneself with other ethnic groups. On the other hand, group animosity and competition may enhance perceived differences observed between groups and may cause increased self-stereotyping (Sinclair, Hardin, & Lowery, 2006). In a study by Brown and Day (2006), ethnic minority African Americans were found to score significantly lower on cognitive ability in a threatening situation (equal to those occurring in selection) and equal to ethnic majority Americans in a low threat situation. One of the reasons for this effect may be that in threatening situations out-group expectations become more salient for ethnic minority members, which may result in lower performance for ethnic minority members but not for ethnic majority members. In personality research, the amount of threat may change the nature of the comparison other, which in turn may influence the type of response on the personality questionnaire. Future research could examine whether minority members will be more strongly inclined to compare themselves to out-group others in a selection situation than in a less threatening environment, such as a counseling situation.

Interestingly, results of the study by De Meijer et al. (2006) suggest that the reference-group effect may occur in varying degrees in sub-samples of ethnic minority members. De Meijer et al. found strong differences between personality profiles of first-generation ethnic minority members and majority members and greatly reduced differences between second-generation ethnic minority members and majority members.
A reason for this difference may be that first-generation ethnic minority members, being exposed to a different culture, are more likely to self-stereotype and thus show enhanced instead of diminished differences in personality profiles from majority members. Instead, second-generation ethnic minority members, although still exposed to the ethnic minority culture, may be less likely to self-stereotype because they are more familiar with the host culture. More research should be conducted to investigate to what extent first- versus second-generation Turkish-Dutch members are influenced by perceptions of comparison others when filling out a personality test.

One of the goals of the current study was to examine whether the RGE was an explanation for the fact that prior research revealed that personality tests only produce small ethnic score differences. However, another explanation for the small ethnic score differences on personality tests may lie with the fact that prior research on ethnicity effects in personality testing primarily focused on broad factor-level traits, such as the Big Five or HEXACO factors, while often failing to focus on narrow traits, such as the facets of the Big Five or HEXACO factors. Several studies found that analyses on the facet-level showed moderate group differences, but that the use of broad personality factors to compare groups appeared to mask these differences (Foldes, Duehr, & Ones, 2008; Van Iddekinge, Taylor, & Eidson, 2005). Future research is encouraged to examine the RGE on the factor- as well as on the facet level.

**Practical implications**

The results of the current study suggest that the use of reference groups to some extent influences the manner in which Turkish-Dutch minorities respond to personality tests. Credé et al. (2010), however, have shown that the use of reference groups may subsequently result in significant decreases in the criterion-related validity, as compared with personality tests that do not specify a reference group. They explain this finding by arguing that respondents are unlikely to perceive the reference group similarly. In our case, the perception of an out-group (that is, the Dutch majority) reference group may be significantly different for one person with a Turkish-Dutch minority background compared to another with the same background, as both persons do not exactly have the same Dutch majority members in their environment. This would suggest that personality tests that do not specify a reference group are to be preferred over personality inventories that do identify a particular reference group. Credé et al. advocate that an alternative possibility would be to refer to natural reference groups. For example, in selection situations they suggest that applicants should fill out a personality inventory in a work setting, which may consequently activate a reference group consisting of employees. This leads us to the positive effects of adding a specific context (frame-of-reference) to personality items. Previous empirical studies confirm the positive effects of providing a specific context in personality tests. In educational settings, adding the tag “at school” to
Conscientiousness items has been shown to increase the prediction of grade point average (Bing, Whanger, Davison, & VanHook, 2004; Lievens, De Corte, & Schollaert, 2008). Moreover, Conscientiousness items with an irrelevant “at work” tag showed lower predictive validities in the prediction of GPA (Lievens et al., 2008). Therefore, instead of identifying a particular reference group, a better way to improve the predictive validity of personality measures is adding a specific context to each personality item.

Conclusion
Until now, no studies have been conducted to test the RGE in personality judgments with different ethnic reference groups. Therefore, our goal was to examine to what extent Turkish-Dutch members are influenced by perceptions of comparison others when filling out a personality test. The results show that when the Turkish-Dutch compared themselves with people from their own Turkish-Dutch minority group (in-group comparison), there were no score differences between the Dutch majority and the Turkish-Dutch minority. This was expected, given that when people compare themselves with similar others, that is, people from their own ethnic group (in-group comparison), true personality differences between ethnic groups may be masked. Yet, when Turkish-Dutch minorities thought about how they behave in comparison to the Dutch majority group (out-group comparison), they saw themselves as less honest and humble. Furthermore, when the Turkish-Dutch used an out-group comparison other, they saw themselves as more emotional as well as less agreeable and less open for new experiences than when they used an in-group comparison other. The findings do suggest that Turkish-Dutch members are influenced by perceptions of comparison others when filling out a personality test.