

## Sex Differences in Personality in Saudi Arabia

Dr. Moneerah Mohammad AlMerab  
*Princess Nourah Bint Abdulrahman University, Riyadh,  
Saudi Arabia*

Edward Dutton  
*Asbiro University, Łódź, Poland*

David Becker  
*Chemnitz University of Technology, Germany*

Salaheldin Farah Bakhiet,\* Yossry Ahmed Sayed Essa  
*King Saud University, Riyadh, Saudi Arabia*

Gerhard Meisenberg  
*Ulster Institute of Social Research, London, UK*

\*Corresponding author: King Saud University, Department of Special Education, College of Education, Saudi Arabia; email: bakhiet@ksu.edu.sa

Research has consistently found that there are sex differences in self-reported personality traits. Studies have also found that there are intercultural variations in the magnitude of these sex-based personality differences, with these differences being the most pronounced in prosperous, gender-egalitarian Western societies. In this study, we contribute to this body of research by exploring sex-related personality differences on the HEXACO personality inventory in a large student sample in Saudi Arabia. Consistent with previous research, we find that sex differences in personality are of rather small magnitude, in line with the Socioecological Complexity Hypothesis whereby personality differentiation is found to be more pronounced the more complex a society is. We also report substantial sex differences in response styles, with acquiescent responding and evidence of “careless responding” higher in males than in females.

**Key Words:** Big 5; HEXACO-PI-R; Trait covariation; Saudi Arabia; Socioecological complexity; Sex/gender differences; Response styles; Acquiescence

Studies in developed countries tend to find that there are consistent sex differences in average personality, based on the Big 5 personality traits. Females typically score higher than males on *Agreeableness*, *Conscientiousness*, *Extraversion* and *Neuroticism*, though sometimes the difference in *Conscientiousness* has only been found at the aspect level (e.g. Weisberg et al., 2011). These sex differences have been observed with both student (e.g. Weisberg et al., 2011) and population (e.g. Yamagata et al., 2006) samples in developed countries. However, the situation is not so clear cut when it comes to developing countries.

Church (2016) has reviewed research finding that the validity of the Big 5 model is particularly difficult to replicate in poorly educated or indigenous peoples, where the validity and trait-related consistency tend to be at best mixed. Meisenberg (2015) has shown that it is very difficult to compare countries in terms of the Big 5. Doing so leads to unexpected results, such as that the Japanese are particularly low in *Conscientiousness*. This may be due to reference group effects (Heine et al., 2002), cultural differences in how the items are understood, and also possible intercultural differences in self-esteem.

Cohen and Deuling (2014) have found when comparing different ethnic groups within France the Big 5 model was broadly useful but that *Conscientiousness* and *Agreeableness* were the least stable. Schmitt et al. (2008) have reviewed sex differences in personality across 55 nations. They found that, overall, women scored higher on *Agreeableness*, *Conscientiousness*, *Neuroticism* and *Extraversion*. These sex differences were larger the more prosperous, healthy and egalitarian the cultures were. In some instances, in particularly poor and conservative cultures, the sex differences were actually reversed, such as in Botswana and Indonesia where males score higher in *Neuroticism* than females. The same phenomenon was found by Costa et al. (2001) across 26 cultures: The differences were largest in Europe and North America. This is particularly surprising because, from the viewpoint of evolutionary psychology, we would expect there to be consistent sex differences in personality. Indeed, these seem to exist in relatively prosperous countries in North America, Europe and Asia (Yamagata et al., 2006).

The HEXACO model of personality is a more recent alternative to the Big Five. It shares with the Big Five the dimensions of *Extraversion* (X), *Agreeableness* (A), *Conscientiousness* (C), and *Openness* (O). Big Five *Neuroticism* (N) is represented by *Emotionality* (E), and *Honesty-humility* (H) is added as a sixth dimension. H can be considered the opposite pole of the Dark Triad traits (*Narcissism*, *Machiavellianism*, *Psychopathy*; Hodson et al., 2018). Sex differences on the HEXACO scales have recently been reported from an

online survey in 48 countries (Lee & Ashton, 2020). This survey found sex differences with higher scores for females on E (close to one standard deviation) and H (approximately 0.4 standard deviations) in most countries. Sex differences on the other scales were smaller and more variable.

In this study, we aim to extend this body of research by exploring personality differences in a student sample from two large universities in a relatively prosperous but also culturally conservative society: the Kingdom of Saudi Arabia. Based on the research reviewed above, we will argue that this is potentially in line with Lukaszewski et al.'s (2017) Socioecological Complexity Hypothesis.

In addition to sex differences on the six HEXACO scales and their facets, we will also examine sex differences in response styles. Response styles, and especially acquiescent responding (the tendency to agree with statements) can change the mean level of the measured constructs, alter the psychometric properties of the measuring scales, and affect their factorial structure (Goldammer et al., 2020; Kam & Meyer, 2015). We know that response styles show variations across countries (Meisenberg & Williams, 2008), but we do not know whether there are sex differences in response styles and whether they have any consequences for sex differences in the substantive constructs that are measured by personality tests.

## **Method**

We used the HEXACO–PI–R (Self-Report Form) from Ashton and Lee (2007, 2009), translated into Arabic. It includes 60 single items, each in form of a short statement which had to be rated on a 1 (strongly disagree) to 5 (strongly agree) scale. From these 60 items, 24 facets can be calculated by averaging the scores from two or three items each, with the scores on some of the items being reversed (R).<sup>1</sup> Combinations of these facets produce six HEXACO–60 scales as follows:

### ***Honesty/humility (H)***

*Sincerity* (6, 30R, 54)

*Fairness* (12R, 36, 60R)

*Greed-Avoidance* (18, 42R)

*Modesty* (24R, 48R)

---

<sup>1</sup> There are 29 reverse-scored items in the English HEXACO. Due to a translation error, the Arabic version used in the present study had the meaning of item 35 inverted, such that it was scored without reversal. Thus, there are 28 reverse-scored items on the Arabic test.

**Emotionality (E)**

*Fearfulness* (5, 29, 53R)

*Anxiety* (11, 35)

*Dependence* (17, 41R)

*Sentimentality* (23, 47, 59R)

**Extraversion (X)**

*Social Self-Esteem* (4, 28R, 52R)

*Social Boldness* (10R, 34, 58)

*Sociability* (16, 40)

*Liveliness* (22, 46R)

**Agreeableness versus Anger (A)**

*Forgiveness* (3, 27)

*Gentleness* (9R, 33, 51)

*Flexibility* (15R, 39, 57R)

*Patience* (21R, 45R)

**Conscientiousness (C)**

*Organization* (2, 26R)

*Diligence* (8, 32R)

*Perfectionism* (14R, 38, 50)

*Prudence* (20R, 44R, 56R)

**Openness to Experience (O)**

*Aesthetic appreciation* (1R, 25)

*Inquisitiveness* (7, 31R)

*Creativity* (13, 37, 49)

*Unconventionality* (19R, 43, 55R)

The full sample consisted of 2,831 randomly selected Saudi Arabian students in higher education aged between 15 and 46 ( $M=22.05$ ;  $SD=2.88$ ) from classes between the 1<sup>st</sup> and 12<sup>th</sup> ( $M=4.79$ ;  $SD=2.03$ ). They were 43.3% male and 56.7% female. The majority (2,144, 75.7%) came from King Saud University (KSU), and the remaining 687 (24.3%) were from Al-Imam Mohammad Ibn Saud Islamic University (IMAM). The processing time was 10 to 20 minutes. Complete replies were available from 2,818 participants, after excluding some with missing or miscoded responses. Data were analyzed using SPSS.

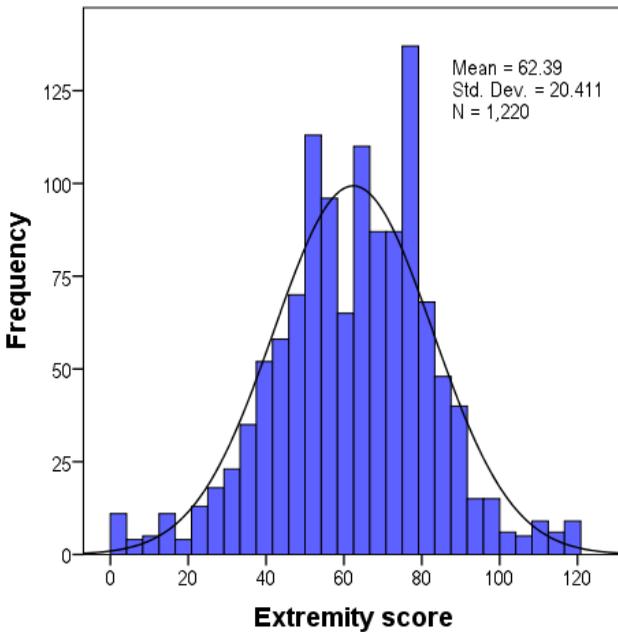
**Results**

1. *Extreme responding*

Data quality is compromised when many respondents have a strong bias to either choose the midpoint of the scale for most responses, or to strongly prefer

the extremes (Strongly agree, Strongly disagree) over the more moderate responses (Agree, Disagree). To quantify this response style, an extremity measure was formed by recoding all 60 items such that *Neutral* became 0, *Agree* and *Disagree* 1, and *Strongly agree* and *Strongly disagree* became 2. The sum of these recoded scores was the measure of extreme responding.

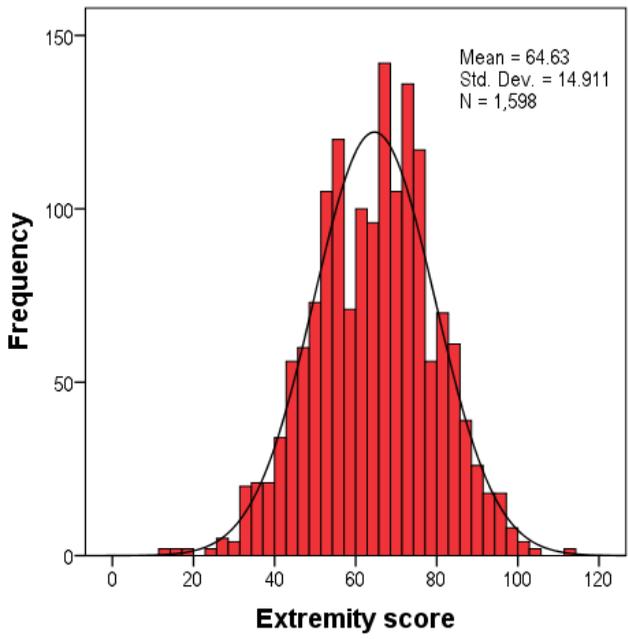
Sex differences can potentially manifest as a general tendency of one sex to respond at the midpoint of the scale, and of the other to respond at the extremes. In this case, the mean of the extremity score distribution will be different for males and females. A second possibility is that one sex has more individuals than the other who respond mainly at the midpoint of the scale, and also more individuals who respond at the extremes. In this case the mean of the extremity score distribution is not necessarily different, but the standard deviation is larger for the sex with the more “unbalanced” responding. Figures 1 and 2 show the results.



**Figure 1.** *Extremity of responding for 1,220 males.*

The figures show that the mean of the distribution is higher in females than in males. The difference of 0.11 male standard deviations or 0.15 female standard deviations is statistically significant at  $p = .001$ , but it should be considered very small. The more important sex difference is in the standard deviation, which is

20.4 in males but only 14.9 in females. Levene’s test for equality of variances produced an F value of 101.7, which is statistically significant at  $p < .001$ . This means that males are substantially overrepresented both among those who habitually choose the midpoint of the scale, and among those who habitually choose the extremes. At the far ends, 5/1,220 males but none of the 1,598 females consistently answered all 60 questions at the midpoint of the scale, and 7 males but none of the females chose an extreme response every time. This indicates that in this sample, “careless responding” (Kam & Meyer, 2015) is much more frequent in males than in females.



**Figure 2.** *Extremity of responding for 1,598 females.*

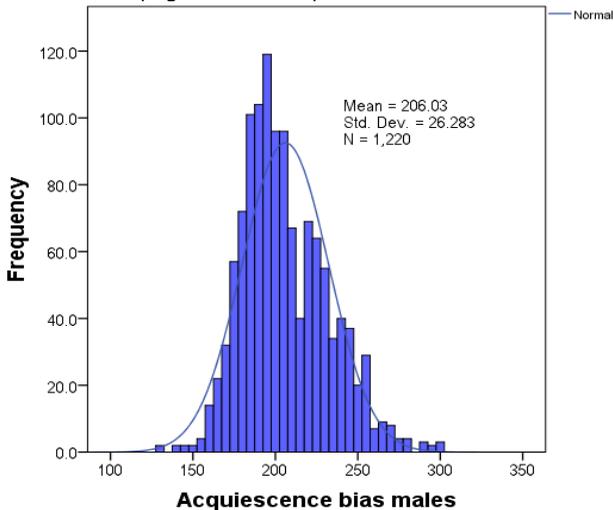
When the responses from KSU and IMAM were analyzed separately, a very similar sex difference in standard deviations was found: male 20.1, female 14.7 at KSU; and 19.9 male, 15.1 female at IMAM. However, the mean was slightly higher for females than males at KSU (male 60.3, female 63.4) but at IMAM the males were slightly more extreme (male 69.6, female 68.1). The difference in means between the two universities, males and females combined, is statistically significant ( $t = 8.8, p < .001$ ).

2. *Acquiescent responding*

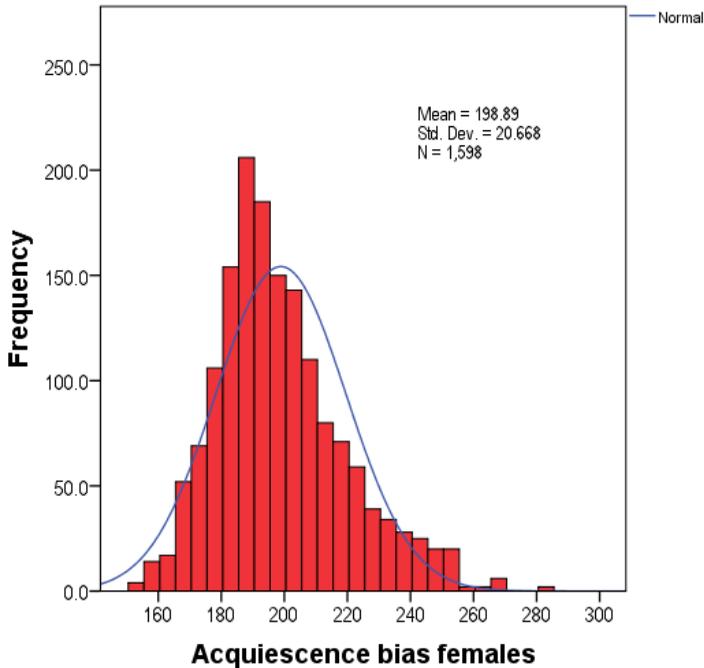
Acquiescence bias, also called agreement bias or acquiescent response style, is defined as the bias to choose the *Agree* over the *Disagree* response — or, conversely, the tendency to disagree rather than agree with offered statements in the case of negative acquiescence bias. It can affect sex differences when the number of reverse-scored items on a scale is different from the number of direct-scored items. In the present case, we have a 60-item scale with answer choices ranging from 1 to 5 for each item. The sum of scores on the unreversed total scale can range between 60 (strongly disagree with everything) and 300 (strongly agree with everything), and in the absence of acquiescence bias, scores would cluster around the midpoint of 180. Thus, a mean score of 180 is expected.

The actual mean of the summed unreversed scores is  $198.9 \pm 20.7$  for females, and  $206.0 \pm 26.3$  for males. Sex differences are highly significant for both the mean ( $t = 8.1$ ) and the standard deviation ( $F = 85.2$ ), both with  $p < .001$ . Thus, both sexes tend to be “conformist”, but males are more conformist than females and are more variable in this respect.

In total, 6/1,220 males (0.5%) and none of the females had an acquiescence score below 150; and 451/1,220 males (37.0%) and 388/1598 females (24.3%) had an acquiescence score above 210 (more than 1.3 standard deviations away from the theoretical mean). These were deemed to have substantial acquiescence bias (Figures 3 and 4).



**Figure 3.** *Distribution of the summed unreversed scores (“acquiescence bias score”) in males.*



**Figure 4.** *Distribution of the summed unreversed scores (“acquiescence bias score”) in females.*

The consequences of acquiescent responding can be seen in the correlations between the direct scored and reverse scored halves of the six scales. After score reversal, these are expected to be positively correlated if individual differences in responses are determined primarily by question content, but negatively if they are determined mainly by acquiescent responding. Table 1 shows the results. In the complete male sample, the correlations between the half-scales (after reversal of the reverse-scored items) are negative rather than positive, indicating that acquiescence is more important than substantive item content for the responses. The correlations switch from negative to positive after removal of those with acquiescence scores below 150 or above 210. All correlations are more positive in females than in males, as is expected based on their lesser tendency for acquiescent responding.

Because acquiescence bias can have profound effects on the properties of the measurement instrument and the obtained results, most of the following analyses were done in duplicate, both with and without exclusion of those with strong acquiescence bias.

**Table 1.** Correlations between direct-scored and reverse-scored parts of the 6 HEXACO scales, before and after exclusion of those with strong acquiescence bias. Shown are the correlations between the half-scales after score reversal. The “reduced” samples have those with acquiescence scores below 150 and above 210 removed.

|   | Males                        |                           | Females                      |                             |
|---|------------------------------|---------------------------|------------------------------|-----------------------------|
|   | Complete sample<br>(N=1,220) | Reduced sample<br>(N=763) | Complete sample<br>(N=1,598) | Reduced sample<br>(N=1,210) |
| H | -.253                        | .175                      | .047                         | .346                        |
| E | -.204                        | .196                      | .114                         | .374                        |
| X | -.183                        | .231                      | .069                         | .336                        |
| A | -.184                        | .300                      | .031                         | .323                        |
| C | -.132                        | .187                      | .043                         | .292                        |
| O | -.170                        | .176                      | .110                         | .285                        |

### 3. Reliability

Table 2 shows the alpha reliabilities of the six HEXACO scales for the complete sample and the reduced sample after removal of those with strong acquiescence bias. Several observations can be made:

1. All reliabilities are low, even considering that each scale has only 10 items. Overall, Cronbach’s alphas are below the norm as reported by Ashton and Lee (2009). However, lower reliabilities are acceptable because the scales are heterogeneous, each having 4 facets defined by 2 or 3 items each.
2. Removal of those with strong acquiescence bias raises the reliabilities substantially.
3. Reliabilities are higher for females than males on all scales in the complete sample, and on 5 of the 6 scales even after removal of those with strong acquiescence bias. Therefore, what appears to be “careless responding” is more frequent in males than in females.

**Table 2.** Cronbach’s  $\alpha$  for the complete sample and the sample after removal of those with strong acquiescent responding (“reduced sample”).

| Scales | Complete sample        |                       |                         | Reduced sample    |                 |                    |
|--------|------------------------|-----------------------|-------------------------|-------------------|-----------------|--------------------|
|        | Total<br>(N=2825-2831) | Male<br>(N=1221-1225) | Female<br>(N=1602-1606) | Total<br>(N=1973) | Male<br>(N=763) | Female<br>(N=1210) |
| H      | .437                   | .359                  | .472                    | .484              | .430            | .490               |
| E      | .467                   | .366                  | .512                    | .565              | .430            | .600               |
| X      | .427                   | .328                  | .496                    | .550              | .488            | .585               |
| A      | .374                   | .347                  | .386                    | .496              | .518            | .480               |
| C      | .459                   | .423                  | .485                    | .507              | .465            | .531               |
| O      | .410                   | .269                  | .488                    | .508              | .413            | .553               |

4. Correlations among scales

A test that is meant to measure a “pure” construct is expected to have not only high internal consistency, but also low overlap with scales measuring different constructs. Table 3 shows that in this Saudi sample, the correlations between the six HEXACO scales are indeed low. They are similar for males and females, but with a tendency to be higher in males than in females. Average inter-scale correlation is .081 for females and .150 for males. Also included is the General Factor of Personality (GFP), defined as the unrotated first principal component of an item-level principal components analysis of the entire inventory. The table also shows the response styles. The correlations of acquiescent response style (AgBias in Table 3) depend on the number of reverse-scored items in the scale: positive for E (3 reversals), and negative for H and C (6 reversals each).

**Table 3.** Correlations among HEXACO scales for 763 males and 1,210 females, after removal of those with strong acquiescence bias. Females shown above and males below the diagonal. GFP, general factor of personality; ExBias, extreme responding; AgBias, agreement bias (acquiescence); \*  $p < .05$ , \*\*  $p < .01$ .

|        | H       | E       | X      | A      | C       | O       | GFP    | ExBias | AgBias  |
|--------|---------|---------|--------|--------|---------|---------|--------|--------|---------|
| H      | -       | -.081** | .014   | .205** | .148**  | -.006   | .449** | -.007  | -.260** |
| E      | -.005   | -       | .022   | .039   | -.071*  | -.076** | -.008  | .065*  | .317**  |
| X      | .308**  | -.087*  | -      | .056   | .131**  | .202**  | .591** | .132** | .133**  |
| A      | .215**  | .046    | .215** | -      | .013    | -.011   | .327** | .051   | -.031   |
| C      | .343**  | -.107** | .370** | .179** | -       | .140**  | .569** | .164** | -.181** |
| O      | -.021   | .031    | .199** | -.008  | .114**  | -       | .417** | .156** | .027    |
| GFP    | .604**  | .001    | .741** | .393** | .668**  | .357**  | -      | .228** | -.121** |
| ExBias | .170**  | -.009   | .175** | .058   | .082*   | .091*   | .264** | -      | .127**  |
| AgBias | -.173** | .242**  | .062   | .084*  | -.130** | .069    | -.043  | .117** | -       |

5. Sex differences on the HEXACO scales

Descriptives for the mean scores and standard deviations are presented in Table 4. All sex differences are small. The largest is on *Emotionality* (E), where females score substantially higher than males. Higher female than male scores are also seen for *Honesty/humility* (H), *Conscientiousness* (C), and *Openness* (O), but males score higher than females on *Agreeableness* (A). *Extraversion* (X) shows no statistically significant sex differences in either the complete or the reduced sample. Comparing the complete and reduced samples, sex differences in the means are similar but different in magnitude. In addition to sex differences

in means, we observe higher standard deviations in females than males except for *Agreeableness* in the reduced sample.

**Table 4.** Descriptives for the six HEXACO scales; *d* = Cohen's *d*.

|   | Total sample | Male         | Female       | <i>d</i> |
|---|--------------|--------------|--------------|----------|
| <i>Complete sample</i>                        |              |              |              |          |
| Age (y)                                       | 22.0 ± 2.8   | 22.8 ± 3.6   | 21.4 ± 1.8   |          |
| H   | 31.52 ± 5.22 | 31.00 ± 4.95 | 31.92 ± 5.38 | -0.18*** |
| E   | 32.97 ± 4.92 | 32.16 ± 4.46 | 33.60 ± 5.16 | -0.29*** |
| X   | 33.06 ± 4.62 | 32.93 ± 4.19 | 33.15 ± 4.92 | -0.05    |
| A   | 31.44 ± 4.55 | 32.08 ± 4.46 | 30.95 ± 4.56 | 0.25***  |
| C   | 31.35 ± 4.69 | 30.87 ± 4.43 | 31.71 ± 4.85 | -0.18*** |
| O   | 31.45 ± 4.90 | 30.94 ± 4.39 | 31.83 ± 5.23 | -0.18*** |
| <i>Those with strong acquiescence removed</i> |              |              |              |          |
| Age (y)                                       | 21.9 ± 2.7   | 22.8 ± 3.6   | 21.3 ± 1.7   | -0.08    |
| H   | 32.63 ± 5.24 | 32.37 ± 4.98 | 32.80 ± 5.40 | -0.41*** |
| E   | 32.44 ± 5.23 | 31.12 ± 4.49 | 33.28 ± 5.48 | -0.04    |
| X   | 33.18 ± 4.87 | 33.05 ± 4.50 | 33.26 ± 5.09 | 0.21***  |
| A   | 31.09 ± 4.77 | 31.71 ± 4.80 | 30.70 ± 4.72 | -0.10*   |
| C   | 32.35 ± 4.67 | 32.06 ± 4.37 | 32.53 ± 4.85 | -0.23*** |
| O   | 31.56 ± 5.19 | 30.83 ± 4.67 | 32.01 ± 5.45 | -0.08    |

The regression models in Table 5 are for the complete sample. They show that sex differences in the means are robust only for E, A, and O. They also show that age is not an important covariate in this student sample and that, other things being equal, students at KSU score higher on *Conscientiousness* than students at IMAM. Acquiescence bias is again a major determinant of scores depending on the number of score reversals in each scale.

**Table 5.** OLS linear regression models predicting scores on the six HEXACO scales for the complete sample, *N* = 2,818. Standardized betas are shown.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ .

|                            | H        | E        | X       | A       | C        | O        |
|----------------------------|----------|----------|---------|---------|----------|----------|
| Male                       | -.014    | -.181*** | .003    | .114*** | -.022    | -.069*** |
| Age                        | .000     | -.034    | -.033   | .012    | .020     | -.022    |
| IMAM                       | -.018    | .038*    | -.004   | .023    | -.065*** | -.039*   |
| Acquiescence bias          | -.435*** | .303***  | -.063** | .075*** | -.436*** | -.063**  |
| Extremity bias             | .121***  | -.020    | .155*** | .069**  | .150***  | .123***  |
| Adj. <i>R</i> <sup>2</sup> | .156     | .108     | .018    | .030    | .160     | .019     |

## 6. Facet-level sex differences

Each of the six HEXACO scales has four facets, each facet being represented by 2 or 3 items. As shown in Tables 6 and 7, many of these show sex differences, which can be different in direction within each scale. For example, in the O scale the facet Inquisitiveness is higher in males than in females, while females are higher on the other three facets; and within the E scale, only Fear and Sentimentality are consistently higher in females than in males. Anxiety and Dependence have virtually no sex difference. In the complete sample, the average sex difference is 0.1638, and in the reduced sample it is 0.1767. The facet-level scores also confirm that in most cases, females have a higher standard deviation than males, both in the complete sample (mean  $d = 0.14$ ,  $p < .001$ , paired-samples  $t$  test) and in the reduced sample after removal of those with strong acquiescence (mean  $d = 0.15$ ,  $p < .001$ ).

For comparison with a typical Western sample, we included in Table 7 the  $d$  values from a Croatian sample on the 100-item form of the HEXACO, reported in Babarović and Šverko (2013). As can be seen, the sex differences are broadly similar. The correlation between the  $d$  values in Saudi Arabia and those in Croatia is .666. However, sex differences tend to be systematically smaller in Saudi Arabia than in Croatia. In Saudi Arabia, the average absolute sex differences on the 24 facets is 0.177 standard deviations, and in Croatia it is 0.352 standard deviations. In an independent-samples  $t$  test, this difference is statistically significant at  $p = .009$  using an independent-samples  $t$  test and  $p = .002$  using a paired-samples  $t$  test. This confirms that gender differences tend to be larger in modern Western countries than in more traditional non-Western countries (e.g., Costa, Terracciano & McCrae, 2001).

**Table 6.** Descriptive statistics of facet-level traits, complete sample;  $d$  = standardized sex difference (Cohen's  $d$ ).

|                    | Male        | Female       | $d$      |
|--------------------|-------------|--------------|----------|
| H: Sincerity       | 9.88 ± 2.19 | 10.12 ± 2.38 | -0.10**  |
| H: Fairness        | 9.42 ± 2.14 | 9.45 ± 2.12  | -0.01    |
| H: Greed avoidance | 5.77 ± 1.77 | 5.36 ± 1.90  | 0.22***  |
| H: Modesty         | 5.08 ± 1.83 | 5.11 ± 1.87  | -0.02    |
| E: Fear            | 9.69 ± 2.20 | 10.56 ± 2.26 | -0.38*** |
| E: Anxiety         | 6.84 ± 1.75 | 6.67 ± 1.87  | 0.09*    |
| E: Dependence      | 6.05 ± 1.59 | 6.17 ± 1.89  | -0.07    |
| E: Sentimentality  | 9.58 ± 1.84 | 10.19 ± 2.30 | -0.29*** |
| X: Self-esteem     | 9.83 ± 2.50 | 10.56 ± 2.30 | -0.30*** |

|                           | Male         | Female       | <i>d</i> |
|---------------------------|--------------|--------------|----------|
| X: Social boldness        | 9.46 ± 1.86  | 9.23 ± 2.26  | 0.11**   |
| X: Sociability            | 7.44 ± 1.56  | 6.98 ± 1.73  | 0.28***  |
| X: Liveliness             | 6.18 ± 1.54  | 6.38 ± 1.79  | -0.12**  |
| A: Forgiveness            | 6.94 ± 1.84  | 6.56 ± 1.77  | 0.21***  |
| A: Gentleness             | 9.72 ± 2.00  | 9.49 ± 2.12  | 0.11**   |
| A: Flexibility            | 9.25 ± 2.12  | 8.85 ± 2.14  | 0.19***  |
| A: Patience               | 6.16 ± 1.63  | 6.05 ± 1.82  | 0.06     |
| C: Organization           | 6.23 ± 1.49  | 6.27 ± 1.66  | -0.03    |
| C: Diligence              | 6.62 ± 1.54  | 6.87 ± 1.67  | -0.15*** |
| C: Perfectionism          | 10.21 ± 1.93 | 10.31 ± 2.06 | -0.05    |
| C: Prudence               | 7.80 ± 2.37  | 8.25 ± 2.41  | -0.19*** |
| O: Aesthetic appreciation | 6.24 ± 1.79  | 6.70 ± 1.84  | -0.25*** |
| O: Inquisitiveness        | 6.36 ± 1.65  | 5.85 ± 1.87  | 0.28***  |
| O: Creativity             | 9.82 ± 2.12  | 10.23 ± 2.38 | -0.18*** |
| O: Unconventionality      | 8.53 ± 1.99  | 9.05 ± 2.21  | -0.24*** |

**Table 7.** Descriptive statistics of facet-level traits, sample after removal of those with strong acquiescent responding.

|                    | Male         | Female       | <i>d</i> | <i>d</i> Croatia |
|--------------------|--------------|--------------|----------|------------------|
| H: Sincerity       | 9.82 ± 2.29  | 10.14 ± 2.46 | -0.13**  | -0.15            |
| H: Fairness        | 9.19 ± 2.08  | 9.41 ± 2.08  | -0.11*   | -0.56            |
| H: Greed avoidance | 5.84 ± 1.84  | 5.35 ± 1.96  | 0.25***  | -0.38            |
| H: Modesty         | 5.79 ± 1.66  | 5.47 ± 1.83  | 0.18***  | -0.12            |
| E: Fear            | 9.52 ± 2.29  | 10.61 ± 2.34 | -0.46*** | -0.81            |
| E: Anxiety         | 6.24 ± 1.56  | 6.36 ± 1.86  | -0.07    | -0.55            |
| E: Dependence      | 6.06 ± 1.66  | 6.19 ± 1.98  | -0.07    | -0.66            |
| E: Sentimentality  | 9.30 ± 1.87  | 10.12 ± 2.33 | -0.37*** | -0.99            |
| X: Self-esteem     | 10.63 ± 2.29 | 11.00 ± 2.15 | -0.18*** | 0.05             |
| X: Social boldness | 9.07 ± 1.81  | 9.03 ± 2.26  | 0.02     | -0.14            |
| X: Sociability     | 7.09 ± 1.57  | 6.75 ± 1.70  | 0.20***  | -0.27            |
| X: Liveliness      | 6.26 ± 1.58  | 6.48 ± 1.80  | -0.13**  | -0.21            |
| A: Forgiveness     | 6.45 ± 1.78  | 6.32 ± 1.72  | 0.07     | 0.05             |
| A: Gentleness      | 9.41 ± 2.06  | 9.24 ± 2.11  | 0.08     | 0.24             |
| A: Flexibility     | 9.63 ± 2.11  | 9.05 ± 2.15  | 0.27***  | 0.25             |
| A: Patience        | 6.22 ± 1.67  | 6.08 ± 1.84  | 0.08     | 0.46             |
| C: Organization    | 6.45 ± 1.53  | 6.38 ± 1.68  | 0.04     | -0.16            |

|                           | Male         | Female       | <i>d</i> | <i>d</i> Croatia |
|---------------------------|--------------|--------------|----------|------------------|
| C: Diligence              | 6.91 ± 1.53  | 7.06 ± 1.63  | -0.09*   | -0.22            |
| C: Perfectionism          | 10.02 ± 2.01 | 10.22 ± 2.12 | -0.10*   | -0.17            |
| C: Prudence               | 8.69 ± 2.03  | 8.87 ± 2.16  | -0.09    | 0.11             |
| O: Aesthetic appreciation | 5.93 ± 1.75  | 6.65 ± 1.88  | -0.39*** | -0.99            |
| O: Inquisitiveness        | 6.52 ± 1.72  | 5.85 ± 1.90  | 0.36***  | 0.08             |
| O: Creativity             | 9.61 ± 2.18  | 10.26 ± 2.40 | -0.28*** | -0.53            |
| O: Unconventionality      | 8.78 ± 2.04  | 9.25 ± 2.23  | -0.22*** | -0.30            |

7. *Items with strong sex differences*

Table 8 shows individual items with strong sex differences, distributed across all six scales. Some items are reverse-scored, but *d* values are given to indicate endorsement of the statement. For example, item A15r is reverse-scored, is endorsed more by females, making females score lower on Flexibility, one of the facets of *Agreeableness* (A).

**Table 8.** *HEXACO items with sex differences of at least 0.2 standard deviations. An r after the item number means it is reverse-scored. Positive d means males are more likely than females to endorse the statement.*

| Item                       | Statement   | <i>d</i> |
|----------------------------|---|----------|
| H18 Greed-avoidance        | Having a lot of money is not especially important to me.                          | .25      |
| H36 Fairness               | I would never accept a bribe, even if it were very large.                         | -.35     |
| H60r Fairness              | I'd be tempted to use counterfeit money, if I were sure I could get away with it. | .29      |
| E5 Fearfulness             | I would feel afraid if I had to travel in bad weather conditions.                 | -.28     |
| E23 Sentimentality         | I feel like crying when I see other people crying.                                | -.45     |
| E29 Fearfulness            | When it comes to physical danger, I am very fearful.                              | -.28     |
| E53r Fearfulness           | Even in an emergency I wouldn't feel like panicking.                              | .38      |
| X10r Social boldness       | I rarely express my opinions in group meetings.                                   | .23      |
| X40 Sociability            | The first thing that I always do in a new place is to make friends.               | .23      |
| X52r Social self-esteem    | I sometimes feel that I am a worthless person.                                    | .23      |
| A15r Flexibility           | People sometimes tell me that I'm too stubborn.                                   | -.45     |
| C14r Perfectionism         | When working on something, I don't pay much attention to small details.           | .35      |
| O1r Aesthetic appreciation | I would be quite bored by a visit to an art gallery.                              | .29      |
| O7 Inquisitiveness         | I'm interested in learning about the history and politics of other countries.     | .44      |
| O13 Creativity             | I would enjoy creating a work of art, such as a novel, a song, or a painting.     | -.23     |
| O25 Aesthetic appreciation | If I had the opportunity, I would like to attend a classical music concert.       | -.29     |
| O43 Unconventionality      | I like people who have unconventional views.                                      | -.36     |

## Discussion

The present study produced a number of novel results. One is that sex differences appeared not only in the scale means, but also in the response styles, or response biases. Generally, males showed more evidence of “careless responding”. They were more likely than females to choose consistently an extreme response, were more likely to choose mainly the midpoint of the scale, and had a greater tendency to agree rather than disagree with statements. The ubiquity of acquiescent responding and other response sets has been widely appreciated since Cronbach’s (1950) classical work, but sex differences in response sets and their effects on sex differences in scale scores have received hardly any attention since that time.

Sex differences in response style, and to acquiescent responding in particular, need to be taken into account. Specifically, when a scale contains reverse-scored items, acquiescent responding will reduce its internal coherence resulting in low Cronbach’s  $\alpha$ . This is because in the presence of acquiescent responding, the correlations of direct-scored items with reverse-scored items on the same scale will be less positive. In our sample, they actually became negative in males indicating that individual differences in acquiescence were more important determinants of responding than individual differences in the measured constructs. In the absence of reverse-scored items, acquiescent responding is expected to increase both the scales’ internal coherence and their correlations with other scales. Sex differences on the scales can be skewed when, as in our case, there is a substantial sex difference in acquiescent responding. Therefore, we recommend to either exclude individuals with strong acquiescence from analysis of the results, or to correct the scores for acquiescent responding as described in Rammstedt, Kemper and Borg (2013).

We know as yet little about the structure of personality in Saudi Arabia when assessed with Western inventories. Farrag (1987) had found previously that Western personality models worked in Saudi Arabia and his results suggested a 3-factor structure interpreted as *Neuroticism*, *Psychoticism* and *Extraversion*, but he did not explore sex differences. Our results show that the internal reliabilities of all six scales were lower than is usually seen in Western samples. For example, Ruch, Bruntsch and Wagner (2017) reported a range of alphas from .69 (*Agreeableness*) to .83 (*Extraversion*) in a predominantly Swiss student sample, which is much higher than those we found (Table 2). This suggests that the personality dimensions defined by the HEXACO may be a poor fit to personality structure in Saudi Arabia. This question, and the related issue of measurement invariance in different cultural contexts, needs to be examined in future studies.

Sex differences on the six scales were of rather small magnitude in our sample. For example, Ashton and Lee (2007) review results showing a nearly one standard deviation difference in *Emotionality*, with higher scores for females, both in Western countries and in South Korea. This was confirmed with a larger number of country samples by Lee and Ashton (2020). In our sample, this sex difference had a magnitude of only 0.29 standard deviations in the complete sample and 0.41 standard deviations after removal of those with strong acquiescence bias. This is the largest sex difference in our sample.

Otherwise, sex differences were most robust on *Agreeableness* and *Openness* as shown in Tables 3 and 4. *Agreeableness* scores are slightly higher for males than females in most of the samples surveyed in Lee and Ashton (2020). In our sample, this difference seems to be stronger than usual. The higher female score on *Openness* in our sample is not a typical finding, since sex differences on this scale fluctuate around zero in Lee and Ashton (2020). The difference is mainly on the Unconventionality facet, where females scored higher than males in our sample, which is the opposite of the usual pattern. Sex differences in *Honesty-humility* were smaller in our sample than in most other studies. This is because two of the four facets, Greed avoidance and Modesty, were higher in males than in females. The weak female advantage that we found on *Conscientiousness* is in line with Lee and Ashton (2020) and other studies, as is the near-zero sex difference on *Extraversion*.

There are two key theories with regard to understanding sex differences in average personality. Evolutionary psychologists such as Buss (1989) aver that males and females have been selected to have different average personalities. Females are higher in *Agreeableness* and *Conscientiousness* because the offspring of such females were more likely to survive. They are higher in *Neuroticism*, and thus anxiety, for the same reasons. Consistent with this, it has been found that sex differences in personality manifest very early and can, to some extent, be explained in terms of androgen exposure (see Baron-Cohen, 2002). Moreover, females suffer from higher levels of clinical depression than do males. This condition is significantly genetic and associated with *Neuroticism*, which is higher in females than in males (Albert, 2015).

As an alternative to evolutionary theory, social psychological theorists argue that children tend to internalize sex roles and this helps to explain the consistent sex differences (e.g., Eagly & Wood, 1991). The social psychological theory is unconvincing because it leaves us asking why these sex roles developed the way they did, and it fails to explain why they manifest at an early age, why they affect traits that have substantial heritability, and why they should be more pronounced in *more egalitarian societies*.

Costa et al. (2001) have argued that it may be useful to distinguish between 'masculine' societies, where innate sex differences are likely to become exaggerated, and 'feminine' societies, where they will be played down. However, the empirical evidence points in the opposite direction: it is in the supposedly more 'feminine' cultures with gender-egalitarian ideologies where the differences are most pronounced. Susan Pinker (2008) has argued that as individuals become more 'equal', their innate behavioral tendencies become expressed to a greater extent. Or, to use a different jargon, emancipation enhances inequality. However, this would imply that non-Western cultures masculinize females or feminize males in some way. It is possible that when females are less free, they are more likely to be inculcated into thinking in a more 'masculine' way — such as by possessing tough-minded social attitudes — and that this would reduce sex differences on personality tests. Even without cultural amplification, economic conditions such as the need for many women to perform strenuous physical labor can conceivably enhance "masculine" personality traits in women. Economically advanced societies have created vast numbers of office jobs which are more in line with female-typical preferences. This may have promoted gender differentiation of female office workers.

Another explanation is that there is a problem with the nature and cross-cultural applicability of the instrument. This issue has most frequently been raised for cognitive testing. For example, research by Dutton et al. (2018) administering Raven's tests to samples from South Sudan has revealed average IQs which are improbably low, in the region of 50 points. Dutton et al. argue that Ravens is highly problematic as a measure of the intelligence of South Sudanese children because these children are likely to be very unused to thinking in the highly abstract way which the test measures, as schooling in such countries, to the extent that it exists at all, tends to teach facts by rote. They are living in an environment somewhat resembling pre-modern England, a country where IQ appears to have continuously increased — on average by 2.5 points per decade — on Raven's-type tests across the twentieth century as the country has become more industrialized and people became more educated and adopted what James Flynn (2012) called "scientific spectacles". In addition, Dutton et al. (2018) argue that South Sudanese pupils are likely to be unused to taking tests, to geometric shapes of the kind found in Raven's tests, and even unused to paper, as they are likely to be taught to write on sand rather than paper. Thus, the instrument is problematic, as other researchers have also observed in Sub-Saharan Africa more broadly (e.g. Wicherts et al., 2010).

It may be that also personality tests are not fully applicable in non-western cultures. One possibility is that the instrument items are Western in their

orientation and this somehow causes confusion. The 60-item test includes a number of items which would be substantially irrelevant in non-western society, thereby reducing the internal consistency and the validity of the test.

Having looked at alternative explanations and found them wanting, it may be that our findings are in line with Lukaszewski et al.'s (2017) Socioecological Complexity Hypothesis (see also Gurven, 2018). According to this model, personality trait covariance is shaped within a particular socio-ecology. The more technologically advanced a society is, the narrower is the degree of each member's specialization. These authors hypothesize that the more specialized the society, the more niches there are to fill, and the more niches there are to fill the greater will be the number of specific personality types that are selected for. Thus, one niche might benefit from a very specific combination of relatively low *Agreeableness*, relatively low *Conscientiousness*, and relatively high *Neuroticism*, as has been proposed is the case of many artists (Post, 1994). In a less advanced society, there are fewer niches, meaning less intense selection for highly specific personality types, leading to more positive correlations between different aspects of personality.

Lukaszewski et al. (2017) have found, based on analysis of data from 55 nations, that aspects of personality are more strongly inter-correlated the less socioeconomically advanced the nation is, as measured by economic development, sectoral diversity and urbanization. Such a model, therefore, predicts that in a country like Saudi Arabia, compared to Western countries, item-level measures of a given Big 5 or HEXACO trait would be relatively strongly correlated. It would also predict a relatively strong relationship between the different traits. We clearly did not find the expected high correlations between items within each of the six scales, as is evident from the mediocre alpha reliabilities of the scales (see Table 2). Nor did we find high correlations between the six scales (Table 3).

Various analyses have found that the Big Five factors can be reduced down to a General Factor of Personality (GFP) which is extracted as the unrotated first principal component of a complex personality inventory. Its interpretation is uncertain, although it has been considered a measure of socially desirable responding or of social effectiveness (van der Linden et al., 2016). There is disagreement about whether the GFP can be extracted from the HEXACO as well, with some researchers claiming that a GFP cannot be found with the HEXACO (e.g., De Vries, 2011), and others finding one (e.g., Veselka et al., 2009). This positive manifold is predicted to be stronger in less developed countries (Rushton & Irwing, 2011). In our complete sample, an item-level factor analysis (after score reversal) produced a strong GFP that explained 11.6% of

ALMERAB, M.M. et al.      *SEX DIFFERENCES IN PERSONALITY IN SAUDI ARABIA*  
the total variance and that correlated at  $r = -.984$  with our measure of acquiescent responding. This means that in the complete sample, the GFP is a measure of acquiescent responding. After excluding respondents with strong acquiescence, we obtained a GFP that correlated only at  $r = .382$  with the GFP extracted from the complete sample and that did not represent acquiescent responding. It rather correlated with the HEXACO scales in a pattern that might represent socially desirable responding (Table 3).

### **Acknowledgement**

This research was funded by the Deanship of Scientific Research at Princess Nourah Bint Abdulrahman University through the Fast-track Research Funding Program.

### **References**

- Albert, P. (2015). Why is depression more prevalent in women? *Journal of Psychiatry and Neuroscience* 40: 219-221.
- Ashton, M.C. & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review* 11: 150-165.
- Ashton, M.C. & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment* 91: 340-345.
- Babarović, T. & Šverko, I. (2013). The HEXACO personality domains in the Croatian sample. *Društvena istraživanja: časopis za opća društvena pitanja* 22: 397-411.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences* 6: 248-254.
- Buss, D. (1989). *The Evolution of Desire: Strategies of Human Mating*. Basic Books.
- Church, A.T. (2016). Personality traits across cultures. *Current Opinion in Psychology* 8: 22-30.
- Cohen, E. & Deuling, J. (2014). Structural analysis of the abridged Big Five Circumplex: A comparison among gender and ethnic groups. *Bulletin of Sociological Methodology* 122: 68-86.
- Costa, P.T., Terracciano, A. & McCrae, R.R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology* 81: 322-331.

Cronbach, L.J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement* 10: 3-31.

de Vries, R.E. (2011). No evidence for a General Factor of Personality in the HEXACO Personality Inventory. *Journal of Research in Personality* 45: 229-232.

Dutton, E., Becker, D., Osman, H., Bakhiet, S., Essa, Y., Ali, H., Alqafari, S., Hamdi, A. & Alfelah, A. (2018). The Raven's test performance of South Sudanese samples: A validation of criticisms of the utility of Raven's among Sub-Saharan Africans. *Personality and Individual Differences* 128: 122-126.

Eagly, A.H. & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin* 17: 306-315.

Farrag, M. (1987). Dimensions of personality in Saudi Arabia. *Personality and Individual Differences* 8: 951-953.

Flynn, J.R. (2012). *Are We Getting Smarter?* Cambridge University Press.

Goldammer, P., Annen, H., Stöckli, P.L. & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *Leadership Quarterly*: 101384.

Gurven, M. (2018). Broadening horizons: Sample diversity and socioecological theory are essential to the future of psychological science. *Proceedings of the National Academy of Sciences USA* 115: 11420-11427.

Heine, S.J., Lehman, D.R., Peng, K. & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology* 82: 903-918.

Hodson, G., Book, A., Visser, B.A., Volk, A.A., Ashton, M.C. & Lee, K. (2018). Is the dark triad common factor distinct from low honesty-humility? *Journal of Research in Personality* 73: 123-129.

Kam, C.C.S. & Meyer, J.P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods* 18: 512-541.

Lee, K. & Ashton, M.C. (2020). Sex differences in HEXACO personality characteristics across countries and ethnicities. *Journal of Personality* 88: 1075-1090.

Lukaszewski, A., Gurven, M., Von Rueden, C. & Schmitt, D. (2017). What explains personality covariation? A test of the socioecological complexity hypothesis. *Social Psychological and Personality Science* 8: 943-952.

Meisenberg G. (2015). Do we have valid country-level measures of personality? *Mankind Quarterly* 55: 360-382.

Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences* 44: 1539-1550.

Pinker, S. (2008). *The Sexual Paradox*. Random House.

Post, F. (1994). Creativity and psychopathology. *British Journal of Psychiatry* 165: 22-34.

Rammstedt, B., Kemper, C.J. & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18-country cross-cultural study. *European Journal of Personality* 27: 71-81.

Ruch, W., Bruntsch, R. & Wagner, L. (2017). The role of character traits in economic games. *Personality and Individual Differences* 108: 186-190.

Rushton, J.P. & Irwing, P. (2011). The General Factor of Personality: Normal and abnormal. In: T. Chamorro-Premuzic, S. von Stumm & A. Furnham (eds.), *The Wiley-Blackwell Handbook of Individual Differences*. Blackwell.

Schmitt, D., Realo, A., Voracek, M. & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology* 94: 168-182.

van der Linden, D., Dunkel, C.S. & Petrides, K.V. (2016). The General Factor of Personality (GFP) as social effectiveness: Review of the literature. *Personality and Individual Differences* 101: 98-105.

Veselka, L., Schermer, J.A., Petrides, K.V., Cherkas, L.F., Spector, T.D. & Vernon, P.A. (2009). A general factor of personality: Evidence from the HEXACO model and a measure of trait emotional intelligence. *Twin Research and Human Genetics* 12: 420-424.

Weisberg, Y., DeYoung, C. & Hirsch, J. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology* 2: 178.

Wicherts, J., Dolan, C., Carlson, J., van der Maas, H. (2010). Raven's test performance of Sub-Saharan Africans: Average performance, psychometric properties and the Flynn Effect. *Learning and Individual Differences* 20: 135-151.

Yamagata, S., Suzuki, A., Ando, J., Ono, Y., Kijima, N., Yoshimura, K. et al. (2006). Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *Journal of Personality and Social Psychology* 90: 987-998.