



Big Five personality dimensions in Italian and Dutch adolescents: A cross-cultural comparison of mean-levels, sex differences, and associations with internalizing symptoms

Theo A. Klimstra^{a,b,*}, Elisabetta Crocetti^c, William W. Hale III^b, Alessandra Fermani^c, Wim H.J. Meeus^b

^a Catholic University Leuven, Belgium

^b Utrecht University, The Netherlands

^c University of Macerata, Italy

ARTICLE INFO

Article history:

Available online 12 March 2011

Keywords:

Cross-cultural
Measurement invariance
Personality
Adolescence
Five-factor model
Big Five
Anxiety
Depression

ABSTRACT

In the present cross-national comparison, self-reported Big Five personality data on large samples of Dutch ($N = 1521$) and Italian ($N = 1975$) adolescents were employed. Results suggest that the personality of Dutch and Italian adolescents can be described by the same Big Five traits, but that these might have slightly different meanings to the Dutch and Italian adolescent respondents. Supplementary analyses uncovered that sex differences are largest among Italian adolescents. Further comparisons reveal subtle cross-national differences in personality–psychopathology relationships, with stronger associations of Emotional Stability with depression for Italian when compared to Dutch adolescents. Results underscore that cross-national comparisons of personality may be alluring to use in research, however the findings of these comparisons should be interpreted with caution.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

In the previous two decades, there has been a growing consensus that the higher-order structure of personality can be described with five broad dimensions (Caspi, Roberts, & Shiner, 2005). These so-called “Big Five” dimensions are Extraversion (activity and dominance in social situations), Agreeableness (investment in maintaining positive and reciprocal relationships with others), Conscientiousness (planful, organized, and responsible behavioral tendencies), Emotional Stability (ability to deal with negative emotions in an effective manner), and Openness (curiosity and creativity) (Caspi et al., 2005; McCrae & John, 1992). The Big Five dimensions have been claimed to be universally replicable in adults (McCrae & Costa, 1997) as well as in adolescents (de Fruyt et al., 2009), although some researchers have expressed doubts with regard to this universality claim (e.g., de Raad et al., 2010; Peabody & de Raad, 2002).

Still, the claim regarding the universality of the Big Five has inspired several cross-cultural comparisons. Several studies,

including dozens of Western and developing countries (McCrae, 2001; Schmitt et al., 2007), have found substantial cross-cultural differences in self-reported mean-levels of personality traits in college students and adults. For example, the study by McCrae (2001) revealed that mean-level differences between countries were related to Hofstede’s (2001) well-known dimensions of culture. Power distance (i.e., acceptance of status differences) was negatively related to Extraversion, and positively related to Conscientiousness, uncertainty avoidance (i.e., engagement in activities to minimize threats in ambiguous situations) was negatively associated with Emotional Stability and Agreeableness, and there were positive associations of individualism (i.e., assertiveness and need for autonomy) with Extraversion and Openness. Only masculinity–femininity (i.e., the degree of sex-role differentiation which is highest in masculine cultures and lowest in feminine cultures) was unrelated to mean-level differences between countries. In a follow-up study that included additional cultures, Allik and McCrae (2004) reanalyzed McCrae’s (2001) data and related the findings (2001) to geographical regions. Allik and McCrae (2004) demonstrated that individuals from European and American cultures tend to report higher levels of Openness and Extraversion, and lower levels of Agreeableness when compared to individuals from Asian and African cultures. This line of research was extended by Schmitt et al. (2007) whom distinguished somewhat more specific geographical regions. Comparatively speaking, they found that East-Asians tend to report the lowest levels on all Big Five

* Corresponding author at: Department School Psychology and Child and Adolescent Development, Catholic University Leuven, Tiensestraat 102, bus 3717, 3000 Leuven, Belgium.

E-mail address: theoklimstra@online.nl (T.A. Klimstra).

¹ The first author is Postdoctoral Researcher at the Fund for Scientific Research Flanders (FWO).

dimensions, that Africans reported the highest levels of Conscientiousness, Agreeableness, and Emotional Stability, while South-Americans reported the highest scores on Openness. Within Europe, they found that Southern Europeans reported higher levels of Agreeableness and Conscientiousness, but lower levels of Emotional Stability when compared to Western Europeans.

However, it should be noted that some caution is warranted in conducting analyses on a geographical region level, as McCrae (2001) demonstrated that there can be substantive differences between neighboring countries. For example, in McCrae's study (2001) the Dutch were among the highest ranking countries in Emotional Stability, whereas their neighbors (Belgium and Germany) were ranked much lower on this trait. As a result, Schmitt et al.'s (2007) findings on Europeans only partly replicate previous work by McCrae (2001), as McCrae also found that Western Europeans (e.g., the Dutch) reported higher levels of Emotional Stability when compared to Southern Europeans (e.g., Italians and Spaniards). For Agreeableness, McCrae's findings contradict Schmitt et al.'s findings, as Western Europeans were more agreeable than Southern Europeans in McCrae's study, while the inverse pattern (with Southern Europeans being more agreeable than Western Europeans) was found in Schmitt et al.'s study. Thus, findings concerning the geographical distribution of personality trait-scores in Europe have not been all that consistent until now.

Sex differences in self-reported personality traits have also been compared across cultures. Generally, sex differences appeared to be larger in more economically prosperous (i.e., European and North-American) countries when compared to less economically prosperous (i.e., African and Asian) countries (Costa, Terracciano, & McCrae, 2001; Schmitt, Realo, Voracek, & Allik, 2008). In these two studies, the pattern of sex differences across cultures was compared with Hofstede's (2001) cultural dimensions. Only one of these cultural dimensions (i.e., individualism–collectivism) was associated with the magnitude of sex differences, indicating that sex differences were larger in more individualistic countries. Intuitively, one might expect sex differences to be larger in more masculine cultures. However, Hofstede's (2001) masculinity–femininity dimension was not associated with sex differences in the studies of Costa et al. (2001) and Schmitt et al. (2008), suggesting that countries with more differentiated sex roles are not necessarily characterized by larger sex differences in personality traits.

All the aforementioned cross-cultural comparisons focused only on adult and college student samples, but recently it has been suggested that these findings can also be replicated among adolescent samples (McCrae et al., 2010). However, instead of simply using adolescent self-reported data, as has been done in adult and college student samples, the study by McCrae et al. (2010) instead relied on third-person ratings of the adolescents' personalities. More specifically, these raters were assigned a target (i.e., an adolescent), whose personality they had to rate. However, a similarly recent work by Wood, Harms, and Vazire (2010) has demonstrated that reports about someone else's personality might be as informative about the personality of the perceiver (the third-person providing the ratings) as they are about the personality of the target (the person being rated). In other words, the value of McCrae et al. (2010) replication of adult and college student findings with an adolescent sample may be somewhat limited. Therefore, it would appear that adolescent self-reported personality data is needed in addition to third-person judgment based data, in order to obtain more insight into cross-cultural differences in adolescent personality traits. However, as of the present, we are unaware of any previous cross-cultural studies that have employed adolescent personality self-reports.

A lack of studies using adolescent self-reports is not the only thing that is missing in regard to cross-cultural research. Extensive literature regarding the linkages between Big Five personality traits and psychopathology in adults (Kotov, Gamez, Schmidt, &

Watson, 2010) and accumulating knowledge regarding these associations in adolescents (Tackett, 2006) have revealed substantive negative associations of psychopathology with Emotional Stability and Extraversion. Additionally, negative associations have been found with Agreeableness and Conscientiousness, albeit weaker, but still significant. Again, despite the substantive number of studies on personality–psychopathology linkages, we are unaware of cross-cultural comparisons of such associations. Because it is well known that people from different countries express psychopathological symptoms in different ways (Lewis-Fernandez & Kleinman, 1994), and personality traits also appear to have a slightly different meanings to persons across cultures (Nye, Roberts, Saucier, & Zhou, 2008), it seems obvious that personality–psychopathology linkages found in one culture may not necessarily apply to other cultures. While cross-cultural comparisons of personality–psychopathology associations could considerably add to our understanding of the meaning that specific personality traits and/or specific psychopathological symptoms may have across the globe, certain conditions to such cross-cultural comparisons must be applied. Specifically, before cross-cultural comparisons of any kind can be conducted, one should first perform several tests to rule out that statistical artifacts are the most likely cause of one's results.

1.1. Methodological issues with cross-cultural comparisons

In interpreting cross-cultural mean-level difference findings it is important to be cautious for several reasons (Perugini & Richetin, 2007). One potential source of error that can influence the validity of cross-cultural comparisons is the “frame-of-reference effect” (Heine, Lehman, Peng, & Greenholtz, 2002). The frame-of-reference effect is that a person will compare him or herself to reference persons which usually happen to be the one's in their direct proximity (i.e., people from their own culture). Because of this effect, an Italian might compare him or herself to other Italians when providing personality ratings, whereas a Dutch person might compare him or herself to other Dutch persons. It is perhaps needless to say that the frame-of-reference effect would obscure the validity of mean-level differences between two cultures (in this case, between the Italians and the Dutch).

However, the most important reason to be cautious with cross-cultural comparisons is that one needs to make sure that the instrument that is used measures the construct in similar ways across cultures. In other words, measurement invariance for the instrument needs to be established before cross-cultural differences can be interpreted (e.g., Cheung & Rensvold, 2002). It is for this reason that Exploratory and Confirmatory Factor Analytic approaches are widely used (Caprara, Barbaranelli, Bermudez, Maslach, & Ruch, 2000). The aforementioned cross-cultural studies (McCrae, 2001; McCrae et al., 2010; Schmitt et al., 2007) have mainly relied on Exploratory Factor Analysis (EFA) techniques to establish measurement invariance. However, it has been argued that EFA is mainly useful to test whether there is bias at the configural level (i.e., examining whether the number of factors and the pattern of factor loadings is roughly equivalent in different groups or cultures; Vandenberg & Lance, 2000), but is somewhat limited in its potential to detect additional forms of bias. In order to overcome these limitations, Item Response Theory (IRT) methods and Confirmatory Factor Analyses (CFAs) have been recommended to test for two additional types of measurement invariance (i.e., metric and scalar invariance). In order to first establish metric invariance, one needs to check whether constraining factor loadings to be equal across groups affects model fit. Secondly, scalar invariance is tested by examining whether model fit is affected by constraining intercepts of latent factor indicators (e.g., items) to be equal across groups/cultures. Together metric and scalar invariance tests are useful in order to check for systematic response bias

(Vandenberg & Lance, 2000). That is, individuals from different cultures might rate themselves differently on an item solely because they interpret it differently, and not because of true mean-level differences. For example, the item “neat” may be more related to being orderly in the Netherlands, and perhaps more to having your hair properly cut in Italy (although we are aware that this particular example is somewhat stereotypical). To put it more technically, one needs to demonstrate that individuals from different cultures with a similar mean-level on a latent factor, also have a similar pattern of scores across the indicators of this latent factor (i.e., the items).

Previous work with IRT-methods revealed that it is difficult to demonstrate full metric and scalar invariance in commonly used personality inventories, such as the Multidimensional Personality Questionnaire (Johnson, Spinath, Krueger, Angleitner, & Riemann, 2008), the Trier Personality Inventory (Ellis, Becker, & Kimmel, 1993), and the NEO Personality Inventory (Huang, Church, & Katigbak, 1997). However, IRT-methods are only sporadically used because they are only recommended for very large samples of more than 1000 cases (Stark, Chernyshenko, & Drasgow, 2006). A good alternative for IRT-methods are Confirmatory Factor Analyses (CFAs), as these analyses can be used to detect the same forms of bias, but are less demanding in terms of the required sample size. Unfortunately, CFAs have also rarely been used in cross-cultural research on personality. One exception is a study by Nye et al. (2008). Similar to the aforementioned IRT-studies, they were also unable to establish full measurement equivalence. Thus, establishing full measurement equivalence seems hard to achieve, even though it is a necessity for assuring that cross-cultural differences reflect true mean-level differences, and cannot be solely attributed to different interpretations of specific items.

1.2. The present study

To the best of our knowledge, the current study will be the first to provide a cross-cultural comparison of adolescent self-reported personality traits. To pursue this goal, large samples of Dutch ($N = 1521$) and Italian ($N = 1975$) adolescents will be employed. However, before this goal can be pursued, a prerequisite is the establishment of measurement invariance, as aforementioned. For that purpose, we will examine configural, scalar, and metric invariance of a 30-item version of Goldberg's Big Five questionnaire (Gerris et al., 1998; Goldberg, 1992) that the Dutch and Italian adolescent samples completed. If measurement invariance can be established, we will be able to pursue the following three research goals: (1) comparing mean-levels of adolescent self-reported Big Five personality traits of Dutch and Italian adolescents, (2) compare sex differences in these traits across the two cultures, and (3) examine cross-cultural differences in personality–psychopathology linkages. In all analyses, we will distinguish among early and middle adolescents, because previous studies found considerable age-related changes in Big Five traits among adolescents (Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2009).

Hence, the first goal of this study will be to assess mean-level cross-cultural differences in adolescent self-reported Big Five personality traits. It is reasonable to expect personality differences between Dutch and Italian adolescents, as they have been shown to differ on various variables that are related to personality. For example, a report of the World Health Organization (WHO) found that Italian adolescents report having more trouble communicating with their parents when compared to Dutch adolescents (Currie et al., 2008). Furthermore, Italian adolescents report having fewer friends when compared to Dutch adolescents. In addition to these cross-cultural differences found in the WHO study, Crocetti and colleagues found that Italian adolescents report having higher levels of anxiety (Crocetti, Hale, Fermani, Raaijmakers, & Meeus, 2009) and a

less mature identity (Crocetti, Schwartz, Fermani, & Meeus, 2010) when compared to Dutch adolescents. Adolescent personality traits have been shown to be related to all the aforementioned. That is, specific adolescents personality traits have been shown to be associated with parent–child relationship quality (Agreeableness; Branje, van Lieshout, & van Aken, 2004), the number of friends one selects and the number of times one is selected as a friend (Extraversion and Agreeableness, respectively; Selfhout et al., 2010), anxiety (Emotional Stability; Krueger, 1999), and identity (Agreeableness and Conscientiousness; Crocetti, Rubini, & Meeus, 2008; Luyckx, Soenens, & Goossens, 2006). As personality traits have been shown to be related to so many variables on which cross-cultural differences have been demonstrated, such differences are also likely to emerge in adolescent personality traits. If cross-cultural differences in mean-levels of personality traits are found, the aforementioned studies suggest that Dutch adolescents are likely to reflect more favorable (i.e., higher) personality trait scores than Italian adolescents, especially for Extraversion, Agreeableness, Conscientiousness, and Emotional Stability. For Openness, we are unsure of what to expect, as this trait has only weak linkages with the variables on which Dutch and Italian adolescents tend to differ.

The second goal of the study is to compare the magnitude of sex difference in the Dutch and the Italian sample. Using different questionnaires, sex differences in Big Five traits have been found both in Dutch (e.g., Klimstra et al., 2009) and Italian adolescents (e.g., Caprara, Barbaranelli, Pastorelli, & Cervone, 2004). Although it is less than ideal to compare results obtained with different questionnaires, a glance at these studies seems to suggest larger sex differences among Italian adolescents when compared to Dutch adolescents. In the current study, we aim to provide a more appropriate cross-cultural test by comparing sex differences in adolescent self-reported personality traits as examined by a Dutch and an Italian version of the same Big Five questionnaire.

The third and final goal of the current study is to compare personality–psychopathology linkages in Italian and Dutch adolescents. We will focus on two types of commonly experienced psychopathological symptoms of adolescents: depressive symptoms and the Generalized Anxiety Disorder (GAD) symptom of worry. These two forms of psychopathology have previously been found to be strongly intertwined with personality traits (i.e., Emotional Stability, Extraversion, and Conscientiousness) of Dutch adolescents (e.g., Hale, Klimstra, & Meeus, 2010; Klimstra, Akse, Hale, Raaijmakers, & Meeus, 2010), but we are unaware of studies examining associations of Big Five personality traits with the GAD symptom of worry and depressive symptoms in Italian samples. A study among late childhood participants did however uncover linkages between the Big Five traits (i.e., Emotional Stability, Extraversion, Conscientiousness, and Openness) and a broad internalizing factor comprised of depressive, anxiety, somatic, and obsessive symptoms (Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003). Therefore, it seems reasonable to expect that the GAD symptom of worry and depressive symptoms will be associated with roughly the same Big Five personality traits (i.e., Emotional Stability, Extraversion, and Conscientiousness) in Dutch and Italian adolescents. However, the present study will provide a more detailed perspective on these linkages, as depressive symptoms and the GAD symptom of worry will be considered separately.

2. Method

2.1. Participants

The Italian sample consisted of 1975 adolescents (902 boys and 1073 girls) attending various junior high and high schools in the

east-central region of Italy. Participants ranged in age from 11 to 19 years ($M_{age} = 14.5$, $SD = 2.4$). Two age groups were represented in the sample: an early adolescent group (aged 11–14 years) of 1050 adolescents ($M_{age} = 12.5$ years, $SD = 1.0$) and a middle adolescent group (aged 15–19 years) of 925 adolescents ($M_{age} = 16.8$ years, $SD = 1.2$).

The Dutch sample consisted of 1521 adolescents (706 boys and 815 girls) attending various junior high and high schools in the province of Utrecht in the Netherlands. Participants ranged in age from 11 to 19 years ($M = 14.2$; $SD = 2.2$). Two age groups were represented in the sample: an early adolescent group (aged 11–14 years) of 880 adolescents ($M_{age} = 12.3$ years, $SD = 0.6$) and a middle adolescent group (aged 15–19 years) of 641 adolescents ($M_{age} = 16.7$ years, $SD = 0.8$).

To ensure that cross-national differences were not confounded with ethnicity, we only focused on indigenous Italian and Dutch adolescents. Additionally, adolescents in both countries attended school full-time, and they were comparable in terms of years and type of education. In fact, while in both countries there is no differentiation within the junior high school system, high schools are differentiated in various tracks (from the highest level represented by schools who prepare pupils for university attendance to the lowest level represented by vocational schools). All these tracks were represented in the Italian and Dutch samples.

2.2. Procedure

Prior to initiating the study, we obtained permission from the school principals to administer questionnaires during class time. Parents were provided with written information about the research and were asked for their consent for the adolescent to participate. After we received parental permission, students were informed about the study and asked whether they wished to participate. Approximately 99% of the approached students chose to participate. Interviewers then visited the schools and asked adolescents to fill out the questionnaire packet. This procedure was followed for both the Italian and Dutch adolescent samples.

2.3. Measures

2.3.1. Personality

A shortened version of Goldberg's Big Five questionnaire (Gerris et al., 1998; Goldberg, 1992) was used. Participants were asked to rate 30 items (6 items for each factor) on a seven-point scale, ranging from 1 (does not apply to me at all) to 7 (applies to me very well). Sample items include: talkative (Extraversion), sympathetic (Agreeableness), systematic (Conscientiousness), nervous (Emotional Stability), and versatile (Openness to Experience). Cronbach's alphas were .70 and .82 for Extraversion, .71 and .85 for Agreeableness, .72 and .83 for Conscientiousness, .72 and .81 for Emotional Stability, and .65 and .75 for Openness to Experience, in the Italian and Dutch samples, respectively.

2.3.2. Depression

The Children's Depression Inventory (CDI; Kovacs, 1985, 1988; Timbremont & Braet, 2002) was used to measure sub-clinical depressive symptoms. The CDI consists of 27 items, each responded to on a three-point scale: 1 (false), 2 (a bit true), and 3 (very true). A sample item is "I am sad all the time". Cronbach's alphas were .88 and .92 in the Italian and Dutch samples, respectively.

2.3.3. Generalized anxiety symptoms

The Generalized Anxiety Symptoms (GAD) subscale from the Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1997; Crocetti et al., 2009; Hale, Raaijmakers, Muris, & Meeus, 2005) was used to assess the generalized anxiety

disorder symptom of worry. The GAD scale consists of 7 items scored on a three-point scale: 1 (almost never), 2 (sometimes), and 3 (often). A sample item is: "I worry about whether others will like me". Measurement invariance for this measure has been established for early and middle adolescent Dutch and Italian boys and girls in a previous study (Crocetti et al., 2009). Cronbach's alphas were .76 and .86 in the Italian and Dutch samples, respectively.

3. Results

3.1. Preliminary analyses: testing measurement equivalence for personality

We tested for three types of measurement equivalence (configural, metric, and scalar invariance) of personality by employing Confirmatory Factor Analysis (CFA) in Mplus 4 (Muthén & Muthén, 2007) using Maximum Likelihood (ML) estimation. Tests for configural invariance are used to establish whether a model that yields an adequate fit in one sample also yields an adequate fit in another sample. No formal model comparisons can be run with regard to configural invariance, as this would imply comparing two non-nested models.

To test for metric and scalar equivalence, we compared the fit of multigroup CFA models (with Dutch early adolescent boys, Dutch middle adolescent boys, Dutch early adolescent girls, Dutch middle adolescent girls, Italian early adolescent boys, Italian middle adolescent boys, Italian early adolescent girls, and Italian middle adolescent girls as groups) without constraints (i.e., models with metric or scalar variance) to constrained models (i.e., models with metric or scalar invariance). For such model comparisons, the use of multiple criteria has been advocated by Vandenberg and Lance (2000), as different criteria can provide information on different sources of potential model misspecification. Because the χ^2 -statistic is well known to be overly sensitive to sample size and model complexity (e.g., Cheung & Rensvold, 2002), we relied on two other commonly used fit indices: the delta (Δ) Comparative Fit Index (CFI) and the delta (Δ) Root Means Square Error of Approximation (RMSEA). We concluded that there was measurement equivalence if Δ CFI was smaller than .010 and Δ RMSEA was smaller than .015 (Chen, 2007). Absolute fit indices of the various models were also considered, with CFIs of .90 and larger and RMSEAs of .08 and smaller considered satisfactory (Kline, 2005).

For CFAs, using items as indicators of latent factors can lead to overly complex models with a large number of parameters to be estimated. In addition, it has been argued that the optimal number of indicators for latent factors is three as it leads to a just-identified model, whereas fewer indicators lead to an under-identified model and more than three indicators yield an over-identified model (Little, Cunningham, Shahar, & Widaman, 2002). To reduce the number of indicators of a latent factor to the optimal number of three and thereby reduce model complexity, it has been recommended to use parcels consisting of multiple items instead of using individual items (e.g., Marsh & Hau, 1999). We used the well-established item-to-construct balance parceling method (Little et al., 2002) to create three two-item parcels for each Big Five trait resulting in a total of 15 parcels. These parcels were used as input for CFAs in which we tested for configural, metric, and scalar invariance. In all models, the five latent factors representing the Big Five personality traits were allowed to correlate with each other.

To test for configural invariance, we first examined a simple-structure five-factor model in which the latent factors were allowed to correlate with one another for the Dutch sample. This model is depicted in Fig. 1. This model yielded a reasonable fit, although the RMSEA was slightly higher than the .08 benchmark advocated by Kline (2005) (see Table 1).

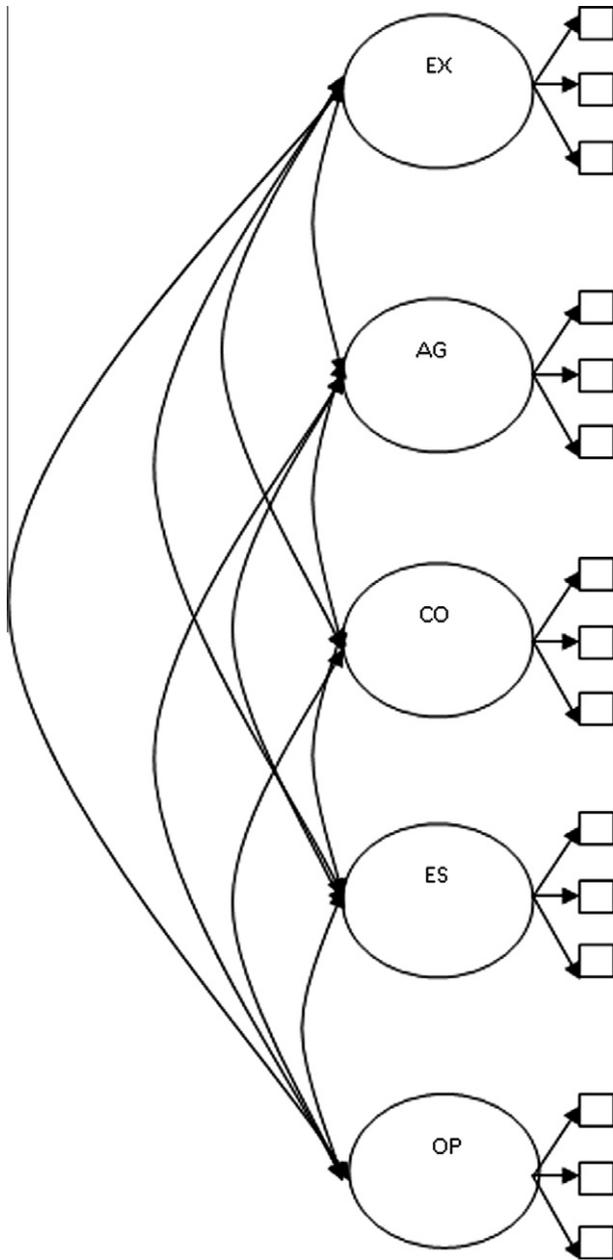


Fig. 1. Estimated Confirmatory Factor Analysis model for establishing configural invariance.

Table 1
Model fit statistics of a five-factor model of personality in Dutch and Italian early and middle adolescent boys and girls.

	χ^2	df	CFI	RMSEA	RMSEA (90% C.I.)
Dutch	1083.55***	77	.910	.093	.088–.098
Males	644.414***	77	.897	.102	.095–.110
Females	567.665***	77	.916	.088	.082–.095
Early Adolescents	688.757***	77	.904	.095	.089–.102
Middle Adolescents	481.546***	77	.920	.091	.083–.098
Italian	883.44***	77	.904	.073	.069–.078
Males	383.915***	77	.913	.066	.060–.073
Females	600.701***	77	.889	.080	.074–.086
Early Adolescents	457.560***	77	.908	.069	.063–.075
Middle Adolescents	521.277***	77	.894	.079	.073–.085

Note. df = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.

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

As Table 1 shows, the same model that we applied to the Dutch sample also yielded a reasonable fit in the Italian sample. Therefore, we established configural invariance for nationality. We proceeded with tests for configural invariance in all gender and age groups. These tests revealed that the model provided a good fit for Dutch girls, Dutch early adolescents, Dutch middle adolescents, Italian boys, and Italian early adolescents. The models for Dutch males, Italian females, and Italian middle adolescents yielded a slightly less optimal fit (see Table 1). However, the fit statistics still seem to indicate that there is reasonable evidence for configural invariance. Therefore, we could proceed with tests for metric and scalar invariance.

We first tested for metric invariance by running two multigroup simple-structure CFAs with eight groups (i.e., Dutch early adolescent boys, Dutch middle adolescent boys, Dutch early adolescent girls, Dutch middle adolescent girls, Italian early adolescent boys, Italian middle adolescent boys, Italian early adolescent girls, and Italian middle adolescent girls) as groups. Because the fit of a model in which factor loadings were freely estimated for each of the groups [metric invariance model: χ^2 (617) = 2570.583 ($p < .001$), CFI = .901, RMSEA = .085 (90% C.I. = .082–.089)] fitted the data as well as a model in which these factor loadings were constrained to be equal for all groups [metric invariance model: χ^2 (687) = 2828.920 ($p < .001$), CFI = .892, RMSEA = .084 (90% C.I. = .081–.088); Δ CFI = .009, Δ RMSEA = -.001], we concluded that our Big Five measure was metrically invariant for the eight groups involved. Therefore, we could proceed to the scalar invariance tests.

Our baseline model for scalar invariance tests was the eight-group model in which factor loadings were constrained to be equal for all groups, but in which intercepts were freely estimated. This baseline model (i.e., the scalar variance model, which is a simple-structure model) was compared to a model in which intercepts were constrained to be equal for all groups (i.e., the scalar invariance model, which is a mean-structure model). The scalar invariance model [χ^2 (757) = 4610.708 ($p < .001$), CFI = .805, RMSEA = .108 (90% C.I. = .105–.111)] had a fit that was much worse than the scalar variance model (Δ CFI = .087, Δ RMSEA = .024). Consequently, we concluded that there was no overall scalar invariance. However, we did proceed to test whether we could establish scalar invariance for specific pairs of gender and age groups within and across cultures. The results of these comparisons appear in Table 2.

Table 2 indicates that scalar invariance was established within the Dutch sample in almost all instances. Comparisons between Dutch middle adolescent boys and girl should, however, be interpreted with some caution as Δ CFI (.011) was just above the .010 benchmark. Within the Italian sample, the situation was the complete opposite, as the only comparison that signified scalar invariance was the comparison between Italian middle adolescent boys and girls. In all other instances, group comparisons within the Italian culture should be interpreted cautiously.

There was no evidence for cross-cultural scalar invariance. Thus, cross-cultural mean-level differences of Big Five traits may be interpreted (because there was configural and metric invariance), but only cautiously (because of a lack of scalar invariance). As there was metric invariance for sex groups but no scalar invariance within the Italian sample, comparisons of sex differences across cultures and mean comparisons within the Italian sample should also be interpreted with some caution.

Cross-cultural comparisons of associations of personality traits with other variables (i.e., problem behavior symptoms) ideally require metric invariance, which we managed to establish across all groups. Therefore, these comparisons can be readily interpreted. In fact, between-group differences in associations of personality with problem behavior symptoms could even provide some insights into the (slight) differences in meaning a specific personality trait

Table 2
Scalar invariance tests for nationality, gender, and age.

	Scalar variance				Scalar invariance			
	χ^2	df	CFI	RMSEA (90% C.I.)	χ^2	df	CFI	RMSEA (90% C.I.)
Overall gender by nationality	2354.093***	338	.896	.083 (.080, .086)	3894.098***	368	.819	.105 (.102–.108)
Dutch boys vs Dutch girls	1254.765***	164	.904	.094 (.089, .098)	1337.788***	174	.897	.094 (.089–.099)
Italian boys vs Italian girls	1015.336***	164	.897	.073 (.068, .077)	1116.132***	174	.885	.074 (.070–.078)
Dutch boys vs Italian boys	1106.512***	164	.895	.085 (.080, .089)	1706.516***	174	.830	.105 (.100–.109)
Dutch girls vs Italian girls	1207.157***	164	.901	.082 (.078, .086)	1946.843***	174	.831	.104 (.100–.108)
Overall age by nationality	2331.162***	338	.899	.082 (.079, .085)	3886.417***	368	.822	.105 (.102–.108)
Dutch early vs Dutch middle	1261.781***	164	.904	.094 (.089, .099)	1365.145***	174	.896	.095 (.090–.100)
Italian early vs Italian middle	989.550***	164	.901	.071 (.067, .076)	1134.960***	174	.884	.075 (.071–.079)
Dutch early vs Italian early	1257.135***	164	.896	.083 (.079, .087)	2173.703***	174	.809	.109 (.105–.113)
Dutch middle vs Italian middle	1015.531***	164	.908	.081 (.077, .086)	1512.983***	174	.855	.099 (.095–.104)
Dutch early boys vs Dutch early girls	879.772***	164	.890	.100 (.093, .016)	913.111***	174	.886	.098 (.092–.105)
Dutch mid boys vs Dutch mid girls	569.000***	164	.920	.088 (.080, .093)	634.868***	174	.909	.091 (.083–.099)
Dutch early boys vs Dutch mid boys	789.853***	164	.888	.105 (.098, .112)	844.480***	174	.881	.104 (.097–.112)
Dutch early girls vs Dutch mid girls	692.858***	164	.910	.089 (.082, .096)	764.585***	174	.900	.091 (.085–.098)
Italian early boys vs Italian early girls	576.982***	164	.899	.069 (.063, .075)	660.327***	174	.881	.073 (.067–.079)
Italian mid boys vs Italian mid girls	640.844***	164	.884	.079 (.073, .086)	671.581***	174	.879	.079 (.072–.085)
Italian early boys vs Italian mid boys	484.591***	164	.910	.066 (.059, .073)	532.629***	174	.899	.068 (.061–.074)
Italian early girls vs Italian mid girls	711.619***	164	.883	.079 (.073, .085)	831.314***	174	.859	.084 (.078–.090)

Note. Comparisons in which scalar invariance was established are printed in bold. *df* = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

might have in these different groups. In other words, between-group differences in personality–problem behavior symptom linkages could potentially clarify why scalar invariance could not be established.

3.2. Sex differences within cultures

Sex differences within cultures were inspected with a Multivariate Analysis Of Variance (MANOVA). In these analyses we accounted for age differences, as we examined sex differences within age cohorts (i.e., Dutch early adolescent boys versus Dutch early adolescent girl, Dutch middle adolescent boys versus Dutch middle adolescent girls, Italian early adolescent boys versus Italian early adolescent girls, and Italian middle adolescent boys versus Italian middle adolescent girls). The results of these analyses appear in Table 3.

Within the Dutch sample, no sex differences were found with regard to Extraversion (see Table 3). However, early adolescent boys tended to be less agreeable than early adolescent girls. This sex difference was replicated in middle adolescence, where Dutch boys were again less agreeable when compared to Dutch girls. Dutch middle adolescent girls also were significantly more

conscientious and open to experience, but less emotionally stable when compared to Dutch middle adolescent boys.

Sex differences in the Italian sample should be interpreted cautiously, as we only managed to establish scalar invariance for Italian middle adolescent boys and girls. That being said, it appears from Table 3 that there were no sex differences in Extraversion among Italian adolescents. Italian girls were more agreeable and conscientious than Italian boys, both in early and middle adolescence. For Emotional Stability, Italian middle adolescent girls reflected lower levels than Italian middle adolescent boys. For Openness, there were no gender or age differences among Italians.

3.3. Mean-level differences across cultures

We assessed mean-level differences across cultures by comparing similar sex and age groups across cultures (e.g., Dutch early adolescent boys versus Italian early adolescent boys, Dutch early adolescent girls versus Italian early adolescent girls, Dutch middle adolescent boys versus Italian middle adolescent boys, and Dutch middle adolescent girls versus Italian middle adolescent girls) with a MANOVA. It should be noted that these comparisons should be interpreted very cautiously, because we did not establish

Table 3
Mean-levels of personality, depression and worry in early and middle adolescent Dutch and Italian boys and girls.

	Early adolescents				Middle adolescents			
	Dutch		Italian		Dutch		Italian	
	Boys (<i>N</i> = 434)	Girls (<i>N</i> = 446)	Boys (<i>N</i> = 520)	Girls (<i>N</i> = 530)	Boys (<i>N</i> = 272)	Girls (<i>N</i> = 369)	Boys (<i>N</i> = 382)	Girls (<i>N</i> = 543)
<i>Personality</i>								
Ex	4.86 (.87) ^{cd}	4.97 (.99) ^d	4.77 (1.20) ^{cd}	4.65 (1.30) ^{bc}	4.65 (1.07) ^{bc}	4.74 (1.14) ^{cd}	4.49 (1.16) ^{ab}	4.37 (1.40) ^a
Ag	5.03 (.98) ^{ab}	5.23 (.80) ^c	4.97 (.90) ^a	5.41 (.84) ^d	5.16 (.86) ^{bc}	5.44 (.69) ^d	5.08 (.92) ^{abc}	5.44 (.78) ^d
Co	4.10 (1.04) ^{bc}	4.22 (1.01) ^{cd}	3.94 (1.13) ^{ab}	4.27 (1.06) ^{cd}	4.08 (1.06) ^{abc}	4.37 (1.13) ^d	3.85 (1.15) ^a	4.14 (1.15) ^{bcd}
ES	4.67 (1.02) ^e	4.56 (.97) ^{de}	4.43 (1.05) ^{cd}	4.22 (1.14) ^{bc}	4.51 (.98) ^{de}	4.22 (.96) ^{bc}	4.04 (1.12) ^b	3.40 (1.06) ^a
Op	4.46 (.99) ^a	4.45 (.90) ^a	4.46 (1.02) ^a	4.63 (1.03) ^{ab}	4.60 (.95) ^{ab}	4.74 (.84) ^b	4.47 (1.03) ^a	4.52 (.94) ^a
<i>Int. problems</i>								
Dep	1.15 (.28) ^a	1.17 (.20) ^{ab}	1.36 (.28) ^c	1.35 (.27) ^c	1.19 (.26) ^{ab}	1.21 (.23) ^b	1.39 (.30) ^{cd}	1.43 (.31) ^d
GAD	1.33 (.35) ^a	1.35 (.30) ^a	1.70 (.43) ^c	1.84 (.45) ^d	1.31 (.33) ^a	1.48 (.43) ^b	1.77 (.43) ^{cd}	2.01 (.45) ^e

Note. Ex = Extraversion, Ag = Agreeableness, Co = Conscientiousness, ES = Emotional Stability, Op = Openness, Int. problems = Internalizing Problems, Dep = Depression; GAD = the Generalized Anxiety Disorder Symptom of Worry. Means of groups that share the same superscript are equal, means of groups with different superscripts are significantly different ($p < .05$).

cross-cultural scalar measurement equivalence (see Table 2). Descriptive statistics for these mean-level comparisons are displayed in Table 3.

Table 3 suggests that Dutch girls (both early and middle adolescents) were more extraverted than their Italian counterparts, whereas there were no cross-cultural mean differences for boys. For Agreeableness, Dutch early adolescent girls exhibited higher mean-levels than Italian early adolescent girls. Italian and Dutch adolescents appear to be equally conscientious. With regard to Emotional Stability, this was by no means the case as all Dutch adolescent groups displayed higher levels on this trait than their Italian counterparts. Finally, for Openness there was one cross-cultural difference as Dutch middle adolescent girls were more open to experience than Italian middle adolescent girls.

3.4. Comparing age and sex differences in personality within cultures across cultures

As a final mean-level analysis, we compared the magnitude of sex differences across cultures with a MANOVA. Again, these results should be interpreted cautiously because of the lack of scalar invariance across cultures. Descriptive statistics concerning these analyses appear in Table 3.

First, a significant multivariate interaction effect of nationality by sex ($F(5, 3484) = 4.30$; $p = .001$; partial $\eta^2 = .006$) indicated that the overall magnitude of gender differences differs between cultures. Univariate tests indicated that this interaction only applied to Extraversion ($F(1, 3488) = 7.23$; $p = .007$; partial $\eta^2 = .002$), Agreeableness ($F(1, 3488) = 7.46$; $p = .006$; partial $\eta^2 = .002$), and Emotional Stability ($F(1, 3488) = 9.64$; $p = .002$; partial $\eta^2 = .003$). An inspection of Table 3 reveals that gender differences in these traits were larger among Italian adolescents when compared to Dutch adolescents.

To examine whether the sex by nationality effects were moderated by age group, we examined a three-way interaction effect of nationality by sex by age group. The multivariate effect was not significant ($F(5, 3484) = 1.96$; $p = .081$; partial $\eta^2 = .003$), indicating that sex by nationality interactions were comparable for early and middle adolescents.

3.5. Comparing associations of Big Five personality traits with internalizing problems across cultures, sex, and age groups

We ran a multigroup structural equation model in Mplus 4 (Muthén & Muthén, 2007) to examine the associations between the Big Five personality traits and two distinct internalizing problems (i.e., depressive symptoms and the Generalized Anxiety Disorder symptom of worry) in the aforementioned eight nationality, gender, and age groups. Maximum Likelihood Robust (MLR) estimation was used as MLR has been shown to be the most accurate estimator when the distribution of scores slightly deviates from a normal distribution (Satorra & Bentler, 1994), which turned out to be the case for the scores on our measures for depression and worry. In the structural equation model, the Big Five personality traits were the independent variables, whereas depression and worry were the dependent variables. Associations among independent variables, among dependent variables, and predictive paths from the independent to the dependent variables were all estimated. We used latent variables which were all indicated by three parcels. In our previous analyses, we had demonstrated that there was metric invariance for personality, whereas a previous study on portions of the same datasets (Crocetti et al., 2009) established metric invariance for our measure of the GAD symptom of worry. Our depression measure was also found to be metrically invariant, as fit differences between a metric

variance model and a metric invariance model were minimal (ΔCFI was .004 and $\Delta RMSEA$ was .001). Therefore, we constrained factor loadings of the parcels of each of the latent variables to be equal across all eight groups. The resulting model is graphically represented in Fig. 2.

The estimated model had an acceptable fit ($\chi^2(1419) = 4345.989$ ($p < .001$), $CFI = .915$, $RMSEA = .069$ (90% C.I. = .066–.071)). To assess whether specific associations were similar or different across culture, sex, and age groups, we compared the confidence intervals of these associations for the different groups. If the confidence intervals did not overlap, we concluded that there was a significant between-group difference. All associations of Big Five personality traits with depressive symptoms and the Generalized Anxiety Disorder symptom of worry, and the between-group comparisons of these associations are depicted in Table 4.

3.5.1. Associations of Big Five personality traits with depressive symptoms

Table 4 indicates that associations between depressive symptoms and Big Five personality in the eight distinguished groups were predominantly negative. These associations will be discussed trait-by-trait.

Extraversion was a significant negative predictor of depression in middle adolescent Italian boys, and early and middle adolescent Italian girls. However, confidence intervals indicated that these associations did not significantly differ from the non-significant associations in the other groups. Agreeableness was a negative predictor for depression in Dutch early adolescent boys and Italian middle adolescent girls. Again, an inspection of the confidence intervals indicated that these associations did not differ significantly from the non-significant associations found in the other groups. Higher levels of Conscientiousness predicted lower levels of depression in Dutch middle adolescent girls, and Italian middle adolescent boys and girls. These associations did again not differ from the non-significant associations found in the other groups. Emotional Stability negatively predicted depression in six of the eight distinguished groups. These associations only failed to reach significance in Dutch early and middle adolescent boys. Associations of Emotional Stability were stronger in Italian early and middle adolescent boys and girls, when compared to Dutch early adolescent girls and Dutch middle adolescent boys. Openness was a significant positive predictor of depression in Dutch girls, but in none of the other groups. In addition, this association was no stronger than the non-significant associations between Openness and depression we found in the other groups.

3.5.2. Associations of Big Five personality traits with the GAD symptom of worry

The GAD symptom of worry was associated with only three of the Big Five personality traits, as Extraversion and Openness were not associated with worry. In addition, there were substantive between-group differences in the associations we found.

Agreeableness was a significant negative predictor of worry, but only in Dutch early adolescent boys. However, this negative association did not significantly differ from the non-significant associations found in all the other groups. Worry was positively predicted by Conscientiousness in early adolescent Dutch girls, but negatively predicted by this very same trait in middle adolescent Dutch girls. The difference between these two associations was significant. Finally, Emotional Stability was a significant negative predictor of worry in all groups, except for Dutch early adolescent boys. In addition, the negative association of Emotional Stability with worry was significantly stronger in Italian middle adolescent girls, when compared to Dutch early adolescent girls.

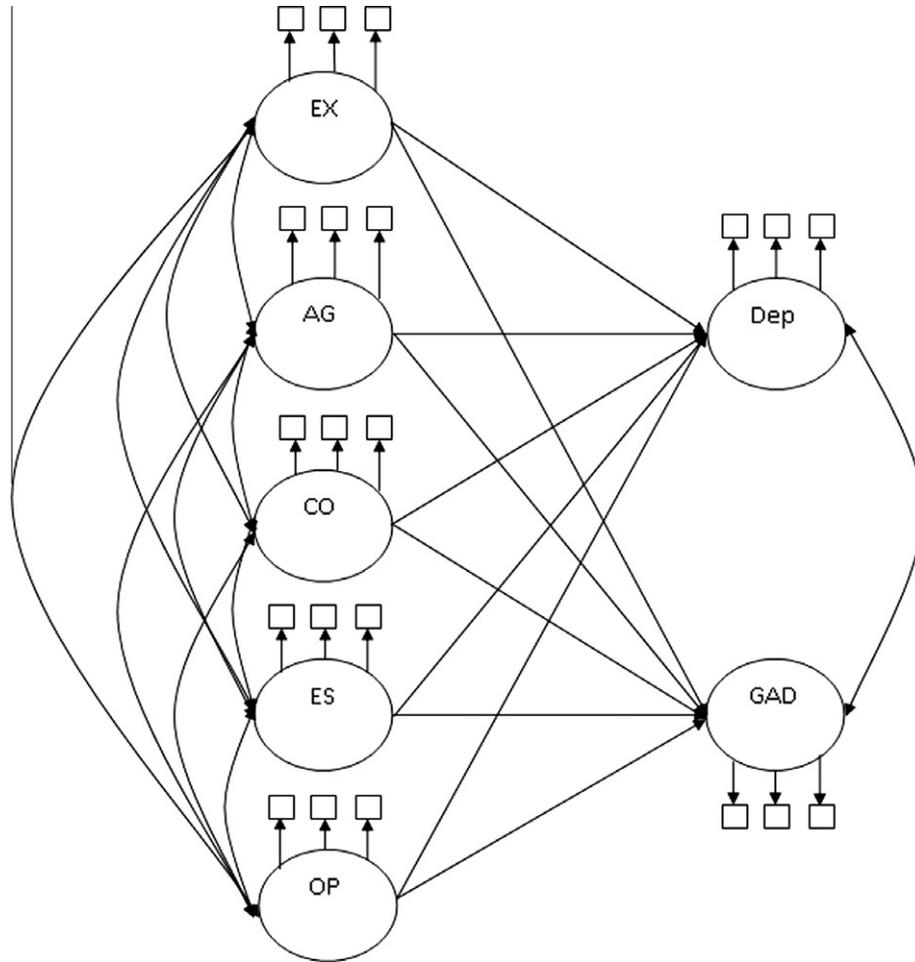


Fig. 2. Estimated structural equation model with the Big personality traits of Extraversion (Ex), agreeableness (Ag), conscientiousness (Co), emotional stability (ES), and openness (Op) as independent latent variables, and depressive symptoms (Dep) and the Generalized Anxiety Disorder symptom of worry (GAD) as dependent latent variables. Circles represent latent variables, squares represent observed item parcels.

Table 4
Associations of Big Five personality traits with depressive symptoms and worry in Italian and Dutch adolescent boys and girls.

	Early adolescents				Middle adolescents			
	Dutch		Italian		Dutch		Italian	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Ex → depression	-.432	-.206	-.049	-.126*	-.127	-.068	-.167*	-.146**
Ag → depression	-.282	-.109	-.064	-.126	-.263**	-.116	-.150	-.129*
Co → depression	-.167	-.040	-.107	-.084	-.114	-.165**	-.160*	-.190***
ES → depression	.166 ^{ab}	-.243*** ^a	-.483*** ^b	-.503*** ^b	-.077 ^a	-.314*** ^{ab}	-.464*** ^b	-.430*** ^b
Op → depression	.323	.033	-.131	-.040	.081	.147 [†]	-.033	.098
Ex → worry	.064	-.071	-.035	-.037	.015	-.040	-.112	.002
Ag → Worry	-.369 ^{ab}	-.024 ^a	-.016 ^{ab}	-.044 ^{ab}	-.269** ^{ab}	-.136 ^{ab}	.027 ^b	-.020 ^{ab}
Co → worry	-.072 ^{ab}	.221 ^a	-.021 ^{ab}	.075 ^{ab}	-.140 ^{ab}	-.130 ^b	.048 ^{ab}	.063 ^{ab}
ES → worry	-.431 ^{ab}	-.500*** ^a	-.482*** ^{ab}	-.591*** ^{ab}	-.420** ^{ab}	-.527*** ^{ab}	-.567*** ^{ab}	-.616*** ^b
Op → worry	.251	-.170	.085	.044	-.001	.110	.017	.023

Note. Ex = Extraversion, Ag = Agreeableness, Co = Conscientiousness, ES = Emotional Stability, Op = Openness. Different superscripts within a line indicate significant ($p < .05$) between-group differences in associations of personality traits with depressive symptoms and worry. Lines without superscript contain no significant between-group differences.

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

4. Discussion

The purpose of the current study was to provide a first cross-national comparison on adolescent self-reported personality traits. Italian adolescents were compared to Dutch adolescents in terms

of mean-levels, sex differences, and personality–psychopathology linkages. In all analyses we also distinguished among early and middle adolescents. However, a prerequisite for an appropriate cross-national comparison is establishing that the instrument one uses measures the construct of interest (i.e., personality) in

the same way in different cultures. In other words, before cross-national comparisons can be interpreted, measurement equivalence must be established.

In the present study, we were able to establish configural invariance. That is, we were able to show that personality as measured with a 30-item Dutch version of Goldberg's (1992) Big Five measure is appropriately described by roughly the same five-factor structure for Italian and Dutch adolescents. Cross-national configural invariance of the Big Five personality traits had previously been established for self-ratings of adult and college student samples (McCrae, 2001; Schmitt et al., 2007), observer ratings on adult and college student samples (McCrae, Terraciano, & 78 members of the personality profiles of cultures project, 2005), and for observer ratings on adolescents (McCrae et al., 2010), but the present study is the first to establish configural invariance for adolescent Big Five self-reports.

However, it has recently been argued that only establishing configural invariance is not sufficient for cross-national comparisons. In fact, it has been recommended to proceed with tests of metric and scalar invariance (Nye et al., 2008). The former can be considered a requirement for comparing associations among variables across groups, whereas the latter is advisable when one wants to compare means (cf. Vandenberg & Lance, 2000). Both metric and scalar invariance can be examined with Confirmatory Factor Analyses. Such analyses can only be applied if the N for each of the investigated countries is at least 100 to 200, but with more complex models with large number of free parameters (such as a five-latent-factor models with usually about 6 to 8 indicators for each of the factors) larger samples ($N > 200$) may be necessary in order to have models converge (Kline, 2005). It is perhaps because such sample sizes are usually not obtained for each of the participating countries that most previous studies only examined configural invariance with exploratory factor analyses, and did not test for metric and scalar invariance with Confirmatory Factor Analyses.

In the present study, we did have a sufficiently large sample size to proceed with tests for metric and scalar invariance. These tests revealed that factor loadings were equivalent across all nationality, sex and age groups, which indicates that we managed to establish metric equivalence. As a result, associations between variables can be compared across these groups. We were, however, unable to establish scalar invariance across all groups. Post-hoc scalar invariance tests within and across cultures revealed that scalar invariance was only attainable within the Dutch sample. Therefore, mean comparisons between Dutch age and sex groups can be readily interpreted, which underscores the validity of results obtained in previous studies that were conducted on the portion of the Dutch participants that was followed longitudinally (e.g., Klimstra et al., 2009; Meenus, van de Schoot, Klimstra, & Branje in press). Moreover, these findings reveal that expecting to establish scalar invariance might not be as unrealistic as has been argued in previous studies (e.g., Byrne, Shavelson, & Muthen, 1989).

Within the Italian sample, scalar invariance could not be established as imposing such constraints led to a significantly worse fit. However, although the changes in the CFI passed the recommended threshold ($>.010$; Chen, 2007), the RMSEAs were similar. Moreover, the CFI threshold was not violated in a dramatic way (Δ CFIs for gender and age invariance tests within the Italian sample were .012 and .017). Thus, caution is warranted in interpreting mean-level differences between sex and age groups within the Italian sample, but our findings would suggest that such differences still can be interpreted, albeit carefully. The same cannot be said of scalar invariance across cultures, as imposing these constraints lead to a dramatic drop in the CFI and a substantive increase in the RMSEA. Thus, when interpreting overall mean-level differences or when comparing the magnitude of mean-level and sex

differences between the Dutch and Italian adolescents in the current study, one should bear in mind that these differences may be due to the slightly different meanings these traits have for Dutch and Italian adolescents. That is, the items included in the scales of our personality measure are endorsed differently by Italian adolescents when compared to Dutch adolescents. Put more simply, a high level of Agreeableness among Italian adolescents may be mainly due to high scores on items like "cooperative", whereas a high level of the same trait may be mainly due to high scores on a slightly different item like "pleasant" in Dutch adolescents. We are not the first that have been unable to establish scalar invariance in a cross-national comparison; several previous studies (Ellis et al., 1993; Huang et al., 1997; Johnson et al., 2008; Nye et al., 2008) that also applied these rigorous tests also were unable to establish scalar invariance in several instances. Given that establishing scalar invariance was mainly problematic across cultures, and less of an issue within cultures, the present study suggests that comparing mean-levels across countries might be problematic. Therefore, the findings of previous cross-national comparisons of mean-levels in which scalar invariance has not been examined (e.g., Allik & McCrae, 2004; McCrae, 2001; McCrae, Terraciano, & 78 members of the personality profiles of cultures project, 2005; McCrae et al., 2010; Schmitt et al., 2007) should be interpreted cautiously. The same applies to the cross-cultural mean-level differences assessed in the present study, which will now be discussed.

Some evidence was found for cross-cultural mean-level differences on Agreeableness and Openness. However, these differences only applied to early adolescent girls for Agreeableness and middle adolescent girls for Openness. For Extraversion, cross-cultural differences were found for both early and middle adolescent girls, with Dutch girls reporting higher levels of Extraversion than Italian girls. Dutch adolescents tend to report having more friends than Italians (Currie et al., 2008), and Extraversion is known to be associated with the number of friends one nominates (Selfhout et al., 2010). Therefore, our findings do not come as a surprise. It is, however, unclear why these anticipated cross-cultural differences in Extraversion only emerged for adolescent girls. One obvious reason could be the lack of scalar invariance across cultures, and within the Italian culture. This could contribute to different adolescent groups endorsing the same items differently. As a result, Extraversion might have a slightly different meaning in boys when compared to girls, and in Italians when compared to Dutch. Thus, before theorizing on these differences one first needs to rule out that findings are due to a lack of scalar invariance. The largest and most consistent cross-cultural differences emerged for Emotional Stability, where all Dutch age and sex groups displayed higher levels than their Italian counterparts. Although an overly ambitious interpretation of these findings would not be in place due to the lack of scalar invariance, our findings are in line with our hypotheses based on cross-cultural mean-level differences in anxiety (Crocetti et al., 2009) and highly consistent across sex and age groups. Therefore, we feel we can cautiously state that Italian adolescents appear to be less successful in dealing with negative emotions effectively, when compared to Dutch adolescents.

We also proceeded with an examination of sex differences within Italian and Dutch samples, and then compared the magnitude of the sex differences across samples. In all these analyses, we distinguished among early and middle adolescents. Given the aforementioned problems with lacking scalar invariance, especially the cross-national comparisons of the magnitude of sex differences need to be interpreted cautiously. In Dutch adolescents, sex differences in Agreeableness mainly applied to early adolescents where girls displayed higher scores than boys. All other prominent sex differences among Dutch adolescents applied to middle adolescents, as middle adolescent girls were more conscientious and open to experience, but less emotionally stable when compared

to middle adolescent boys. Sex differences in Agreeableness, Conscientiousness, and Emotional Stability were replicated among Italian adolescents, but for Openness no sex differences were found. In addition, sex differences in Agreeableness and Conscientiousness applied to both early and middle adolescents. Our findings for Italian adolescents with regard to Emotional Stability perfectly replicate our findings obtained among Dutch adolescents, as these findings also only applied to middle adolescents in our Italian sample.

A cross-national comparison of the magnitude of sex differences revealed that, given cross-national differences in the endorsement of specific items (signified by a lack of scalar invariance), sex differences were somewhat larger in Italy than in The Netherlands for Extraversion, Agreeableness, and Emotional Stability. A first explanation for this finding would be that the items are endorsed in such a way that trait-scores are closely linked to sex stereotypes in Italy, whereas these traits may have a somewhat more neutral connotation in The Netherlands. Alternatively, sex differences may be larger in Italy due to “machismo”. Indeed, Hofstede’s (2001) study on dimensions of culture revealed that Italy was one of the most masculine countries (rank 4 out of 53), whereas The Netherlands was among the least masculine countries (rank 51 out of 53). On the other hand, effect sizes of sex by nationality interactions were only small and should therefore not be over-interpreted. In addition, definite explanations of cross-national differences in the magnitude of sex differences can only be drawn once it has been established how the connotation of Big Five personality traits exactly differs for Dutch and Italian adolescents. A cross-national comparison of personality–psychopathology linkages may provide some preliminary insights into this issue.

Personality–psychopathology linkages were found to be quite similar for Italian and Dutch adolescents with regard to depressive symptoms, as all but one of the Big Five traits (i.e., Openness) were negatively associated with depressive symptomatology in both samples. Openness positively predicted depressive symptoms, but only in Dutch middle adolescents. However, this marginally significant predictive path was not significantly different from the non-significant predictive paths found in the other sex, age, and nationality groups. Cross-national differences did emerge for linkages between Emotional Stability and depressive symptoms, as these were stronger among Italian early adolescent girls than among Dutch early adolescent girls, and stronger among Italian middle adolescent boys when compared to Dutch middle adolescent boys. Thus, Emotional Stability appears to be a better predictor of depressive symptomatology among Italians than among the Dutch. This may suggest that depressive symptoms are more likely to be caused by stable dispositions such as personality among Italians, whereas external causes (e.g., life events, family climate, quality of peer relations) may be a more common cause of depressive symptoms among Dutch. Future studies with a longitudinal focus are needed to replicate our findings before further inferences can be drawn. However, there would be profound implications for the treatment of depression if the extent to which depressive symptoms are caused by stable dispositions instead of more changeable environmental factors indeed differs from one culture to another.

For the Generalized Anxiety Disorder symptom of worry we found fewer linkages with personality traits than for depressive symptoms. Moreover, there were few cross-national differences in these associations. Extraversion and Openness were unassociated with worry, whereas there was only a significant (negative) association between Agreeableness and worry in Dutch middle adolescent boys. Conscientiousness was significantly, but only marginally, associated with worry in Dutch girls. However, whereas Conscientiousness appeared to amount to worry in early adolescence, it seemed to protect against worry in middle

adolescence. Thus, even though there was scalar invariance for Conscientiousness for early adolescent girls when compared to middle adolescent girls, the meaning of this trait still seems to differ slightly for these two age groups. While we do not want to over-interpret the marginally significant positive versus negative associations in Dutch early and middle adolescent girls, respectively, our findings do seem to suggest that establishing scalar invariance does not yet mean that the meaning of a certain trait is perfectly identical for two groups. Thus, we would recommend considering associations with external variables such as problem behavior in addition to measurement invariance tests when examining whether the connotation of traits is similar or different across age, sex, and nationality groups.

Of all personality traits, Emotional Stability had the strongest relations with worry. These associations appeared to be largely equal across cultures. Despite these strong relations, a previous study demonstrated that Emotional Stability and worry are two distinct constructs (Hale et al., 2010). In that study of Dutch adolescents, more emotionally stable individuals were further found to worry less on the subsequent measurement occasion, but less worried individuals were also found to be more emotionally stable on the subsequent measurement occasion. Thus, worry and Emotional Stability affected one another in a reciprocal fashion. An interesting endeavor for future studies would be to investigate whether the same developmental (i.e., longitudinal) processes with regard to Emotional stability and worry can be replicated in different cultures.

Overall, we found fewer linkages of personality traits with problem behavior symptoms than previous studies (Barbaranelli et al., 2003; Hale et al., 2010; Klimstra et al., 2010), as there were very few associations of Conscientiousness and Extraversion with worry and depressive symptoms in the current study. In fact, only Emotional Stability was consistently associated with problem behavior symptoms. This contrast with previous studies is probably caused by the fact that we accounted for the associations among Big Five personality traits. That is, the associations we present here are correlations between problem behavior symptoms and the unique aspect of each Big Five trait, corrected for the overlap between Big Five traits. As such, our findings could suggest that previous studies merely found associations of Conscientiousness and Extraversion with problem behavior symptoms, because these two traits partly overlap with Emotional Stability.

Another important observation regarding the personality–psychopathology linkages in the current study is that the pattern of associations of personality with depression seems largely similar to the pattern of associations of personality with worry. There are, however, subtle differences. The most prominent one refers to Extraversion that seems to protect (Italian) adolescents against depressive symptoms, whereas this trait does not serve such a role with regard to worry. Subtle differences also emerge for Emotional Stability. This trait appears to be stronger related to worry than to depressive symptoms. Moreover, whereas there are relatively consistent cross-national differences in associations of depressive symptoms with Emotional Stability, such cross-national differences do not seem to emerge with regard to the associations of worry with Emotional Stability. Together, these subtle differences in linkages between personality with depressive symptoms when compared to linkages between personality with worry suggest that a broad internalizing problem behavior factor might lead to somewhat too general conclusions regarding personality–psychopathology linkages. As such, our findings seem to favor the specificity of research on psychopathology guided by DSM-IV-TR criteria (Diagnostic and Statistical Manual of the Mental Disorders, Fourth Edition, Text Revision; American Psychiatric Association, 2000) in which the symptoms of disorders such as Depression and the Generalized Anxiety Disorder symptom of

worry are examined separately, over approaches in which these disorders are (together with other related disorders) collapsed into one broad internalizing factor (e.g., Watson & Clark, 1984).

Despite some subtle cross-national differences, personality–psychopathology linkages do not sufficiently clarify why cross-national scalar invariance could not be established in the current study. Hence, it is unclear how the connotation of personality traits differs in The Netherlands and Italy. Therefore, we would recommend further research into cross-national differences in the meaning of traits. In this regard, research on the motivational underpinnings of personality traits might be very relevant. Denissen and Penke (2008) have recently subsumed descriptions of these motivational underpinnings, and provided their own perspective on this topic. A brief glance at the overview and their own ideas reveals that there is quite a bit of variation in descriptions of motivational underpinnings of the Big Five. For example, Agreeableness has been described as a disposition to react cooperatively in resource conflict (Denissen & Penke, 2008), but also as a trait facilitating positive family relationships (MacDonald, 1995, 1998). Similarly, Conscientiousness has been described in motivational terms as the tenacity of goal pursuit (Denissen & Penke, 2008), but also as trustworthiness and dependability (Hogan, 1996). It is hard to assess which description of motivational underpinning is the most accurate, but it could very well be that these motivations are culture-specific. Therefore, one suggestion for future studies would be to examine whether and how motivational underpinnings of Big Five personality traits differ from one culture to another. To address this topic more thoroughly, there is at least qualitative data on cultural differences in motivational underpinnings needed. For this purpose, cultural anthropologists could be a great help to psychologists in exploring cross-cultural differences in the meaning of personality traits.

4.1. Strengths and limitations

Besides the already discussed advantages of the present study, there are several additional strengths. A first strength concerns our samples. The advantages of the sizes of our samples has been described previously, but our samples were also more representative than the college student samples that are typically employed in cross-cultural research. More specifically, we sampled adolescents representing all various educational levels of high school available in The Netherlands and Italy. Because high school attendance is obligatory in both these countries, our samples were quite representative for the general population of adolescents. Previous studies usually employed college student samples (Schmitt et al., 2007, 2008) that are not necessarily representative for the general population. Even though it has been shown that mean-level differences and sex differences found among college students are quite similar to those found among more representative adults samples (Costa et al., 2001; McCrae, 2001), future cross-cultural studies should seek to employ larger and more representative samples, like the ones employed in the present study.

A second strength concerns the examination of cross-cultural differences of personality with two distinct types of psychopathological symptoms (i.e., depressive symptoms and the Generalized Anxiety Disorder symptom of worry). Cross-cultural research on this topic was lacking, while the present study shows that there can be differences between countries in such linkages that could potentially provide insights into the different connotation personality traits may have across the globe.

Besides these strengths, the present study also has several limitations. A first limitation concerns our focus on only two European countries. In the present study, we tentatively attributed cross-cultural differences in the magnitude of sex differences to masculinity. To be more certain that masculinity indeed plays a role in

the magnitude of sex differences, multiple masculine cultures should be compared to multiple feminine cultures.

Our reliance on a measure with adjectives may be considered a second limitation. Phrase items may be more appropriate for cross-cultural comparisons (McCrae & Costa, 1997), but the content of such phrase items is often also heavily dependent on adjectives (Nye et al., 2008). That is, the whole meaning of the phrase is often predominantly determined by the meaning of the one adjective in that phrase. Therefore, it is unlikely that we would have found scalar invariance if we would have used phrase items.

Overall, our Italian–Dutch comparison can be considered an illustration of a more general problem. That is, cross-cultural research may be appealing, but extreme caution needs to be warranted when interpreting the findings of such studies. As previous cross-cultural research has not always taken the appropriate precautions, we would recommend the use of rigorous tests of measurement invariance and careful analyses of the meaning of items in different cultures before proceeding to interpreting cross-cultural differences in personality traits.

References

- Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology*, 35, 13–28.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association Inc. [Text Revision (DSM-IV-TR)].
- Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the big five in late childhood. *Personality and Individual Differences*, 34, 645–664.
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., et al. (1997). The screen for child anxiety related emotional disorders (scared): Scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 545–553.
- Branje, S. J. T., van Lieshout, C. F. M., & van Aken, M. A. G. (2004). Relations between big five personality characteristics and perceived support in adolescents' families. *Journal of Personality and Social Psychology*, 86, 615–628.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 436–456.
- Caprara, G. V., Barbaranelli, C., Bermudez, J., Maslach, C., & Ruch, W. (2000). Multivariate methods for the comparison of factor structures in cross-cultural research: An illustration with the Italian Big Five Questionnaire. *Journal of Cross Cultural Psychology*, 31, 437–464.
- Caprara, G. V., Barbaranelli, C., Pastorelli, C., & Cervone, D. (2004). The contribution of self-efficacy beliefs to psychosocial outcomes in adolescence. Predicting beyond global dispositional tendencies. *Personality and Individual Differences*, 37, 751–763.
- Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, 56, 453–484.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Sex differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81, 322–331.
- Crocetti, E., Hale, W. W., III, Fermani, A., Raaijmakers, Q. A. W., & Meeus, W. H. J. (2009). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in the general Italian adolescent population: A validation and a comparison between Italy and The Netherlands. *Journal of Anxiety Disorders*, 23, 824–829.
- Crocetti, E., Rubini, M., & Meeus, W. H. J. (2008). Capturing the dynamics of identity formation in various ethnic groups: Development and validation of a three-dimensional model. *Journal of Adolescence*, 31, 207–222.
- Crocetti, E., Schwartz, S. J., Fermani, A., & Meeus, W. H. J. (2010). The Utrecht-management of identity commitments scale (U-MICS): Italian validation and cross-national comparisons. *European Journal of Psychological Assessment*, 26, 172–186.
- Currie, C., Gabbaini, S. N., Godeau, E., Roberts, C., Smith, R., Currie, D., et al. (2008). *Inequalities in young people's health. Health behaviour in school-aged children international report from the 2005/2006 survey*. Copenhagen: World Health Organization.
- De Fruyt, F., de Bolle, M., McCrae, R. R., Terracciano, A., & Costa, P. T. Collaborators of the Adolescent Personality Profiles of Cultures Project. (2009). Assessing the universal structure of personality in early adolescence. The NEO-PI-R and NEO-PI-3 in 24 cultures. *Assessment*, 16, 301–311.
- De Raad, B., Barelds, D. P. H., Levert, E., Ostendorf, F., Mlacic, B., Di Blas, L., et al. (2010). Only three factors of personality description are fully replicable across

- languages: A comparison of fourteen trait taxonomies. *Journal of Personality and Social Psychology*, 98, 160–173.
- Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the five-factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, 42, 1285–1302.
- Ellis, B. B., Becker, P., & Kimmel, H. D. (1993). An item response theory evaluation of an English version of the Trier Personality Inventory (TPI). *Journal of Cross-Cultural Psychology*, 24, 133–148.
- Gerris, J. R. M., Houtmans, M. J. M., Kwaaitaal-Roosen, E. M. G., Schipper, J. C., Vermulst, A. A., & Janssens, J. M. A. M. (1998). *Parents, adolescents and young adults in dutch families: A longitudinal study*. Nijmegen: Institute of Family Studies University of Nijmegen.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4, 26–42.
- Hale, W. W., III, Klimstra, T. A., & Meeus, W. H. J. (2010). Is the generalized anxiety disorder symptom of worry just another form of neuroticism? A five-year longitudinal study of adolescents from the general population. *Journal of Clinical Psychiatry*, 71, 942–948.
- Hale, W. W., III, Raaijmakers, Q. A. W., Muris, P., & Meeus, W. H. J. (2005). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders in the general adolescent population. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 283–290.
- 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 problem. *Journal of Personality and Social Psychology*, 82, 903–918.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (second ed.). Beverly Hills, CA: Sage.
- Hogan, R. (1996). A socioanalytic perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five factor model of personality: Theoretical perspectives* (pp. 163–179). New York: Guilford Press.
- Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, 28, 192–218.
- Johnson, W., Spinath, F., Krueger, R. F., Angleitner, A., & Riemann, R. (2008). Personality in Germany and Minnesota: An IRT-based comparison of MPQ self-reports. *Journal of Personality*, 76, 665–706.
- Klimstra, T. A., Akse, J., Hale, W. W., III, Raaijmakers, Q. A. W., & Meeus, W. H. J. (2010). Longitudinal associations between Big Five personality traits and problem behavior in adolescence. *Journal of Research in Personality*, 44, 273–284.
- Klimstra, T. A., Hale, W. W., III, Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. H. J. (2009). Maturation of personality in adolescence. *Journal of Personality and Social Psychology*, 96, 898–912.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: Guilford.
- Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "Big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, 136, 768–821.
- Kovacs, M. (1985). The children's depression inventory (cdi). *Psychopharmacology Bulletin*, 21, 995–998.
- Kovacs, M. (1988) Children's depression inventory. Questionario di autovalutazione [Children's depression inventory. Self-report questionnaire] (Adattamento italiano a cura di M. Camuffo, R. Cerutti, L. Lucarelli, R. Mayer) [Italian adaptation by M. Camuffo, R. Cerutti, L. Lucarelli, R. Mayer]. Firenze: O.S. Organizzazioni Speciali.
- Krueger, R. F. (1999). Personality traits in late adolescence predict mental disorders in early adulthood: A prospective-epidemiological study. *Journal of Personality*, 67, 39–65.
- Lewis-Fernandez, R., & Kleinman, A. (1994). Culture, personality and psychopathology. *Journal of Abnormal Psychology*, 103, 67–71.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173.
- Luyckx, K., Soenens, B., & Goossens, L. (2006). The personality-identity interplay in emerging adult women: Convergent findings from complementary analyses. *European Journal of Personality*, 20, 195–215.
- MacDonald, K. B. (1995). Evolution, the Five-Factor Model, and levels of personality. *Journal of Personality*, 63, 525–567.
- MacDonald, K. B. (1998). Evolution, culture, and the five-factor model. *Journal of Cross-Cultural Psychology*, 29, 119–149.
- Marsh, H. W., & Hau, K.-T. (1999). Confirmatory factor analysis: Strategies for small sample sizes. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 252–284). Thousand Oaks, CA: Sage.
- McCrae, R. R. (2001). Trait psychology and culture: Exploring intercultural comparisons. *Journal of Personality*, 69, 819–846.
- McCrae, R. R., & Costa, P. T. Jr., (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509–516.
- McCrae, R. R., & John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175–215.
- McCrae, R. R., & Terracciano, A. (2005). Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, 88, 547–561.
- McCrae, R. R., Terracciano, A., De Fruyt, F., De Bolle, M., Gelfand, M. J., Costa, P. T., et al. (2010). The validity and structure of culture-level personality scores: Data from ratings of young adolescents. *Journal of Personality*, 78, 815–838.
- Meeus, W. H. J., van de Schoot, R., Klimstra, T. A., & Branje, S. J. T. (in press). Personality types in adolescence: Change and stability and links with adjustment and relationships. A five-wave longitudinal study. *Developmental Psychology*.
- Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (fourth ed.). Los Angeles, CA: Muthén & Muthén.
- Nye, C. D., Roberts, B. W., Saucier, G., & Zhou, X. (2008). Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality*, 42, 1524–1536.
- Peabody, D., & de Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of Personality and Social Psychology*, 83, 983–997.
- Perugini, M., & Richetin, J. (2007). In the land of the blind, the one-eyed man is king. *European Journal of Personality*, 21, 977–981.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage Publications.
- Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martínez, V., Alcalay, L., Ault, L., et al. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38, 173–212.
- Schmitt, D. P., Raelo, 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.
- Selfhout, M., Burk, W., Branje, S. J. T., Denissen, J. J. A., van Aken, M. A. G., & Meeus, W. H. J. (2010). Emerging late adolescent friendship networks and big five personality traits: A social network approach. *Journal of Personality*, 78, 509–538.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with CFA and IRT: Toward a unified strategy. *Journal of Applied Psychology*, 91, 1292–1306.
- Tackett, J. L. (2006). Evaluating models of the personality-psychopathology relationship in children and adolescents. *Clinical Psychology Review*, 26, 584–599.
- Timbremont, B., & Braet, C. (2002). *Children's depression inventory: Nederlandstalige versie [children's depression inventory: Dutch version]*. Lisse: Swets & Zeitlinger.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465–490.
- Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social Psychology*, 99, 174–190.