Title: Development and Initial Validation of the Five Factor Model Adolescent Personality Questionnaire (FFM-APQ)

Authors: Mary E. Rogers[1] and A. Ian Glendon[2]

Author Affiliations: [1] Griffith University, Griffith Business School, Department of Employment Relations and Human Resources, and Centre for Work, Organisation and Wellbeing
[2] Griffith University, School of Applied Psychology, Menzies Health Institute Queensland and Centre for Work, Organisation and Wellbeing

Corresponding Author   Dr Mary Rogers
Griffith University
Griffith Business School, Department of Employment Relations and Human Resources
Gold Coast Campus, QLD, 4222, Australia.

Telephone: +61 7 5552 7070
Facsimile: +61 7 5552 9206

Author Email Addresses: mary.rogers@griffith.edu.au
i.glendon@griffith.edu.au

Development and Initial Validation of the Five-Factor Model Adolescent Personality

Questionnaire (FFM-APQ)

ABSTRACT

This research reports on the 4-phase development of the 25-item Five-Factor Model Adolescent

Personality Questionnaire (FFM-APQ). The purpose was to develop and determine initial evidence

for validity of a brief adolescent personality inventory using a vocabulary that could be understood by

adolescents aged up to 18 years. Phase 1 ($N = 48$) consisted of item generation and expert ($N = 5$)

review of items; Phase 2 ($N = 179$) involved item analyses; in Phase 3 ($N = 496$) exploratory factor

analysis assessed the underlying structure; in Phase 4 ($N = 405$) confirmatory factor analyses resulted

in a 25-item inventory with five subscales.

Keywords: five factor model of personality; exploratory factor analysis; confirmatory factor analysis;

test development, personality inventory

Development and Initial Validation of the Five-Factor Model Adolescent Personality

Questionnaire (FFM-APQ)

Adolescence is a critical period in an individual's development when young people strive to establish their identity to understand aspects of themselves, including their personality, or those individual differences comprising their core self. The nature and development of adolescent personality is under-researched (Shiner & Caspi, 2003). This could be partly due to a lack of self-report measures designed for adolescents, or that some measures are too long, or may include questions that adolescents don't understand or cannot answer. The aim of this research was to develop an adolescent personality inventory, based on the five-factor model (FFM) of personality using a vocabulary that could be understood by contemporary adolescents aged up to 18 years, and provide some initial evidence of validation. While developing a useful and accessible measure is an aim of this paper, an aligned purpose is to make a theoretical contribution to personality research that further elucidates adolescent personality development.

Five-Factor Model of Personality

A widely accepted model of personality, the FFM integrates a number of personality constructs that can be used in applied settings (McCrae & John, 1992). The five personality factors are typically referred to as: agreeableness (A), extraversion (E), conscientiousness (C), neuroticism (N), and openness to experience (O). While most research is based on adults, the FFM has been applied to samples of adolescents and children (e.g., Deal, Halverson, Martin, Victor, & Baker, 2007; Goldberg, 2001; McCrae et al., 2002; Roberts & DelVecchio, 2000; Shiner, 1998, 2000; Shiner & Caspi, 2003).

Adolescent Personality and Relationships

FFM traits have been related to interpersonal processes during adolescence. Peer acceptance is important to adolescent development as friendships provide an environment for learning interpersonal skills, gathering information about the self, and fulfilling social and belongingness needs (Buhrmester, 1996). Agreeableness and extraversion are key in developing healthy adolescent relationships. As they tend to avoid interpersonal conflict, agreeable individuals are more likely to be cooperative, maintain positive social relationships, be chosen as friends, and have better interpersonal adjustment. High-extraversion young people are more socially competent, active, and motivated to seek out

friends (Asendorpf & van Aken, 2003; Knack, Jacquot, Jensen-Campbell, & Malcolm, 2013; Selfhout et al., 2010). Conscientiousness and openness have been associated with maintaining and enhancing existing relationships. Conscientious individuals are more likely to maintain friendships, while high-openness adolescents tend to choose similarly openness friends (Jensen-Campbell & Malcolm, 2007; Selfhout et al., 2010). Neuroticism tends to be negatively associated with relationships. Adolescents who are worried, anxious, and withdrawn may view social situations as threatening, and be more likely to experience social difficulties and victimisation (Jensen-Campbell & Malcolm, 2007).

Adolescent Personality Stability

Research on the stability of personality during adolescence has produced mixed results. While some have identified personality changes during adolescence (Kawamoto & Endo, 2015), others have argued that it is reasonably stable. For example, in reviewing longitudinal studies, McCrae et al. (2002) found that neuroticism appeared to increase in girls, and that openness increased in boys and girls, while extraversion, agreeableness, and conscientiousness were relatively more stable. Pullmann, Raudsepp and Allik (2006) also found an increase in openness. Roberts and DelVecchio (2000) reported that personality traits were stable between the ages of 6 and 18, with a gradual increase in trait consistency between 18 and 22 years. Judge, Higgins, Thoresen, and Barrack (1999) found that adolescent personality traits were stable over time, with conscientiousness being the most, and agreeableness the least stable. Donnellan, Conger and Burzette (2007) proposed that more mature adolescents experienced fewer personality changes. These findings suggest that while there was evidence for consistency among some traits during adolescence, some dispositional attributes become more stable during the transition to adulthood. More research on the stability of personality traits during adolescence is needed and shorter personality measures might assist in this endeavor.

Personality Measures

Personality measures developed for adults have been used with adolescents. De Fruyt, De Bolle, McCrae, Terraciano, and Costa (2009), and De Fruyt, Mervielde, Hoekstra, and Rolland (2000) used the NEO-PI-R (Costa & McCrae, 1992b). Limont, Dreszer-Drogorob, Bedynska, Sliwinska, and Jastrzebska (2014), and Sneed, Gullone, and Moore (2002) used the NEO-FFI (Costa & McCrae, 1992a). Kawamoto and Endo (2015) used the Yatabe-Guilford Personality Inventory (Yatabe, 1975),

Randler (2008) used the Big Five Inventory 10 item (BFI-10; Rammstedt & John, 2007), and

Baldasaro, Shanahan, and Bauer (2013) used the Mini IPIP, a 20-item short form of the 50-item

International Personality Item Pool (IPIP; Donnellan, Oswald, Bairds, & Lucas, 2006). The most

comprehensive measure is the NEO-PI-R (Costa & McCrae, 1992b) – a 240-item version of the FFM

requiring an adult vocabulary level, is time consuming to administer. A possible reason for using adult

measures is that few validated personality inventories have been designed for adolescents.

It is beyond this review to describe inventories developed for clinical adolescent populations and

children. Inventories developed to measure personality traits in general adolescent populations include

the High School Personality Questionnaire (Cattell & Cattell, 1975), and the 48-item California Child

Q-set (CCQ) personality scale – designed for boys aged 12-13 (John, Caspi, Robins, Moffitt, &

Stouthamer-Loeber, 1994). FFM-based measures include the 56-item Adolescent Personality Style

Inventory (Lounsbury et al., 2003), and the 240-item NEO-PI-3 (McCrae, Costa, & Martin, 2015),

which assesses both adult and adolescent personality. Of these, only the CCQ is freely available to

researchers, and while relatively short, was developed for use with young adolescent boys.

The importance of developing a valid, FFM measure that can be understood by young people from

age eleven was identified by Sneed et al. (2002), and by Soto, John, Gosling, and Potter (2008) who

both assessed the reliability and factor structures of FFM measures among adolescent populations.

Administering a short personality measure to young people in research and school settings is practical

when considering that assessment time is often limited, and that some participants will return

incomplete questionnaires due to fatigue or boredom. As interest in the FFM continues, a need for

brief measures is likely to increase (Donnellan et al., 2006; Gosling, Rentfrow, & Swan, 2003;

Rammstedt & John, 2007). As little adolescent research uses the FFM, we aimed to develop a short

FFM-based personality measure that could be used to study personality in adolescent populations.

Method and Results

The research involved four phases. Phase 1 comprised item generation and expert review of items;

Phase 2 involved item analyses; in Phase 3 exploratory factor analysis assessed the underlying

structure; in Phase 4 confirmatory factor analyses were applied. Convenience samples were used and

almost all participants were Caucasian, which is typical of high school populations in Australia where large racial or ethnic groupings are relatively rare. Scale development followed a standard multi-phase trajectory, similar to Holt, Armenakis, Field, and Harris (2007), and was consistent with processes described by the American Educational Research Association (2014). All phases were conducted under the auspices of the authors' University Human Research Ethics Committee.

Phase 1: Item Generation and Expert Review

The purpose of Phase 1 was to generate a pool of items that reflected a contemporary adolescent vocabulary to represent FFM components. To provide initial evidence for validity of content, independent experts who had knowledge of scale development, personality theory, and test administration, rated the items. We used both a deductive (theoretical classification), and inductive (respondents described behaviors relating to FFM adjectives) approach to item generation (Hinkin, 1998). To develop items that included language familiar to the target respondents (Hinkin, 1998), we conducted focus groups with adolescents. As one aim was to produce a scale that contained no more than 30 items, to err on the side of inclusiveness the pool of items to be developed needed to contain approximately twice as many items as targeted for the final form of the instrument (Clark & Watson, 1995). Items needed to be as simple and as short as possible (Hinkin, 1998).

Participants

Focus groups. Participants were 48 girls, aged 12-17 years, enrolled in grades 8, 9, 10, and 11 at a public state high school in SE Queensland attracting students from low to middle socio-economic status (SES). While access was to a girls-only convenience sample, there were eight 6-person focus groups; two groups from each grade.

Expert reviewers. These were three research psychologists and two practising psychologists who were independent of the authors and skilled in test development and/or personality test administration.

Materials

FFM adjectives were generated as a basis for open-ended questions in the focus groups. Based on personality trait research (Goldberg, 1990; McCrae & Costa, 1987; McCrae & John, 1992), there were four adjective sets for each personality dimension. Sample questions on each dimension were:

- A: *Think of someone you know who is trustworthy. What do they do or say that makes you think they are trustworthy?*

- E: *Think of someone you know who is friendly and outgoing. What do they do or say that makes you think they are friendly and outgoing?*

- C: *Think of someone you know who is conscientious. What do they do or say that makes you think they are conscientious?*

- N: *Think of someone you know who is worried and anxious a lot of the time. What do they do or say that makes you think they are worried and anxious?*

- O: *Think of someone you know who is curious about things. What do they do or say that makes you think they are curious about things?*

Procedure and Results

The semi structured focus group discussions were of approximately 60 minutes duration, facilitated by the first author, and audio-taped. Participants' names were placed in a $10 cash prize draw and discarded after the prize draw to retain anonymity. Transcribed recordings contained all of the participants' comments categorized by the five personality dimensions. The first author reviewed the content to highlight comments common to all year levels. Both authors then reviewed the content and wrote 88 items based on: 1) comments common to all year levels, 2) comments that discriminated between factors, and 3) comments that were behavioral in content (Hinkin, 1998). No negatively worded (reverse scored) items were included.

The expert reviewers evaluated the content of the 88 items according to how adequately they judged each item tapped the domain indicated (5-point response format; endpoints 0 *not at all adequately*, 4 *extremely adequately*), and to comment on item readability (i.e., wording, length). Items rated either 0 or 1 by a majority of reviewers were excluded from subsequent phases. Some expressions were re-worded to reduce the likelihood that items might become dated. After deleting items that might lead to floor or ceiling effects, 82 items remained (N 14, E 18, O 15, A 16, C 19).

Phase 2: Item Analysis

The primary aim of this phase was to reduce the number of items representing each dimension (to approximately 10 per scale). A secondary aim was to obtain further evidence for test validity.

Participants

To maintain consistency during the early stages of the study, we accessed a new convenience sample similar to those who participated in Phase 1. Participants were 179 females (aged 12-18 years, $M = 14.91$, $SD = 1.44$), in grades 8 ($n = 31$, 17.3%), 9 ($n = 35$, 19.6%), 10 ($n = 50$, 27.9%), 11 ($n = 34$, 19.0%), and 12 ($n = 29$, 16.2%) at two middle level SES high schools in SE Queensland.

Procedure and Results

Participants responded to the 82 personality items from Phase 1 using a 5-point Likert scale (1 *strongly disagree* to 5 *strongly agree*) using a paper-based format. To control for order effects, items were presented randomly. Participants remained anonymous and were required to return a signed parental consent form before participating.

SPSS v22 was used for all Phase 2 analyses, entering the correlation matrix in each case. As recommended by Klein (2000), item analysis was used to select items most related to each personality construct. Distributions for each item (e.g., skew, kurtosis), correlation matrices, and inter-item and item-total correlations within each of the five subscales were examined. Items were removed one at a time and the subsequent factor structure was rechecked after each item deletion. Reasons for deletion were: non-normal distribution, loading on an "incorrect" factor, cross-loading, low correlations (< .30) on a theorized factor, or items being highly correlated (> .80) with other items. Separate exploratory factor analyses (using principal axis factoring; PAF) were conducted to analyse and refine the factor structures (Conway & Huffcutt, 2003). As a result, 11 items were deleted from the E scale, nine from the C scale, six each from the A and N scales, and five from the O scale. This left 45 items: 10 in each of the A, O, and C scales, 8 in the N scale, and 7 in the E scale. Coefficients alpha for the scales were: A .82, E .76, C .88, N .69, and O .78. As we wished to have at least 10 items in each scale for Phase 3, minor modifications were made to the questionnaire. Five items tapping aspects of N (moody, impulsive), and E (sociable), with a discrimination index of .20–.29 were rewritten for inclusion in Phase 3.

Phase 3: Exploratory Factor Analysis

The purpose of this phase was to pilot test the 50-item form of the instrument with a representative target group comprising students who had not previously participated, and to test the factor structure of the scales using exploratory factor analysis (EFA).

Participants

Participants were a convenience sample of 496 boys (44%) and girls aged 15-18 years ($M = 16.12$, $SD = 0.68$) in grades 11 (76%) and 12 (24%) at a co-educational mid-SES SE Queensland high school.

Materials and Procedure

The 50 personality items and demographic questions (gender, age, year level) were included in a paper-based questionnaire. Participants responded on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Sample items were: "I see myself as someone who…: "…will put others' needs before my own" (A), "…is very sociable" (E), "… manages their time well" (C), "…is easily annoyed" (N), and "…is very inquisitive" (O). Participants remained anonymous and were required to return a signed parental consent form before participating.

Results

EFA can be used to develop and refine an instrument and is used in hypothesis testing (Conway & Huffcutt, 2003). Hinkin (1998) argued for both strong theoretical justification and quantitative results for determining the number of factors to be extracted. Item loadings on latent factors serve to confirm theory-based expectations (Hinkin, 1998). Addressing the limitations of EFA, Tabachnick and Fidell (2013) reported that decisions about number of factors to be extracted and rotation method could be based on pragmatic rather than theoretical criteria. The commonly used eigenvalue >1 rule does not consistently give an accurate number of factors (tending to indicate too many factors), so cannot be relied on (e.g., Conway & Huffcutt, 2003; Gorsuch, 1997, Zwick & Velicer, 1986).

SPSS v22 was used for all Phase 3 analyses entering the correlation matrix in each case. The Kaiser–Meyer–Olkin measure of sampling adequacy (.89) and Bartlett's test of sphericity ($F = 9394.39$, $p < .001$) indicated the 50 items as suitable for an EFA (Tabachnick & Fidell, 2013). The sample size was deemed adequate to obtain an accurate solution (MacCallum, Widaman, Zhang, &

Hong, 1999). To produce meaningful distinctions between the factors by analysing only shared variance between variables, and to eliminate redundant or unclear items, a series of EFAs using PAF was conducted with oblimin rotation, as the factors were expected to be correlated. Before imposing a theoretically indicated factor structure, the Phase 3 data were allowed to vary freely within an EFA. While 13 factors were indicated by the eigenvalue >1 criterion, most being uninterpretable, the scree plot clearly indicated a 5-factor solution. Items were removed as described in Phase 2. The theoretical rationale for forcing a 5-factor solution upon these data was thereby reinforced by their inherent statistical properties. When five factors were selected for extraction, item loadings provided further support for the underlying theory (Conway & Huffcutt, 2003; Hinkin, 1998). However, researchers have deemed that further analyses are required to confirm the number of factors to extract, the most used being minimum average partial procedures, and parallel analysis (Courtney, 2013; Hayton, Allen, & Scarpello, 2004; Ledesma & Valero-Mora, 2007; O'Connor, 2000; Ruscio & Roche, 2012; Warne & Larsen, 2014). A parallel analysis was undertaken using the syntax documented by Hayton et al. (2004), modified for PAF with oblimin rotation. The parallel analysis results are in Table 1.

While Table 1 shows a clear rationale for extracting at least five factors, with the eigenvalues for factors 1-5 from the original analysis being clearly greater than those for the mean eigenvalues of comparable factors extracted from 50 runs of the randomized data, the case for extracting a sixth factor was problematic. The eigenvalue for the sixth extracted factor at 1.475 was just above that for the sixth factor extracted from the parallel analysis (1.444), and also its upper 95% confidence interval (1.450). However, when a 6-factor solution was imposed on the 50-item inventory, the sixth factor comprised three extraversion items, two with loadings of < .40, both of which cross-loaded with the main extraversion factor. All three items were deleted as part of the exploratory phase of the study to produce the 30-item version of the inventory. Nevertheless, the possibility that six factors might have been extracted from this sample could indicate that a facet-level analysis might have been possible with a larger initial item pool, at least for extraversion.

Items loading (> .40) on the factor for which they were written with no substantial cross-loadings (> .30) were retained. Twenty items either loading on an "incorrect" factor, or that had multiple loadings, or that loaded at < .40 on their designated factor, were deleted (Klein, 2000). Table 2 shows

items, factor loadings, summary data, and coefficients alpha. Table 3 shows factor inter-correlations (3a), and correlations between aggregated scaled variables (3b).

Phase 4: Confirmatory factor analyses

While there are no strict guidelines as to an acceptable percentage of the total item variance, as the variance accounted for in the phase 3 analysis was less than 60%, which is considered a minimal target (Hinkin, 1998; Holt et al., 2007), we tested the remaining 30 items on another independent sample (Anderson & Gerbing, 1991). The purpose was to conduct a further test of the factor structure using confirmatory factor analysis (CFA; Brown, 2006).

Participants

Participants were a convenience sample consisting of 405 boys (45%) and girls in grades 9 (31%), 10 (23%), 11 (21%), and 12 (25%), aged 13-18 years (M = 15.29, SD = 1.28), at one other co-educational middle SES high school in southeast Queensland..

Materials and Procedure

A paper-based questionnaire comprising the 30 items used in Phase 3 was administered to participants, who remained anonymous and provided a signed parental consent form.

Results

The six items listed under each of the FFM dimensions in Table 2 were entered into five separate CFAs using AMOS v23 with maximum likelihood estimation (Byrne, 2001). No data were missing. The item that fitted each model least well was excised to leave the final form of the FFM-APQ factors shown in Table 4(a-e). Removal of these items improved the fits of each model. For each of the five scales, Table 4 shows the final five items, and their respective standardized regression weights for the whole sample ($N = 405$), as well as separately for the male ($n = 184$), and female ($n = 221$) subsamples, with the corresponding coefficients alpha. Table 4 also shows fit statistics for the same three models for agreeableness (4a), extraversion (4b), conscientiousness (4c), neuroticism (4d), and openness (4e). In Table 4, all item regression weights were significantly associated with their respective factors ($p < .001$ for all items). For each scale, Table 4 also displays the results of metric and configural invariance analyses, which showed that the regressions weights for the male and female subsamples were equivalent for all five scales.

While all item distributions approximated to normal, as the scales were technically ordinal, the 25 items were further tested with a robust maximum likelihood estimator (MLR) in M*plus* v7.31 (Rhemtulla, Brosseau-Liard, & Savalei, 2012). A modified version of the program in Wang and Wang (2012) was used to derive fit statistics separately for the five factors, which are shown in Table 5 for the complete Phase 4 sample. M*plus* has a smaller range of fit statistics than AMOS, and those available are shown in Table 5. Almost identical fit statistics were produced from the M*plus* MLR estimator. The near-normal item distributions mean that Table 5 shows a similar array of fit statistics to the corresponding AMOS output. The rationale for the qualitative summary of the goodness-of-fit statistics shown in Table 6 is expanded in the Appendix to this paper.

The full CFA model is shown in Figure 1. The final standardized regression weights are almost identical with those shown in Table 4. Correlations between the latent variables, shown in Table 7, are consistent with the equivalent values for variables from the Phase 3 sample data. In Figure 1, the same fit statistics are shown as each set of items representing the five factors is added to the model in order of goodness-of-fit for each of the corresponding 1-factor congeneric models.

Discussion

Confirmatory factor analyses

While some researchers have found reasonable FFM model fits using CFA (Socha, Cooper, & McCord, 2010), due to inconclusive or poorly-fitting models, others have queried the utility of CFA for deriving the FFM dimensions (Borkenau & Ostendorf, 1990; Donnellan et al., 2006; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996; Parker, Bagby, & Summerfeldt, 1993; Vassend & Skrondal, 1997). Even EFA-developed scales may not be supported by CFA (Marsh et al., 2009, 2010; McCrae et al., 1996). However, it is possible that these findings apply mainly to longer (50+ items) versions of the FFM (Marsh, 2007; Marsh, Hau, Balla, & Grayson, 1998; Marsh, Hau, & Grayson 2005; Marsh et al., 2010), so that a 25-item inventory might be appropriate for CFA. It is prudent to select a variety of fit statistics when judging the goodness-of-fit of a model with respect to the null hypothesis (Wang & Wang, 2012). The fit statistics selected to test the models in the current study represented a range of criteria.

General summary on the CFA models

It would be surprising if the hypothesized model fits for all five personality dimensions for the male and female subsamples, and total samples were all good and similar to one another. One possibility is that the inherent variability in the five dimensions' scaled items, represented in the variety of fit statistics described, reflects at least in part a developmental or maturity process within the Phase 4 sample (Soto et al., 2008). This would be consistent with research revealing greater correlational stability of personality traits from mid to late adolescence (Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2009; Lüdtke, Trautwein, & Husemann, 2009; Roberts & DelVecchio, 2000).

While the 5-item scales individually show good fits for their respective factors, as the full CFA model is compiled scale-by-scale, the degree of fit diminishes with each additional set of items representing its corresponding latent factor. While the model comprising the Agreeableness and Extraversion items was a good fit for the data, when the Conscientiousness and Neuroticism items were included in the model, the fit statistics showed a more marginal fit for the data. Adding the Openness items resulted in an overall poor fit for the data, reflecting the more problematic construct validity of the items comprising this scale. It can be concluded that while the content and construct validities for the A, E, C, and N scales are acceptable, for the Openness scale these features are less robust. Comparing our findings with those of Lounsbury et al. (2003), who undertook a similar CFA, their final model fit, which was almost identical to ours, they variously described as "good" and "acceptable". However, while the model structure was similar, there were differences in the way in which the items had been derived, the respective sample sizes, sample age ranges, the software used for the analyses, the estimator used, the number of fit statistics presented, and the number of items representing each factor/scale. Lounsbury et al. used Emotional stability in place of Neuroticism, which resulted in positive correlations between all five scales.

Perhaps particularly within a school environment, friendship is a vital component in the life of an adolescent. Therefore, it might seem unsurprising that the best fit statistics are for the agreeableness dimension, this being the personality trait that is most highly valued – and the one against which one is most likely to be judged by ones' peers (Knack et al., 2013). Thus, of FFM traits, agreeableness is the one to gain salience, and therefore to crystallize first, within adolescent samples. This would be

consistent with Soto et al.'s (2008) finding that agreeableness showed relatively trivial gains in coherence between late childhood and adolescence, with no substantial gender differences. The social context of an adolescent's world is also likely to be strongly reflected in the extraversion dimension, as it is partly in behaviors expressed through this trait that initial and continuing peer acceptance is generated. Therefore, the scaled items representing extraversion, as indicted by the fit statistics for this trait, are those next most likely to provide a coherent structure for an adolescent sample. This would be consistent with Soto et al.'s (2008) finding that extraversion showed large gains in coherence between late childhood and adolescence, with no substantial gender differences.

Like agreeableness, conscientiousness is subject to socialization processes, particularly in home and school environments. For example, children are taught to conform to behaviors represented by conscientiousness, including setting personal goals, and being well organized, as well as staying focused in their school work, and in other activities. Thus, it could be postulated that while agreeableness and extraversion can be identified relatively early in a young person's life, conscientiousness is more subject to socialization processes over a slightly later and longer period. This could help to explain the decrease in fit of the conscientiousness model based on the scales derived in this study, compared with those representing agreeableness and extraversion. This would be consistent with Soto et al.'s (2008) finding that conscientiousness showed trivial gains in coherence between late childhood and adolescence, with no substantial gender differences.

Being assessed less by behavioral markers, the trajectory for neuroticism (and its converse, emotional stability) is perhaps more aligned with maturational processes, which would be consistent with the negative (or near zero) correlations typically found between this trait and the other four. It would also be consistent with Soto et al.'s (2008) finding that N showed moderate gains in coherence between late childhood and adolescence, with no gender differences. Given the relative disparities in the fit statistics for the FFM traits in the Phase 4 sample, a possible rationale for these diverse 1-factor model fits may be that our study reflected the adolescent developmental stage at which each of the FFM dimensions crystallize as coherent personality traits. From these sample-specific analyses developmental order appears to be: agreeableness, extraversion, conscientiousness, neuroticism, and openness.

The more problematic nature of openness compared with the other four dimensions, even for adult samples, has been well-documented (e.g., Sneed et al., 2002). In contrast with the socializing context for the other four dimensions, an individual must achieve a certain amount of life experience before the features characterizing openness are likely to become evident. For example, before it is possible to know about "lots of things", it is necessary to have been exposed to a corresponding amount of information. While being inquisitive or curious is a natural feature of a young person's development, some behaviors representing these concepts may need to be stimulated by appropriate environments, perhaps over an extended period. A similar case could be made for having lots of ideas and being inventive. Thus, of the FFM traits, openness is likely to be the trait that develops most slowly and over the longest period. For some adolescents at least, it could remain a relatively undifferentiated concept, which would be reflected in the fit statistics for this dimension. It would also be consistent with the finding that between late childhood and adolescence, openness showed only moderate gains in coherence, with no substantial gender differences (Soto et al., 2008). These authors also hinted at multiple processes operating to differentiate and crystallize the FFM traits during adolescence. These suggestions are consistent with our reflections on the relative model fits for the five dimensions. Table 8 summarizes our suggested theoretical drivers for development of the FFM during adolescence.

Limitations, strengths, and future research

The limitation of accessing girls only in phases 1 and 2 of this study was partially offset by the mixed gender samples available for psychometric development of the inventory in phases 3 and 4 – although all were convenience samples. As Soto et al. (2008) found no coherence or differentiation gender disparities in a large diverse adolescent sample, it is likely that the items used in this study were comparable across genders, which would be supported by the Phase 4 pattern of fit statistics, and by the invariance analyses. Potential sources of error included random variation in responding, participants' mood at the time of testing, and miscellaneous effects (e.g., testing environment, morningness-eveningness $\times$ time of day interactions).

While they are theoretically orthogonal, inflated correlations between the FFM traits are a well-documented feature (McCrae et al., 1996; Marsh et al., 2010; Parker et al., 1993). In an attempt to minimize this effect, strict criteria as would be consistent with deriving a 25-item inventory were

applied in Phase 3. This included checking the initial EFA solution using different extraction and rotation methods (Conway & Huffcutt, 2003) to ensure that nearly all 30 items remaining after Phase 3 had extremely low cross-loadings on other factors. Removing a further item from each factor in Phase 4 defined each factor even more stringently, which would have assisted the CFA derivation of at least acceptable fits for most of the 1-factor models.

Development of any scaled version of the FFM that is less than the 240-item NEO-PI-R (Costa & McCrae, 1992b), or the IPIP-NEO-120 (Johnson, 2014), must trade the advantages of brevity for ease of application and analysis, against representation of the facet-level components underlying the NEO-PI-R's theoretical construction. In adolescent samples, adequately reflecting the sophistication of facet-level personality characteristics might be problematic. Therefore, in developing the FFM-APQ we did not seek to explicitly represent any FFM facet-level characteristics.

While we obtained evidence for construct and discriminant validity, this study did not seek evidence for convergent validity. This limitation provides opportunities for researchers to compare the FFM-APQ with other adolescent personality measures. As its development was based on Australian high school students who were predominantly Caucasian, future studies could test the applicability of the FFM-APQ on samples from more diverse populations.

Strengths of the study include that EFAs and CFAs were carried out at item level, allowing for item discriminability, which is essential for test development. The large initial item pool permitted the staged refinement to produce a set of scales for further testing on large, independent samples. Low question complexity and high accessibility minimized construct-irrelevant variance (AERA, 2014).

A brief FFM measure with evidence for validity in non-clinical adolescent samples, as described here, could assist in the conceptual task of identifying the development of coherence and differentiation (Soto et al., 2008). It could also be used to measure changes in the FFM five dimensions across this theoretically important part of the lifespan (Branje, Van Lieshout, & Gerris, 2007; Canals, Vigil-Colet, Chico, & Martí-Hennenberg, 2005; De Fruyt et al., 2006; Kawamoto & Endo, 2015; Klimstra et al., 2009; McCrae et al., 2002; Pullmann et al., 2006; Roberts & DelVecchio, 2000; Roberts, Walton, & Viechtbauer, 2006; Soto, John, Gosling, & Potter, 2011; Van den Akker,

Deković, Asscher, & Prinzie, 2014). Other potential applications include assessments of individual awareness, psychological health, and personal growth.

Future research could usefully focus on gathering further evidence for test usage validity, including convergent/concurrent (correlations with other personality measures), and criterion/predictive (e.g., academic performance) evidence for the final 25-item test. Test-retest coefficients and norm referencing would also be desirable (AERA, 2014). The test could be either paper-based, or delivered online, or administered by computer. The last two options can incorporate item randomization.

Conclusion

Adopting a best-practice approach to item development, we established a link between item development and the theoretical domain, and conducted focus groups with a sample drawn from the target population to describe aspects of behavior to explore aspects of each personality dimension (Hinkin, 1998). A brief personality inventory with evidence for its validity will be helpful when researchers include other construct measures in surveys. As personality plays an important role in adolescent development, the FFM-APQ could be useful in career guidance for non-clinical adolescent samples, and in identifying personality features impacting social behaviors, and well-being.

We have reported the 4-phase development of a 25-item adolescent personality inventory based on the FFM. Results from the final phase revealed that items comprising the five personality factors established evidence for validity through focus group derivative extraction, expert review, item analysis, and EFA, and by CFA, a coherent factor structure. Evidence for validity of the factor structure was sustained for both male and female subsamples for all five factors, with agreeableness and extraversion being the most robust dimensions. Whilst agreeableness has often been found to be the most problematic factor (Caruso, 2000), this finding might reflect the important role that agreeableness plays in developing and maintaining friendships during adolescence (Knack et al., 2013). This study provides initial psychometric evidence for the validity of the FFM-APQ, designed specifically for adolescents aged up to 18 years.

References

American Educational Research Association. (2014). *Standards for educational and psychological testing* (7th edn.). Washington, DC: AERA.

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, *76*(5), 732–740. doi:10.1037/0021-9010.76.5.732

Asendorpf, J. B., & van Aken, M. A. G. (2003). Personality–relationship transaction in adolescence: Core versus surface personality characteristics. *Journal of Personality, 71*(4), 629–666. doi.10.1111/1467-6494.7104005

Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the Mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment, 95*(1), 74–84. doi:10.1080/00223891.2012.700466

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences, 11*(5), 515–524. doi:10.1016/0191-8869(90)90065-Y

Branje, S. J. T., Van Lieshout, C. F. M., & Gerris, J. R. M. (2007). Big Five personality development in adolescence and adulthood. *European Journal of Personality, 21*(1), 45–62. doi:10.1002/per.596

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.

Buhrmester, D. P. (1996). Need fulfillment, interpersonal competence, and the developmental contexts of early adolescent friendship. In W. M. Bukowski, A. F. Newcomb, & W. W. Hartup (Eds.), *The company they keep: Friendship in childhood and adolescence* (pp. 158–185). New York, NY: Cambridge University Press.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Canals, J., Vigil-Colet, A., Chico, E., & Martí-Hennenberg, C. (2005). Personality changes during adolescence: The role of gender and pubertal development. *Personality and Individual Differences, 39*(1), 179–188. doi:10.1016/j.paid.2004.12.012

Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and*

   Psychological Measurement, 60(2), 236–254. doi:10.1177/00131640021970484

Cattell, R. B., & Cattell, M. D. (1975). *Handbook for the Junior and Senior High School Personality*

   Questionnaire. Champaign, IL: Institute for Personality and Ability Testing.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development.

   Psychological Assessment, 7(3), 309–319. doi:10.1037/1040-3590.7.3.309

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis

   practices in organizational research. *Organizational Research Methods, 6*(2), 147–168.

   doi:10.1177/1094428103251541

Costa, P. T., Jr., & McCrae, R. R. (1992a). *Revised NEO Personality Inventory (NEO–PI–R) and*

   NEO Five-Factor Inventory (NEO–FFI) professional manual. Odessa, FL: Psychological

   Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992b). *NEO-PI-R professional manual*. Odessa, FL:

   Psychological Assessment Resources.

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-

   Menu v2.0 to make more judicious estimations. *Practical Assessment, Research and Evaluation,*

   18(8). Available online: http://pareonline.net/getvn.asp?v=18&n=8

Deal, J. E., Halverson, C. F., Jr., Martin, R. P., Victor, J., & Baker, S. (2007). The Inventory

   of Children's Individual Differences: Development and validation of a short version.

   Journal of Personality Assessment, 89(2), 162–166, doi:10.1080/00223890701468550

De Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Merivielde, I. (2006).

   Five types of personality continuity in childhood and adolescence. *Journal of Personality and*

   Social Psychology, *91*(3), 538–552. doi:10.1037/0022-3514.91.3.538

De Fruyt, F., De Bolle, M., McCrae, R. R., Terraciano, A., & Costa, P. T. (2009). Assessing the

   universal structure of personality in early adolescence: The NEO-PI-R and NEO-PI-3 in 24

   cultures. *Assessment, 16*(3), 301–311. doi:10.1177/1073191109333760

De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J. P. (2000). Assessing adolescents'

    personality with the NEO PI-R. *Assessment, 7*(4), 329–345. doi:10.1177/107319110000700403

Donnellan, M. B., Conger, R. D. & Burzette, R. G. (2007). Personality development from late

    adolescence to young adulthood: Differential stabilityk, normative maturity, and evidence for

    maturity-stability hypothesis. *Journal of Personality, 75*(2), 237–263. doi: 10.1111/j.1467-

    6494.2007.00438.x

Donnellan, M. B., Oswald, F. L., Bairds, B. M., & Lucas, R. E. (2006). The Mini-IPIP Scales: Tiny-

    yet-effective measure of the Big Five factors of personality. *Psychological Assessment, 18*(2),

    192–203. doi:10.1037/1040-3590.18.2.192

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure.

    Journal of Personality and Social Psychology, 59(6), 1216–1229. doi:10.1037//0022-

    3514.59.6.1216

Goldberg, L. R. (2001). Analyses of Digman's child-personality data: Derivation of Big-Five factors

    scores from each of six samples. *Journal of Personality, 69*(5), 709–743. doi:10.1111/1467-

    6494.695161

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality

    Assessment*, *68*(3), 532–560. doi:10.1207/s15327752jpa6803_5

Gosling, S. D., Rentfrow, P. J., & Swan, W. B., Jr. (2003). A very brief measure of the Big Five

    personality domains. Journal of Research in Personality, 37(6), 504–528. doi:10.1016/S0092-

    6566(03)00046-1

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor

    analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205.

    doi:10.1177/1094428104263675

Hinkin, T. R. (1998). A brief tutorial on the development of measures for the use in survey

    questionnaires. *Organizational Research Methods, 1*(1), 104–121.

    doi:10.1177/109442819800100106

Holt, D. T., Armenakis, A. A., Feild, H. S., & Harris, S. G. (2007). Readiness for organizational

    change the systematic development of a scale. *The Journal of Applied Behavioral Science*, *43*(2),

    232–255. doi:10.1177/0021886306295295

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:

    Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary*

    Journal, 6(1), 1–55. doi:10.1080/10705519909540118

Jensen-Campbell, L. A. & Malcolm, K. T. (2007). The importance of conscientiousness in adolescent

    interpersonal relationships. *Personality and Social Psychology Bulletin, 33*(3), 368–383.

    doi:10.1177/0146167206296104

John, O. P., Caspi, A., Robins, R. W., Moffitt, T. E., & Stouthamer-Loeber, M. (1994). The "Little

    Five": Exploring the nomological network of the five-factor model of personality in adolescent

    boys. *Child Development, 65*(1), 160–178. doi:10.2307/1131373

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public

    domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*,

    78–89. doi:10.1016/j.jrp.2014.05.003

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrack, M. R. (1999). The big five personality traits,

    general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621–

    652. doi:10.1111/j.1744-6570.1999.tb00174.x

Kawamoto, T., & Endo, T. (2015). Personality change in adolescence: Results from a Japanese

    sample. *Journal of Research in Personality, 57*(1), 32–42. doi:10.1016/j.jrp.2015.03.002

Klein, P. (2000). *The handbook of psychological testing* (2nd ed.). London, UK: Routledge.

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*(4),

    898–912. doi:10.1037/a0014746

Kline, R. B. (2005). *Principles and practice of structural equation modelling* (2nd. ed.). New York,

    NY: Guilford.

Knack, J. M., Jacquot, C., Jensen-Campbell, L. A., & Malcolm, K. T. (2013). Importance of having agreeable friends in adolescence (especially when you are not). *Journal of Applied Social Psychology, 43*(12), 2401–2413. doi:10.1111/jasp.12188

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation, 12*(2), 1–11.

Limont, W., Dreszer-Drogorob, J., Bedynska, S., Sliwinska, K., & Jastrzebska, D. (2014). 'Old wine in new bottles'? Relationships between overexcitabilities, the Big Five personality traits and giftedness in adolescents. *Personality and Individual Differences, 69*, 199–204. doi:10.1016/j.paid.2014.06.003

Lounsbury, J. W., Tatum, H., Gibson, L. W., Park, S. H., Sundstrom, E. D., Hamrick, F. L., & Wilburn, D. (2003). The development of a Big Five adolescent personality inventory. *Journal of Psychoeducational Assessment*, *21*(2), 111–133. doi:10.1177/073428290302100201

Lüdtke, O., Trautwein, U., & Husemann, N. (2009). Goal and personality trait development in a transitional period: Assessing change and stability in personality development. *Personality and Social Psychology Bulletin, 35*(4), 428–441. doi:10.1177/0146167208329215

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. doi:10.1037/1082-989X.4.1.84

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modelling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774–798). Hoboken, NJ: Wiley.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181–220. doi:10.1207/s15327906mbr3302_1

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modelling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Psychometrics: A festschrift to Roderick P. McDonald* (pp. 225–340). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341. doi:10.1207/s15328007sem1103_2

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modelling. *Psychological Assessment, 22*(3), 471–491. doi:10.1037/a0019227

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modelling, integrating CFA and EFA: Applications to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 439–476. doi:10.1080/10705510903008220

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81–90. doi:10.1037/0022-3514.52.1.81

McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2015). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of Personality Assessment, 84*(3), 261–270. doi:10.1207/s15327752jpa8403_05

McCrae, R. R., Costa, P. T., Jr. , Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology, 83*(6), 1456–1468. doi:10.1037//0022-3514.83.6.1456

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., Zonderman, A. B., Costa, P. T. Jr., Bond, M. H., & Paunonen, S. (1996). Evaluating the replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology, 70*(3), 552–566. doi:10.1037/0022-3514.70.3.552

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396–402. doi:10.3758/BF03200807

Parker, J. D. A., Bagby, R. M., & Summerfeldt, L. J. (1993). Confirmatory factor analysis of the Revised Neo-personality Inventory. *Personality and Individual Differences, 15*(4), 463–466. doi:10.1016/0191-8869(93)90074-D

Pullmann, H., Raudsepp, L., & Allik, J. (2006). Stability and change in adolescents' personality: A longitudinal study. *European Journal of Personality*, *20*(6), 447–459. doi:10.1002/per.611

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203–212. doi:10.1016/j.jrp.2006.02.001

Randler, C. (2008). Morningness–eveningness, sleep–wake variables and big five personality factors. *Personality and Individual Differences, 45*(2), 191–196. doi:10.1016/j.paid.2008.03.007

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. doi:10.1037/a0029315

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantititative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3–25. doi:10.1037/0033-2909.126.1.3

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*(1), 1–25. doi:10.1037/033-2909.132.1.1

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of a known factorial structure. *Psychological Assessment, 24*(2), 282–292. doi:10.1037/a0025697

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and Big Five personality traits: A social network approach. *Journal of Personality, 78*(2), 509–538. doi:10.1111/j.1467-6494.2010.00625.x

Shiner, R. L. (1998). How shall we speak of children's personalities in middle childhood?

   Psychological Bulletin, 124(3), 308–332. doi:10.1037/0033-2909.124.3.308

Shiner, R. L. (2000). Linking childhood personality with adaptation evidence for continuity and

   change across time into late adolescence. *Journal of Personality and Social Psychology, 78*(2),

   310–325. doi:10.1037/0022-3514.78.2.310

Shiner, R. L., & Caspi, A. (2003). Personality differences in childhood and adolescence:

   Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry,*

   44(1), 2–32. doi:10.1111/1469-7610.00101

Sneed, C. D., Gullone, E., & Moore, S. (2002). Reliability and factor structure of the NEO-Five-

   Factor Inventory for Australian adolescents. *Behaviour Change, 19*(2), 121–126.

   doi:10.1375/bech.19.2.121

Socha, A., Cooper, C. A., & McCord, D. M. (2010). Confirmatory factor analysis of the M5-50: An

   implementation of the International Personality Item Pool item set. *Psychological Assessment*,

   22(1), 43–49. doi:10.1037/a0017371

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big

   five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20.

   Journal of Personality and Social Psychology, *94*(4), 718–737. doi:10.1037/0022-3514.94.4.718

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from

   10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and*

   Social Psychology, 100(2), 718–737. doi:10.1037/a0021717

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Allyn

   & Bacon.

Van den Akker, A. L., Deković, M., Asscher, J., & Prinzie, P. (2014). Mean-level personality

   development across childhood and adolescence: A temporary defiance of the maturity principle

   and bidirectional associations with parenting. *Journal of Personality and Social Psychology,*

   107(4), 736–750. doi:10.1037/a0037248

Vassend, O., & Skrondal, A. (1997). Validation of the NEO Personality Inventory and the five-factor

   model: Can findings from exploratory and confirmatory factor analysis be reconciled? *European*

Journal of Personality, 11(2), 147–166. doi:10.1002/(SICI)1099-0984(199706)11:2<147::AID-PER278>3.0.CO;2-E

Wang, J., & Wang, X. (2012). *Structural equation modelling applications using Mplus*. Chichester, UK: Wiley.

Warne, R. G., & Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychological Test and Assessment Modeling, 56*(1), 104–123.

Yatabe, J. (1975). Yatabe-Guilford personality inventory. In T. Okado (Ed.), *Psychological assessment: Basic foundations* (pp. 269–281). Tokyo, Japan: Kakinouchi Press.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. doi:10.1037/0033-2909.99.3.432

Table 1

Parallel Analysis Results Compared with Original Factor Analysis (N = 496)

Factor	Eigenvalues from initial PAF analysis	Mean eigenvalues from 50 parallel PAF analyses	Upper 95% CI
1	10.440	1.680	1.692
2	3.437	1.616	1.623
3	3.165	1.560	1.567
4	2.361	1.516	1.522
5	1.917	1.479	1.485
6	1.475	1.444	1.450
7	1.340	1.410	1.414
8	1.249	1.379	1.384
9	1.219	1.350	1.355
10	1.128	1.320	1.325

Table 2

Items, Factor Loadings, and Summary Data for the FFM-APQ 30-item Scale in Phase 3 (N = 496)

Factor	1	2	3	4	5
Extraversion items (M 21.59, SD 3.45, α .77)					
Is very sociable	**.68**	.01	-.01	.16	.11
Is very outgoing	**.62**	.00	.01	.05	-.15
Always has something to talk about	**.60**	.04	-.00	.13	.06
Will always be the one to say "Let's do something"	**.55**	-.05	-.04	.12	-.07
Has more friends than other people do	**.50**	.03	.09	-.11	-.03
Is active and energetic	**.46**	-.06	.04	.06	-.09
Neuroticism items (M 17.42, SD 4.15, α .78)					
Is easily annoyed	.16	**.70**	-.05	-.19	.02
Gets snappy over little things	.07	**.68**	.03	-.21	-.00
Can get annoyed at the smallest of things	.09	**.66**	-.08	.01	-.01
Is quite anxious a lot of the time	-.17	**.61**	.10	.08	-.08
Is moody a lot of the time	.02	**.55**	-.03	-.10	-.05
Always gets nervous about things	-.13	**.46**	.04	.22	.06
Conscientiousness items (M 19.42, SD 4.29, α .81)					
Manages their time well	.07	-.04	**.78**	.00	.09
Organises themselves well	.15	.04	**.78**	.03	.12
Is not easily distracted	-.13	.01	**.59**	-.02	-.02
Sets goals for myself	-.02	.04	**.57**	.10	-.11
Can stay focused on a task	-.01	-.04	**.54**	.09	-.06
Always has a plan	.07	.04	**.54**	-.03	-.12
Agreeableness items (M 22.69, SD 3.47, α .84)					
Will put others' needs before my own	.08	.00	.03	**.72**	.05
Likes to help others	.08	-.12	-.02	**.63**	-.15
Is considerate of others	.09	-.03	.11	**.63**	-.02
Always thinks about other people's feelings	-.03	-.04	.10	**.63**	.02
Gives things to others	.15	.04	-.01	**.62**	.00
Understands where others are coming from	.09	-.05	.02	**.50**	-.19
Openness items (M 20.78, SD 3.11, α .70)					
Is very inquisitive	-.07	-.01	.05	.04	**-.57**
Is inventive	.05	-.04	.05	-.07	**-.56**
Is rather curious	-.04	.16	-.17	.16	**-.56**
Knows about a lot of things	.00	-.14	.17	.03	**-.43**
Has lots of ideas	.25	-.05	.07	.04	**-.43**
Is always asking questions about things	.12	.11	.04	-.03	**-.41**

Note. Factor loadings > .40 are in bold.

Table 3

a) Inter-factor correlations

Factor	N	C	A	O
Extraversion	-.07	.18	.27	-.30
Neuroticism		-.09	-.10	-.04
Conscientiousness			.26	-.28
Agreeableness				-.26

Table 4

b) Correlations between Aggregated Scaled Variables (N = 496)

Scale	E	C	N	O
Agreeableness	$.39^{**}$	$.33^{**}$	$-.19^{**}$	$.32^{**}$
Extraversion		$.24^{**}$	-.04	$.32^{**}$
Conscientiousness			$-.09^{*}$	$.30^{**}$
Neuroticism				.00

*** p < .001, * p < .05*

Table 5

Personality Items, Standardized Regression Weights, Cronbach's Alphas, CFA Model Fit Statistics,
Gender Invariance Test Fit Statistics for the Complete Phase 4 Sample (N = 405), and separately for
Males (n = 184), and Females (n = 221)

	Standardized regression weights		
a) Agreeableness items	*Whole sample*	*Males*	*Females*
Always thinks about other people's feelings	.73	.74	.69
Will put others' needs before my own	.72	.70	.72
Likes to help others	.70	.69	.71
Is considerate of others	.67	.71	.64
Gives things to others	.58	.65	.48
Cronbach's Alpha	*.81*	*.82*	*.78*
CFA model fit statistics *p*	.310	.781	.404
Normed χ^2	1.193	0.494	1.020
SRMR	.016	.011	.012
AGFI	.983	.984	.973
CFI	.998	1.000	1.000
RMSEA	.022	.000	.009
RMSEA 90% CIs	.000–.075	.000–.068	.000–.095
Invariance tests: Fit statistics for gender* *Metric*	*p* .0213, χ^2 20.8, df 15		
Configural	*p* .671, Normed χ^2 0.757, SRMR .015, AGFI .978, CFI 1.000, RMSEA .000 (.000–.043)		
b) Extraversion items	*Whole sample*	*Males*	*Females*
Is very sociable	.84	.89	.80
Is very outgoing	.74	.74	.72
Always has something to talk about	.66	.65	.65
Will always be the one to say "Let's do something"	.59	.66	.53
Has more friends than other people do	.51	.53	.54
Cronbach's Alpha	*.80*	*.82*	*.78*
CFA model fit statistics *p*	.088	.097	.024
Normed χ^2	1.918	1.609	2.057
SRMR	.022	.027	.031
AGFI	.971	.953	.943
CFI	.992	.990	.981
RMSEA	.048	.039	.049
RMSEA 90% CIs	.000–.093	.000–.072	.017–.079
Invariance tests: Fit statistics for gender* *Metric*	*p* .5460, χ^2 20.110, df 15		
Configural	*p* .097, Normed χ^2 1.609, SRMR .026, AGFI .953, CFI .990, RMSEA .039 (.000–.072)		
c) Conscientiousness items	*Whole sample*	*Males*	*Females*
Organises themselves well	.84	.86	.85
Manages their time well	.80	.87	.74
Can stay focused on a task	.52	.54	.49
Always has a plan	.47	.39	.55
Sets goals for myself	.40	.42	.38
Cronbach's Alpha	*.74*	*.76*	*.74*
CFA model fit statistics *p*	.007	.236	.038
Normed χ^2	3.198	1.360	2.356
SRMR	.016	.034	.041
AGFI	.953	.955	.940
CFI	.977	.993	.973
RMSEA	.074	.044	.079

	Whole sample	Males	Females
RMSEA 90% CIs	.035–.116	.000–.119	.017–.138
Invariance tests: Fit statistics for gender* Metric	*p* .0992, χ^2 27.853, df 15		
Configural	*p* .046, Normed χ^2 1.858, SRMR .037, AGFI .947, CFI .983, RMSEA .046 (.006–.078)		

d) Neuroticism items

	Whole sample	Males	Females
Is easily annoyed	.82	.88	.77
Gets snappy over little things	.82	.74	.89
Can get annoyed at the smallest of things	.79	.77	.80
Is moody a lot of the time	.57	.50	.61
Is quite anxious a lot of the time	.34	.31	.37
Cronbach's Alpha	*.81*	*.78*	*.82*
CFA model fit statistics *p*	.003	.022	.271
Normed χ^2	3.627	2.634	1.275
SRMR	.037	.048	.030
AGFI	.950	.924	.966
CFI	.982	.971	.997
RMSEA	.081	.095	.035
RMSEA 90% CIs	.043–.122	.033–.158	.000–.105
Invariance tests: Fit statistics for gender* Metric	*p* .0379, χ^2 31.338, df 15		
Configural	*p* .034, Normed χ^2 1.955, SRMR .047, AGFI .947, CFI .987, RMSEA .049 (.013–.081)		

e) Openness items

	Whole sample	Males	Females
Has lots of ideas	.70	.76	.65
Is inventive	.63	.64	.59
Knows about a lot of things	.61	.74	.50
Is very inquisitive	.59	.67	.52
Is rather curious	.40	.39	.42
Cronbach's Alpha	*.73*	*.78*	*.67*
CFA model fit statistics *p*	.002	.012	.033
Normed χ^2	3.879	2.946	2.433
SRMR	.030	.035	.033
AGFI	.940	.902	.935
CFI	.961	.960	.948
RMSEA	.084	.103	.081
RMSEA 90% CIs	.047–.126	.044–.166	.021–.140
Invariance tests: Fit statistics for gender* Metric	*p* .1377, χ^2 35.252, df 15		
Configural	*p* .003, Normed χ^2 2.690, SRMR .042, AGFI .920, CFI .955, RMSEA .065 (.036–.095)		

*Regression weights for male and female subsamples deemed equivalent

Table 6

Fit Statistics for the Phase 4 Sample (N = 405) from the Mplus MLR Estimator

Trait	Normed χ^2	p	RMSEA	RMSEA 90% CIs	CFI	TLI*	SRMR
Agreeableness	0.835	.525	.000	.000–.063	1.000	1.004	.014
Extraversion	1.516	.181	.036	.000–.084	.994	.988	.021
Conscientiousness	2.529	.027	.061	.019–.105	.978	.956	.032
Neuroticism	3.104	.008	.072	.033–.114	.977	.955	.030
Openness	2.835	.015	.067	.027–.110	.965	.929	.033

* Tucker-Lewis Index – like the CFI, an incremental fit index that compares the lack of fit of a specified model with the lack of fit of the null model (Wang & Wang, 2012).

Table 7

Qualitative Summary of Goodness-of-fit Statistics for the Final FFM-APQ Scales

Scale	*Agreeableness*	*Extraversion*	*Conscientiousness*	*Neuroticism*	*Openness*
p value	Good (T, M, F)	Good (T, M, F)	Good (M, F)	Good (M, F)	Good (M, F)
Normed χ^2	Good (T, M, F)	Good (T, M, F)	Good (M, F); Acceptable (T)	Good (M, F); Acceptable (T)	Good (M, F); Acceptable (T)
SRMR	Good (T, M, F)	Good (T, M, F)	Good (T, M, F)	Good (T, M, F)	Good (T, M, F)
AGFI	Good (T, M, F)	Good (T, F); Acceptable (M)	Good (T, M); Acceptable (F)	Good (T, M); Acceptable (F)	Acceptable (T, M, F)
CFI	Good (T, M, F)	Good (T, M, F)	Good (T, M, F)	Good (T, M, F)	Good (T, M, F)
RMSEA	Good (T, M, F)	Good (T, M, F)	Good (M); Acceptable (T, F)	Good (F); Acceptable (T); Mediocre (M)	Mediocre (T, F); Poor (M)

T = Total sample; M = Male subsample; F = Female subsample

Table 8

Correlations between Latent Factors in the CFA Model (N = 405)

Scale	E	C	N	O
Agreeableness	.44	.36	-.10	.41
Extraversion		.32	-.09	.49
Conscientiousness			-.04	.38
Neuroticism				.04

Table 9

Main Theoretical Drivers Proposed for Developing the FFM Personality Traits during Adolescence

Trait	Main Driver
Agreeableness	Social (e.g., initial friendships, peer acceptance)
Extraversion	Social (e.g., proactive seeking of new friendships)
Conscientiousness	Socialization (e.g., rewarding desired behaviors)
Neuroticism	Maturity (e.g., enhancing emotional stability, reducing moodiness)
Openness	Experience (e.g., exposure to multiple life opportunities)

APPENDIX: Expanded summary of rationale for fit statistics for the Phase 4 sample

p **value.** The probability that a model differs significantly from the null model is assessed by the chi-square statistic. Non-significant *p* values indicate a good fit for the data to the hypothesized model. On a 1% cut-off criterion (as 15 models were tested), good fits can be seen for A and E (total sample, male and female subsamples), while good fits for the separate male and female subsamples were obtained for O, N, and C. Because the *p* value is sensitive to sample size, while the fits for the total samples for these three dimensions exceeded the 1% level, all were within the .1% level, indicating acceptable model fits. The MLR estimator output broadly confirmed these results.

Normed chi-square. Because the chi-square statistic is sensitive to sample size, normed chi-square, which divides the raw chi-square value by the sample degrees of freedom, is a standard goodness-of-fit measure. With no consensus on the acceptable range for this statistic, limits of 2.00, or 3.00 have been recommended (Kline, 2005). On this criterion, the fits for all separate male and female subsamples for all five dimensions were good. However, while the fits for the total samples for A and E were good fits, those for N, O, and C fell just outside the 3.00 limit. On this criterion, all dimensions might be considered to be at least acceptable fits to the sampled data. The MLR estimator produced similar results, with only the value for N being just over 3.00.

Standardized root mean square residual (SRMR). This measure represents the average discrepancy between the observed sample and the hypothesized correlation matrices (Byrne, 2001), being a measure of the mean absolute correlation residual, with values less than .10 considered acceptable (Kline, 2005). On this criterion, all dimensions indicated good fit for male, female, and total samples, which were confirmed for the total sample by the MLR estimator results.

Adjusted goodness-of-fit index (AGFI). This absolute fit index, which adjusts for the degrees of freedom in the specified model, compares the hypothesized model with no model, and measures the relative amount of variance and covariance in the sample covariance matrix that is jointly explained by the population covariance matrix (Byrne, 2001). Values above .90 represent acceptable fit, while values exceeding .95 are considered to represent good model fit. On this criterion, A showed a good fit for the male, female, and total samples. N showed good fit for the female and total samples, and

acceptable fit for the male subsample. E and C showed good fit for the male and total samples, and acceptable fit for the female subsample, while O showed acceptable fit for the three samples.

Comparative fit index (CFI). Taking account of sample size and assuming complete independence between the variables, the CFI is derived by comparing the hypothesized model with the independence (baseline or null) model, which assumes zero covariances among the observed variables, being a measure of complete covariance in the data (Byrne, 2001; Socha et al., 2010; Wang & Wang, 2012). A generally supported cut-off value of .95 has been recommended (Hu & Bentler, 1999). On this index, all samples for the five dimensions were good fits for the sampled data, which were confirmed by the MLR estimator output.

Root mean square error of approximation (RMSEA). This index, which is sensitive to the number of estimated model parameters (model complexity) and degrees of freedom, estimates the extent to which the hypothesized model would fit the deemed population covariance matrix (Byrne, 2001), estimating the extent to which population covariances can be reproduced from the model parameters (Socha et al., 2010). The approximation error refers to the lack of fit of the specified model to the population (Wang & Wang, 2012). Values less than .05 indicate good fit, values up to .08 represent reasonable fit, values between .08 and .10 indicate mediocre fit, while values exceeding .10 indicate poor fit (Byrne, 2001; Marsh, Hau, & Wen, 2004; Wang & Wang, 2012). Ideally the lower 90% confidence interval (CI) would be zero, indicating the possibility of perfect fit, while the upper CI would ideally be within the acceptable range. The narrower the CI range, the greater the confidence that the true value of the RMSEA is close to the estimated value. Being sensitive to sample size and number of estimated parameters, it is difficult to obtain narrow confidence intervals with small samples and large numbers of estimated parameters. In these analyses, the sample sizes – particularly for the separate male and female subsamples – were not large, making it difficult to obtain good fits from RMSEA estimates.

Bearing these criteria in mind, the RMSEA values for A were good for the male, female, and total samples, with all three lower CIs being zero, and only the upper CI for females being above the .08 acceptable limit, but still within the "mediocre" limit. The MLR estimator confirmed the result for the total sample. RMSEA values for E (males, females, total sample) also represented good fits to the

data, all being under the .05 cut-off value. Lower CIs for the male and total samples were zero, while that for the female subsample was also low. While the upper CIs for the male and female subsamples were within the .08 acceptable cut-off, that for the total sample fell within the mediocre cut-off. The MLR estimator output confirmed the result for the total sample, although the upper CI was .084.

RMSEA values and CIs for the other three dimensions, while mixed were not as good. While the female subsample RMSEA indicated a good fit for N, and that for the total sample a just about acceptable fit, the male sample showed only a mediocre fit. While all the lower-bound CIs were under the good fit cut-off, all the upper CIs were in the poor fit range – findings that were confirmed by the MLR estimator output. The picture was similar for C, which indicated a good fit for the male subsample, and acceptable fits for the female subsample and the total sample. However, while the lower bound CIs were all within the good fit range, all the upper bound CIs were in the poor fit range. The finding for the total sample was confirmed by the MLR estimator output. O showed the poorest RMSEA fit, with the female subsample and the total sample showing mediocre fits, and the male subsample showing a poor fit. While the lower bound CIs were all within the good fit range, all upper bound CIs were in the poor fit range, this result again being broadly confirmed by the MLR output.
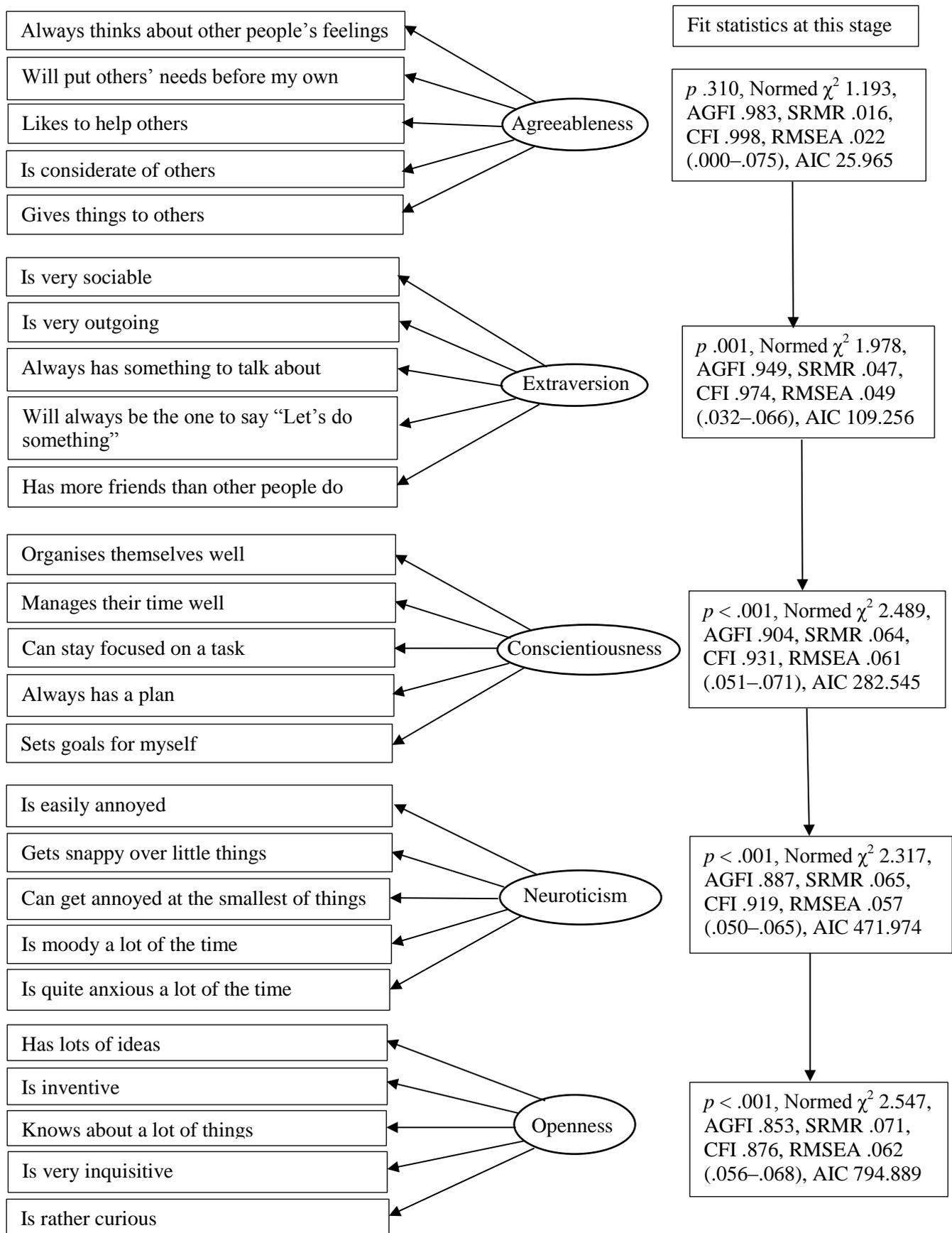
Figure 1. Full CFA model showing change in goodness-of-fit statistics as items representing each personality dimension are added.