

**ADOLESCENT PERSONALITY:
DEVELOPMENT, INTERRATER
AGREEMENT, AND RELATION
TO ACADEMIC ACHIEVEMENT**

KAIA LAIDRA



TARTU UNIVERSITY
PRESS

Department of Psychology, University of Tartu, Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in Psychology) on December 6, 2006, by the Doctoral Committee of the Department of Psychology, University of Tartu

Supervisor: Jüri Allik, PhD, Professor
University of Tartu, Estonia

Opponent: Marcel A. G. van Aken, PhD, Professor
Utrecht University, The Netherlands

Commencement: January 22, 2007

ISSN 1024–3921
SBN 978–9949–11–522–8 (trükis)
ISBN 978–9949–11–523–5 (PDF)

Autoriõigus Kaia Laidra, 2007

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr 687

CONTENTS

LIST OF ORIGINAL PUBLICATIONS.....	6
INTRODUCTION	7
1. Concept of Personality Trait.....	7
2. Five-Factor Model	9
3. FFM in Children and Adolescents.....	10
3.1. Measuring FFM in Children and Adolescents.....	11
4. Personality Development.....	12
5. Cross-Observer Agreement	13
6. External Validity and Predictive Utility of Personality Traits.....	14
7. Summary and Conclusions	15
ACKNOWLEDGEMENTS.....	17
REFERENCES	18
SUMMARY IN ESTONIAN.....	21
PUBLICATIONS.....	23

LIST OF ORIGINAL PUBLICATIONS

This dissertation is based on the following publications and an unpublished manuscript which will be referred to in the text by their respective Roman numerals:

- I Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality, 18*, 445–462.
- II Laidra, K., Allik, J., & Pullmann, H. (2006). Personality development in adolescence: Analyses of longitudinal and cross-sectional sequences. Manuscript under review.
- III Laidra, K., Allik, J., Harro, M., Merenäkk, L., & Harro, J. (2006). Agreement among adolescents, parents, and teachers on adolescent personality. *Assessment, 13*, 187–196.
- IV Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42*, 441–451.
- V Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2002). A study of intelligence in Estonia. *Psychological Reports, 91*, 1022–1026
- VI Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2004). Sex differences on the progressive matrices among adolescents: Some data from Estonia. *Personality and Individual Differences, 36*, 1249–1255.

INTRODUCTION

The developmental period of adolescence is typically portrayed as the most difficult or at least confusing time in life when individuals are more likely to experience “storm and stress” than at other ages (Arnett, 1999). Being marked by rapid physical and cognitive growth as well as by great social transitions, adolescence is a particularly interesting time to study how all these developmental changes are reflected in personality traits (Donnellan, Trzesniewski, & Robins, 2006).

Adolescent personality is examined from three aspects in this dissertation. First, both cross-sectional and longitudinal data are used to describe the developmental trajectories of personality traits in adolescence (Studies I and II). Then, different perspectives on adolescents’ personalities as perceived by teachers, mothers, fathers, and adolescents themselves are compared (Study III). Finally, the extent to which personality traits in children and adolescents are related to a real-life outcome, such as academic achievement, is assessed (Study IV). As academic achievement is known to depend strongly on cognitive abilities, it was important to determine the independent contribution of personality traits above that of intelligence. In order to do this, students’ intellectual abilities were measured with Raven’s Standard Progressive Matrices (Raven, 1981), which was standardized in Estonia as reported in Studies V and VI.

Besides being of interest in their own right, the three topics about the adolescent personality (Studies I–IV) also bear direct relevance to the fundamental issue in personality psychology concerning the utility of the concept of *personality trait*. In the following sections, I will outline the general theoretical background within which the empirical studies were conducted and briefly discuss their results in relation to the background.

1. Concept of Personality Trait

The assumption that people have personality traits provides a basis for many decisions in every-day life. For instance, someone described as “dependable” is more likely to be offered job and someone described as “honest” is more likely to be lent money. Traits are recognized not only by laypersons, who use thousands of trait terms for describing themselves and others, but this approach is also the dominant perspective in personality psychology (McCrae & Costa, 1999). In fact, personality measures, whatever their theoretical origins, are usually measures of traits (McCrae & Costa, 2003, p. 20).

Personality traits can be defined as “dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings, and actions” (McCrae & Costa, 2003, p. 25). Although there are several other variants of the definition of trait, almost all of them refer to the relative consistency and

ubiquity of traits (Roberts & Caspi, 2001). Temporal stability and cross-situational generalizability are the two critical features of traits that allow individual's behavior to be predicted, at least to a certain extent. However, traits are only dispositions to act in one way or another; they are not the absolute and only determinants of one's behavior.

The fact that the same person acts differently on different occasions and different situations is challenging to trait psychology. "The more that the typical individual acts differently from occasion to occasion, the less useful it is to label him or her as acting a particular way, that is, to describe his or her personality" (Fleeson & Jolley, 2006, p. 41). Although the cross-situational consistency of traits was most hotly disputed about 30–40 years ago, during the person-situation debate (see Epstein & O'Brien, 1985; Kenrick & Funder, 1988), the same or very similar questions concerning the existence of traits are still raised, every now and then (e.g., Lewis, 2001). The "situational" hypothesis is often viewed as an alternative to the trait position, but actually the two approaches need not be at odds with one another, because there are many ways in which persons and situations interact (Kenrick & Funder, 1988).

Trait approach implicitly assumes that people are generally rational and capable of understanding themselves and others (Funder, 1995). Traits are never directly observed (Allport, 1966), but they are inferred through what people say and do (Funder, 1994). As laypersons are deemed to be extraordinarily sophisticated judges with respect to their own and others' personalities (McCrae & Costa, 1999), traits are typically measured by questionnaires, asking persons to report on their own or others' thoughts, feelings, and typical behaviors.

The most fundamental assumption underlying the trait approach is the assertion that traits are real and exist within a person (Funder, 1991). Viewed as real, traits have to meet the following criteria: There must be evidence of longitudinal stability, cross-observer and cross-measure agreement; also, traits must be generalizable across different age groups and cultures, and they have to show relations to independent criteria (Costa & McCrae, 1992). In this dissertation, four of these criteria (i.e., applicability across age groups – Study I; longitudinal stability – Study II; cross-observer agreement – Study III; and external validity – Study IV) are addressed in the framework of the five-factor model (FFM).

2. Five-Factor Model

Despite the abundance of words that natural languages provide for describing individual differences in personality, these differences can adequately be represented by a small number of higher-order factors. Although the controversy about the number of factors seems to never end, many researchers have converged on the five-factor or “Big Five” model (Goldberg, 1993; John & Srivastava, 1999; McCrae & John, 1992) as the model that represents personality at the broadest level of abstraction. The FFM is the “meeting point” of lexical and questionnaire traditions: Factor analyses of ratings on both adjective scales and questionnaire items typically yield the same five factors.

John and Srivastava (1999) summarize the content of the five factors as follows: *Neuroticism* contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense. *Extraversion* implies an energetic approach toward the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality. *Openness to Experience*, as opposed to closed-mindedness, describes the breadth, depth, originality, and complexity of an individual’s mental and experiential life. *Agreeableness* contrasts a prosocial and communal orientation towards others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty. Finally, *Conscientiousness* describes socially prescribed impulse control that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks.

The FFM does not represent a particular theoretical perspective; it provides a primarily descriptive rather than explanatory account of personality. During the last decade, the FFM has become a standard framework for studying personality. According to McCrae and John (1992), the appeal of the FFM to researchers is threefold: First, as other personality constructs can easily be mapped onto the five factors, the model integrates a wide array of personality constructs, and facilitates therefore communication among researchers of many different orientations. Second, it is comprehensive, giving a basis for systematic exploration of the relations between personality and other phenomena. Third, it is efficient, providing a global description of personality with only five scores.

Personality descriptors can be organized hierarchically. While the Big Five dimensions represent the highest level of this hierarchy, each of them summarizes a large number of distinct, more specific personality characteristics (Goldberg, 1993; John & Srivastava, 1999). Thus, the FFM does not imply that personality differences can be reduced to only five traits. The unique position of the FFM traits in the personality hierarchy is reflected by the finding that other Big Trait models (e.g., Eysenck’s Gigantic Three) can be derived from the Big Five in some way. Markon, Krueger, and Watson (2005) have recently argued that the superordinate personality hierarchy is unbalanced, i.e., the Big Five

traits differ in their level of abstraction; Neuroticism in particular appears to exist at a more abstract level than the other FFM traits. Markon et al. (2005) suggested that other trait models might be related to the FFM in a hierarchical manner. For example, the three factors of Eysenck's model may occupy a level of the hierarchy above the five factors of the FFM.

3. FFM in Children and Adolescents

The growing consensus about adult personality structure provides now greater clarity about the traits that developmental studies should try to predict over time (Shiner & Caspi, 2003). Early-emerging individual differences in children have traditionally been studied within the framework of various temperament models and until quite recently, there has been little communication between temperament and personality researchers. Fortunately, there is increasing recognition that temperament and personality models share many traits in common (Shiner & Caspi, 2003). The studies summarized in Halverson, Kohnstamm, and Martin (1994) were the first attempt to link the models of childhood temperament/personality with the adult FFM.

On the basis of various temperament and personality models in children and adolescents, Shiner (1998, 2006; Shiner & Caspi, 2003) has proposed a preliminary taxonomy of traits that can be measured in children since preschool-age. The following higher- (and lower-) order traits are included in this taxonomy: Extraversion/Positive Emotionality (social inhibition, sociability, dominance, activity level), Neuroticism/Negative Emotionality (anxious distress, irritable distress), Conscientiousness/Constraint (attention, inhibitory control, achievement motivation), and Agreeableness (antagonism, prosocial tendencies); thus, four of the FFM factors are clearly represented. Openness to Experience does not appear to be as broad in childhood as its adult conceptualizations suggest; in children, Openness is more closely related to intellect (Shiner, 2006). Empirical studies measuring the FFM traits in children and adolescents tend to find evidence for the five-factor structure (e.g., Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003; De Fruyt, Mervielde, Hoekstra, & Rolland, 2000; Del Barrio, Carrasco, & Holgado, 2006; John, Caspi, Robins, Moffitt, & Stouthamer-Loeber, 1994; McCrae et al., 2002), although there is indeed some indication that Openness to Experience may not emerge until adolescence (e.g., Markey, Markey, & Tinsley, 2004).

Another line of evidence for the salience of the FFM in children comes from a cross-cultural study of parental free descriptions (Kohnstamm, Halverson, Mervielde, & Havill, 1998; Halverson, et al., 2003). Investigators in eight countries interviewed over 3000 parents and asked them to provide free-language descriptions of their 2–14-year old children. Convincing support for the usefulness of the FFM for coding parental descriptions was found, with

more than 80% of descriptors fitting into the FFM. On the basis of this “active vocabulary” for describing children’s traits, a new age-specific personality measure – *Inventory of Children’s Individual Differences* (ICID; Halverson, et al., 2003) – was constructed.

3.1. Measuring FFM in Children and Adolescents

Generally, there are three strategies to assess the FFM in non-adult age groups (De Fruyt et al, 2000). First, a “bottom-up” strategy can be used, which starts with a careful study of the range of individual differences that can be reliably observed in the target age group and results in the construction of a new FFM inventory for the particular age group (e.g., ICID; Halverson, et al., 2003). A second strategy is to derive FFM scores from inventories that originally intended to assess other personality models (e.g., John et al., 1994, derived five-factor scores via a re-analysis of California Child Q-set data). The third approach is to administer an adult FFM measure, such as the NEO Five-Factor Inventory (e.g., Parker & Stumpf, 1998) or the Revised NEO Personality Inventory (e.g., De Fruyt et al., 2000). The use of adult inventories enables comparisons across developmental periods, but, of course, this approach can not be used with younger children.

In Studies I–IV, adult personality measures NEO-FFI and NEO-PI-R (Costa & McCrae, 1992; Kallasmaa, Allik, Realo, & McCrae, 2000) were used for adolescents’ self-reports on the FFM dimensions. In Study I, adolescents’ mean levels and factor structure of self-reported traits were compared with those of Estonian adults. Small mean-level differences between adolescent and adult samples were found, with a bigger gap in Openness. With increasing age, factor structures of adolescents’ self-reported traits became increasingly similar to the adult target structure. However, the five-factor structure of adolescents’ traits was clearly recognizable already in the group of 6th graders, suggesting that the FFM is an appropriate framework for studying personality differences in adolescence.

In addition to the well-known and well-validated NEO-FFI and NEO-PI-R, two more age-specific Big Five measures were used for observer ratings (and self-reports in Wave 1) in Study III and for pre-adolescents’ self-reports in Study IV. The Estonian Brief Big Five Inventory (Study III) was constructed following the example of the “Common Language” California Child Q-Set (John et al., 1994), while the Estonian Big Five Questionnaire for Children (Study IV) was inspired by the Big Five Questionnaire – Children version by Barbaranelli et al. (2003). Both measures demonstrated acceptable convergence with the NEO questionnaires in separate samples of adults and adolescents.

4. Personality Development

By definition, traits are expected to exhibit certain levels of consistency over time, therefore, the question of stability and change can be considered crucial in trait approach to personality. Consistency is typically evaluated as the simple correlation between two measurements taken at two different points in time. As noted by Paunonen (2001), longitudinal stability entails consistency both across situations and over the long term. Conclusions regarding the temporal stability of personality may be rather contradictory. While some researchers see personality traits as practically fixed, at least in adulthood (e.g., McCrae & Costa, 2003), there is also a minority of those who argue that longitudinal studies find only weak correlations of personality variables over time (e.g., Lewis, 2001). Researchers belonging to the first group conclude that evidence concerning stability of personality across the lifespan supports the existence of global traits (Funder, 1991); others, on the other hand, call into question the meaning of personality (Lewis, 2001). This discrepancy probably reflects different appraisals of the same data – “a case of the cup being half full versus half empty” (Paunonen, 2001, p. 91). One question is, of course, how much stability is required in order to claim that a personality variable is consistent over time? There seems to be a tendency for researchers to compare the magnitude of longitudinal correlations with unity (e.g., Roberts & DelVecchio, 2000); however, it should be kept in mind that the inherent unreliability of personality measures sets limits to the maximum correlations that can be expected (Paunonen, 2001).

As there is evidence of both stability and change (Caspi & Roberts, 2001), it is suggested to replace the often-used phrase “stability *or* change”, which refers to the mutual exclusivity of these two phenomena, with the phrase “stability *and* change” when it is used to describe personality development (Mroczek, Almeida, Spiro, & Pafford, 2006). Similarly, Watson (2004) calls upon to abandon the simplistic opposition between the two concepts and recommends quantifying the change instead.

In Studies II and III, moderate levels of rank-order stability over 2–4 years were found in adolescents’ self reports (Studies II and III) as well as in parental ratings (Study III). The magnitude of test-retest correlations over time was comparable to previous studies (Roberts & DelVecchio, 2000) and in line with the *cumulative continuity principle* which states that “the relative consistency of personality traits continues to increase throughout the life span” (Caspi, Roberts, & Shiner, 2005, p. 469). On the other hand, there were high levels of mean-level continuity, as the only factor that demonstrated remarkable normative increase in adolescence was Openness (Study II). Considering that adolescents’ mean trait levels differed from those of adults most with respect to Openness (Study I), the rise of Openness was to be expected.

5. Cross-Observer Agreement

The issue of how different judges see someone's personality is important both on a practical and theoretical level. Theoretically, cross-observer agreement is another argument for the concept of personality trait, demonstrating an empirical basis for individual differences. On a practical level, it is related to the question who should provide personality ratings?

There are primarily two ways to access information about people – what they say about themselves and what others say about them. However, both ways have their flaws and strengths (McCrae, 1982; Roberts & Wood, 2006). While self-reports possess the unique advantage of access to the private thoughts and fantasies of the individual, they are typically derogated for being biased by response sets, such as social desirability responding. Observer methods, on the other hand, are regarded as more objective and free from the particular biases of self-reports, but their validity is undermined by the fact that observers do not have complete access to a person's thoughts, feelings, and behaviors. Furthermore, observer ratings may also suffer from certain biases such as halo effect. Although self-reports occupy the leading position in personality research, observer ratings are generally afforded greater respect (Roberts & Wood, 2006). One reason to prefer observer ratings comes from the possibility to increase their reliability by averaging the ratings provided by several observers. Hofstee (1994, p. 149) suggests that “the averaged judgment of knowledgeable others provides the best available point of reference both for the definition of personality structure in general and for assessing someone's personality in particular.” However, another position considers self and other perspectives of two distinct ways to understand a person and neither of them should be dominant in defining personality (Meyer et al., 2001; Roberts & Wood, 2006).

According to McCrae (1982, p. 294), a comparison of self-report personality measure with ratings by individuals who know the subject well provides “a direct test of the proposition that our impressions of personality reflect veridical attributes of the person”. Again, the same question can be raised: How much agreement between self-report and observer ratings would be sufficient to support the trait concept? Is the typical .50 enough (cf. Pervin, 1994)? While McCrae (1982, p. 302) concludes that self-other agreement is “certainly high enough to overturn the notions that personality traits are pure fictions”, Lewis (2001) believes that the observed level of interjudge agreement is too low for such conclusions. Nevertheless, the latter view can be considered marginal, because it is not well-substantiated but rather appears to reflect a personal conviction.

Another, yet related, perspective on interrater agreement is to consider it as a possible criterion for accuracy of person perception (e.g., Funder, 1995; Kenny, 1991). It can be argued that if a group of judges were to be perfectly accurate, they would necessarily manifest perfect agreement with one another. In social

perception, however, exact accuracy is rare, and only partial accuracy is the norm (Kenny, 1991). Because many people can agree and still all be wrong, Funder (1995) claims that agreement is a necessary but not sufficient criterion for accuracy. Kenny (1991), on the other hand, argues that consensus is neither a necessary nor sufficient condition: If there are independent sources of variance to which two judges have differential access, it is possible for two judges not to agree, with both being partially accurate (Kenny, 1991). However, in the case of personality perception, these conditions would be met only if there was absolutely no cross-situational consistency.

Study III revealed both agreement and disagreement in the way adolescents' personality characteristics were perceived by different observers. There was a high consensus between mothers and fathers, moderate agreement between parents and adolescents, but only low agreement between teachers and all other raters. These findings suggest that there are definitely consistencies in adolescents' behavior upon which different raters can agree, but adolescents also vary their behavior as a function of setting, and the behavioral consistencies they exhibit at home are not the same (or cannot be reliably observed) at school. While other researchers have argued that teacher ratings are "the ideal method for measuring the personality traits of children too young to provide self-reports" (Hampson & Goldberg, 2006, p. 764), Study III does not support this notion with regard to adolescents.

6. External Validity and Predictive Utility of Personality Traits

A further argument for the veridicality of traits comes from their ability to predict external criteria. If traits were only cognitive fictions, it would be difficult to understand how they can explain real-life outcomes (McCrae & John, 1992). Thus, personality traits must demonstrate external and predictive validity in order to be regarded as "real".

There is ample evidence that childhood individual differences predict important outcomes in adulthood (Roberts & Caspi, 2001). In a recent review, Caspi et al. (2005) summarized the evidence concerning associations between early personality traits and later outcomes in three domains: social relationships, educational and work performance, and health. While all of the higher-order traits appear to be significant predictors of social competence among children, in adulthood, Neuroticism and Agreeableness are more consistently related to relationship outcomes than other traits. Regarding educational and work performance, Conscientiousness-related traits are widely found to be the best non-cognitive predictors of academic and occupational achievement; associations with other traits depend more on the specifics of the job. With respect to health outcomes, life-span studies have documented links between longevity and high

Positive Emotionality and Conscientiousness, while other research reveals that individuals high in anger and hostility (traits inversely related to Agreeableness) are at greater risk of developing cardiovascular disease. Caspi et al. (2005, p. 471) interpret these findings as “robust evidence that early-emerging individual differences in personality shape how individuals experience, interpret and respond to the developmental tasks they face across the life course.”

The prediction afforded by personality traits is typically modest in size, providing a basis for criticism against the trait approach (e.g., Pervin, 1994). In response to such critiques, Shiner and Caspi (2003) remind that even small effect sizes are of theoretical and practical significance. They also argue that large effect size for a single trait would be improbable, taking into account that every behavior is a product of multiple additive and interacting traits. Furthermore, focusing on a single outcome variable measured at a single point in time may underestimate the contribution of personality traits to the course of development because the effect of traits accumulates over lifetime (Shiner & Caspi, 2003).

In Study IV, associations between children’s and adolescents’ personality traits, intelligence (see Studies V and VI on the standardization of intelligence test), and concurrent academic achievement were examined. Although academic achievement was most strongly related to intelligence, personality traits showed incremental validity over the intelligence measure. In particular, Agreeableness in elementary school and Conscientiousness in secondary school were the best personality predictors of students’ grade point average. These results support the conclusion that personality traits can be used to predict academic achievement and that they add to the prediction of achievement even when cognitive abilities are taken into account.

7. Summary and Conclusions

In sum, the main results and conclusions of this dissertation are as follows:

- ✓ The five-factor model is an adequate taxonomy for describing personality differences in adolescence. The five-factor structure of self-reported personality traits was clearly recognizable already at age 12 and became practically indistinguishable from adult personality structure by age 16 (Study I).
- ✓ Personality development during adolescence is characterized by generally small mean-level changes and moderate rank-order stability (Studies I, II, and III). From Grade 6 to 12, both longitudinal and cross-sectional data indicated an increase in self-reported Openness, as well as small declines in Agreeableness and Conscientiousness (Studies I and II). However, different mean-level changes were found in observer ratings provided by parents and teachers (Study III).

- ✓ Developmental effects found in longitudinal and cross-sectional data might be confounded by practice or cohort effects, respectively. Examination of these effects revealed no clear evidence of a practice effect on self-reported personality scores, but there was a cohort effect for Openness with younger cohorts scoring higher on this dimension (Study II). Thus, cross-sectional research may underestimate the increase of Openness in adolescence.
- ✓ Not all adolescents are characterized by the same developmental trajectory. Growth curve modeling revealed significant individual variation in the growth trajectories for four personality factors, which could not be explained by adolescents' sex or cohort (Study II).
- ✓ Raters from different settings may describe adolescents' personality characteristics quite differently. Ratings provided by adolescents, their parents, and teachers on adolescents' personality traits converged only to a modest extent (Study III). While there was a high consensus between mothers and fathers in terms of both differential agreement and mean ratings, parental ratings were only moderately correlated to adolescents' self-reports. Nevertheless, teacher ratings demonstrated even less agreement with all other raters.
- ✓ Teachers' ratings may not be the most reliable source for deciding about adolescents' personalities. Besides diverging from all other raters, teacher ratings also showed less consistency over time than the ratings provided by other informants (Study III). Three-year test-retest correlations were significant only for teacher-rated Extraversion and Openness, whereas self- and parent ratings yielded moderate test-retest correlations on all five factors.
- ✓ Adolescents' personality traits are meaningfully related to many real-life outcomes, including academic achievement. Although students' grade point average was best predicted by intellectual abilities, four personality factors had significant, albeit small, incremental effects on GPA (Study IV). While Agreeableness was the strongest predictor among personality factors in elementary school, in adolescence, its place was taken over by Conscientiousness.
- ✓ As a part of the research, Raven's Standard Progressive Matrices (Raven, 1981) were standardized in Estonian adolescents (Studies V and VI). The mean intelligence of Estonian adolescents was comparable to that of their British peers (Study V). Analysis of sex differences revealed higher scores for girls aged 12–15 and for boys aged 16–18, confirming Lynn's developmental theory of sex differences in intelligence (Study VI).

ACKNOWLEDGEMENTS

I have been fortunate to have worked with the right people and on the right research projects. For this, I am most grateful to my supervisor, Professor Jüri Allik, who believed in me and gave me the opportunity to collaborate on his projects. I highly appreciate his guidance and support throughout my studies.

Many thanks are due to all the people who contributed to the research presented in this dissertation, either as researchers or participants. My special gratitude belongs to my co-authors (in alphabetical order): Jaanus Harro, Maarike Harro, Richard Lynn, Liis Merenäkk, Helle Pullmann, and Anu Realo. Without their contribution, this dissertation would not exist.

I would also like to give my warmest thanks to all the friendly and helpful colleagues at the Department of Psychology. And last but not least, I thank my family and friends for being there for me.

REFERENCES

- Allport, G. W. (1966). Traits revisited. *American Psychologist*, *21*, 1–10.
- Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist*, *54*, 317–326.
- Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the big five in late childhood. *Personality and Individual Differences*, *32*, 645–664.
- Caspi, A. & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, *12*, 49–66.
- Caspi, A. & Roberts, B. W., & Shiner (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484.
- Costa, P. T., Jr. & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*, 653–665.
- Costa P.T., Jr., and McCrae R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and the NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J.-P., (2000). Assessing adolescents' personality with the NEO PI-R. *Assessment*, *7*, 329–345.
- Del Barrio, V., Carrasco, M. A., & Holgado, F. P. (2006). Factor structure invariance in the Children's Big Five Questionnaire. *European Journal of Personality Assessment*, *22*, 158–167.
- Donnellan, B. M., Trzesniewski, K. H., & Robins, R. W. (2006). Personality and self-esteem development in adolescence. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 285–309). Mahwah, NJ: Erlbaum.
- Epstein, S. & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, *98*, 513–537.
- Fleeson, W. & Jolley, S. (2006). A proposed theory of the adult development of intraindividual variability in trait-manifesting behavior. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 41–59). Mahwah, NJ: Erlbaum.
- Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, *2*, 31–39.
- Funder, D. C. (1994). Describing traits. *Psychological Inquiry*, *5*, 125–127.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34.
- Halverson, C. F., Havill, V. L., Deal, J. Baker, S. R., Victor, J. B., Pavlopoulos, V., Besevegis, E., & Wen, L. (2003). Personality structure as derived from parental ratings of free descriptions of children: The Inventory of Child Individual Differences. *Journal of Personality*, *71*, 995–1026.
- Halverson, C. F., Kohnstamm, G. A., & Martin, R. P. (Eds.), (1994). *The developing structure of temperament and personality from infancy to adulthood*. Hillsdale, NJ: Erlbaum.
- Hampson, S. E. & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology*, *91*, 763–779.

- Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality*, 8, 149–162.
- 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, 160–178.
- John, O. P. & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.
- Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the five-factor model. *European Journal of Personality*, 14, 265–278.
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98, 155–163.
- Kenrick, D. T. & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, 43, 23–34.
- Kohnstamm, G. A., Halverson, C. F., Mervielde, I., & Havill, V. L. (Eds.), (1998). *Parental descriptions of child personality: Developmental antecedents of the Big Five?* Mahwah, NJ: Erlbaum.
- Lewis, M. (2001). Issues in the study of personality development. *Psychological Inquiry*, 12, 67–83.
- Markey, P. M., Markey, C. N., & Tinsley, B. J. (2004). Children's behavioural manifestations of the Five-Factor Model of personality. *Personality and Social Psychology Bulletin*, 30, 423–232.
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88, 139–157.
- McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, 43, 292–303.
- McCrae, R. R. & Costa, P. T., Jr. (1999). A Five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 139–153). New York: Guilford.
- McCrae, R. R. & Costa, P. T., Jr. (2003). *Personality in adulthood: A Five-Factor theory perspective* (2nd ed). New York: Guilford Press.
- 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, 1456–1468.
- McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175–215.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128–165.
- Mroczek, D. K., Almeida, D. M., Spiro, A., III, & Pafford, C. (2006). Modeling intraindividual stability and change in personality. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 163–180). Mahwah, NJ: Erlbaum.

- Parker, W. D. & Stumpf, H. (1998). A validation of the five-factor model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences, 25*, 1005–1025.
- Paunonen, S. (2001). Inconsistencies in the personality consistency debate. *Psychological Inquiry, 12*, 91–93.
- Pervin, L. A. (1994). A critical analysis of current trait theory. *Psychological Inquiry, 5*, 103–113.
- Raven, J. (1981). *Manual for Raven's Progressive Matrices and Mill Hill vocabulary scales*. Oxford: Oxford Psychologists Press.
- Roberts, B. W. & Caspi, A. (2001). Personality development and the person-situation debate: It's déjà vu all over again. *Psychological Inquiry, 12*, 104–109.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3–25.
- Roberts, B. W. & Wood, D. (2006). Personality development in the context of the neo-socioanalytic model of personality. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 11–39). Mahwah, NJ: Erlbaum.
- Shiner, R. (1998). How shall we speak of children's personalities in middle childhood? A preliminary taxonomy. *Psychological Bulletin, 124*, 308–332.
- Shiner, R. (2006). Temperament and personality in childhood. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 213–230). Mahwah, NJ: Erlbaum.
- Shiner, R. & Caspi, A. (2003). Personality differences in childhood and adolescence: measurement, development, and consequences. *Journal of Child Psychology and Psychiatry, 44*, 2–32.
- Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality, 38*, 319–350.

SUMMARY IN ESTONIAN

Teismeliste isiksus: areng, hindajatevaheline üksmeel ja seosed akadeemilise edukusega.

Väitekirja käsitleb teismelise isiksust kolmest aspektist. Esmalt kirjeldatakse läbilõike- ja longituudandmetele tuginedes isiksusejoonte arengut teismeliseas (I ja II uuring). Seejärel kõrvutatakse erinevaid vaatepunkte, kuidas näevad teismelise isiksust õpetajad, emad, isad ja teismelised ise (III uuring). Viimasena on vaatluse all see, kuidas seostuvad laste ja teismeliste isiksusejooned nende edukusega koolis (IV uuring). Uuringute teoreetiliseks raamistuks on viie faktori mudel, mis kirjeldab inimestevahelisi isiksuse erinevusi viiel dimensioonil: neurootilisus, ekstravertsus, avatus, sotsiaalsus ja meelekindlus.

Väitekirja peamised tulemused ja järeldused on järgmised:

- ✓ Viie faktori mudel on sobiv taksonoomia teismeliste isiksuse kirjeldamiseks. Viiefaktoriline struktuur oli äratuntav juba 12-aastaste enesekohastes isiksusehinnangutes ja 16. eluaastaks muutus hinnangute struktuur praktiliselt eristamatuks täiskasvanute isiksuse struktuurist (I uuring).
- ✓ Teismeliste isiksuse arengut iseloomustavad üldiselt väikesed muutused isiksusejoonte keskmistes tasemetes ja mõõdukas suhteline stabiilsus (I, II ja III uuring). Enesekohased longituud- ja läbilõikeandmed näitasid 6.–12. klassini tõusu avatuse dimensioonil ning väikest langust sotsiaalsuse ja meelekindluse dimensioonidel (I ja II uuring). Need muutused ei langenud aga kokku muutustega, mis ilmsid vanemate ja õpetajate hinnangutes (III uuring).
- ✓ Longituuduuringu tulemusi võib mõjutada korduv testimine, läbilõikeuuringu tulemusi aga kohordi efekt. Nende efektide uurimisel ei leitud, et korduv testimine avaldaks enesekohastele isiksusehinnangutele selget mõju. Küll aga ilmsid kohordi efekt avatuse dimensioonil, kus nooremad kohordid said kõrgemaid skoori (II uuring). Seega võivad läbilõikeuuringud teismeliste avatuse kasvu alahinnata.
- ✓ Mitte kõigi teismeliste isiksuse areng ei ole ühesugune. Kasvukõverate mudeldamisel leiti olulised individuaalsed erinevused nelja isiksusefaktori arenguteedes, mis ei olnud seletatavad teismeliste soo või kohordiga (II uuring).
- ✓ Teismelistega erinevas kontekstis kokku puutuvad hindajad võivad kirjeldada nende isiksust üsna erinevalt. Teismeliste endi, nende vanemate ja õpetajate antud hinnangud langesid kokku vaid vähesel määral (III uuring). Kui emade ja isade arvamused olid üksteisele väga lähedased (nii suhtelise kokkulangevuse kui keskmiste hinnangute mõttes), siis teismeliste enesekohaste hinnangutega korreleerusid nad vaid mõõdukalt. Õpetajate hinnangud langesid kõigi teiste hindajate arvamustega veelgi vähem kokku.

- ✓ Õpetajate hinnangud ei pruugi olla väga usaldusväärseks allikaks, mille alusel teismeliste isiksuse üle otsustada. Lisaks sellele, et õpetajate hinnangud olid nõrgalt seotud teiste hindajate omadega, olid nad ka ajaliselt vähem püsivad kui teiste hindajate arvamused (III uuring). Kolmeaastase vahega toimunud testimiste skooride korreleerimisel olid olulised vaid õpetajate hinnatud ekstravertsuse ja avatuse test-retest korrelatsioonid, samas kui enesekohaste ja vanemate hinnangute puhul olid kõigil viiel faktoril mõõdukad test-retest korrelatsioonid.
- ✓ Teismeliste isiksusejooned on ootuspärasel viisil seotud mitmete eluliste näitajatega, sealhulgas kooliedukusega. Ehkki õpilaste keskmist hinnet ennustas kõige paremini nende intellektuaalne võimekus, ilmnis neljal isiksusefaktoril oluline, ehkki väike, võimetest ja teistest isiksusejoontest sõltumatu panus keskmise hinde ennustamisel (IV uuring). Kui algklassides ennustas isiksusefaktoritest õppeedukust kõige paremini sotsiaalsus, siis teismelistel võttis selle koha üle meelekindlus.
- ✓ Uurimistöö ühe osana standardiseeriti Eesti teismeliste valimil Raveni Progresseeruvad Maatriksid (Raven, 1981; V ja VI uuring). Eesti teismeliste keskmine intelligentsus oli sarnane Briti eakaaslaste omale (V uuring). Kooskõlas Lynni intelligentsuse sooliste erinevuste arengu teooriaga olid 12–15-aastaselt tüdrukute intelligentsuse skoorid kõrgemad kui poistel, 16–18-aastaselt aga poiste skoorid kõrgemad kui tüdrukutel (VI uuring).

PUBLICATIONS

Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004).
Personality development from 12 to 18 years of age:
Changes in mean levels and structure of traits.
European Journal of Personality, 18, 445–462.

Laidra, K., Allik, J., & Pullmann, H. (2006).
Personality development in adolescence:
Analyses of longitudinal and cross-sectional sequences.
Manuscript under review.

Laidra, K., Allik, J., Harro, M., Merenäkk, L., & Harro, J. (2006).
Agreement among adolescents, parents, and teachers on adolescent personality.
Assessment, 13, 187–196.

Laidra, K., Pullmann, H., & Allik, J. (2007).
Personality and intelligence as predictors of academic achievement:
A cross-sectional study from elementary to secondary school.
Personality and Individual Differences, 42, 441–451.

Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2002).
A study of intelligence in Estonia.
Psychological Reports, 91, 1022–1026

Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2004).
Sex differences on the progressive matrices among adolescents:
Some data from Estonia.
Personality and Individual Differences, 36, 1249–1255.

Personality Development in Adolescence: Analyses of Longitudinal and Cross-Sectional Sequences

Kaia Laidra, Jüri Allik, and Helle Pullmann

Department of Psychology, University of Tartu, Tartu, Estonia
The Estonian Centre of Behavioural and Health Sciences

Abstract. Development of the five-factor personality traits in adolescence was examined using longitudinal and cross-sectional data from three waves of the Longitudinal Study of Estonian Schoolchildren. Samples of 6th, 8th, 10th, and 12th graders completed the NEO Five-Factor Inventory three times with 2-year intervals. The longitudinal sample comprised 323 adolescents who participated each time, while cross-sectional and time-lag comparisons were performed on data of a total of 3212 adolescents. Results of longitudinal and cross-sectional analyses converged in demonstrating small age-related mean-level changes in personality, with an exception of a more pronounced increase in Openness by 0.17 standard deviations per year. Openness was also the only trait that exhibited a clear time-lag increase, suggesting a cohort effect on this dimension. There were significant individual differences in the developmental trajectories of most of the five-factor personality traits, but they were not related to adolescents' sex or cohort.

Keywords: Five-factor model; adolescence; personality development; longitudinal and cross-sectional sequences; growth curve analysis

INTRODUCTION

Personality traits are not static, invariant over time; they rather show systematic and meaningful changes during the whole human life span (Caspi & Roberts, 2001; Watson, 2004). These changes, however, are relatively slow and modest in their magnitude, suggesting that continuity is the governing principle of personality development (McCrae & Costa, 2003; Terracciano, Costa, & McCrae, 2006). Because there is evidence for both stability and change in personality, researchers should avoid the simplistic opposition between the two concepts; instead, the main research task is to quantify the changes in personality traits (Watson, 2004). In order to take a quantitative approach, a

more sophisticated research methodology must be adopted than it is typical to stability studies. In particular, it is necessary to have a sufficiently large sample, several different scales assessing a wide spectrum of personality traits, and at least two different retest intervals extending across a time period that is sufficient for true change to occur (cf. Watson, 2004). In this study, the proposed approach is applied to examine the change and stability of the five-factor personality traits (Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness; McCrae & John, 1992) in adolescence.

Continuity and change in personality are most commonly operationalized as rank-order stability and mean-level change, although other types of continuity are also occasionally examined (e.g., De Fruyt et al., 2006; Robins, Fraley, Roberts, & Trzesniewski, 2001). *Rank-order* or *differential stability* refers to the extent that individuals' relative standing on a trait dimension remains invariant across time, whereas *mean-level* or *normative change* refers to the extent that average personality scores change over time. Rank-order stability and mean-level change should be thought of as conceptually and empirically independent or orthogonal constructs that can exist simultaneously (Roberts, Walton, & Viechtbauer, 2006). Previous longitudinal research with regard to these two types of continuity has been summarized in recent meta-analyses by Roberts and colleagues, which revealed that over the life span personality traits become increasingly rank-order stable (Roberts & DelVecchio, 2000) and show a clear pattern of trait-specific mean-level change in the direction of greater maturity (Roberts et al., 2006).

Although adolescence is commonly regarded as a time of transition marked by relatively dramatic biological, cognitive, and social changes (Donnellan, Trzesniewski, & Robins, 2006), the meta-analysis conducted by Roberts et al. (2006) showed few robust mean-level changes in the Big Five personality traits during these transitional years. The only traits that increased significantly from ages 10 to 18 were Social Dominance (a facet of Extraversion related to independence and dominance) and Emotional Stability, while there was also a small but insignificant increase in Openness. On the other hand, mean-level changes in Agreeableness, Conscientiousness, and Social Vitality (another facet of Extraversion, related to positive affect, activity level, and sociability) were negligible during this period. Whereas over the adolescence mean levels remain relatively stable, differential stability is moderate, with a mean test-retest correlation of 0.43 (Roberts & DelVecchio, 2000). This indicates that adolescents still undergo substantial change but its direction may vary from person to person resulting in little or no change in mean levels, or as formulated by Donnellan et al. (2006, p. 289) "adolescence is best characterized as a period of individual, rather than normative change in personality."

Methods typically employed to estimate rank-order and mean-level stability in personality conceal important information on individual differences in stability and change (Mroczek, Almeida, Spiro, & Pafford, 2006). New statistical techniques that allow to estimate individual differences in stability and

change, such as growth curve modeling based on structural equation models or multilevel modeling (see Mroczek et al., 2006, for an overview of growth curve modeling in personality), are now used by researchers on adult personality development (e.g., Helson, Jones, & Kwan, 2002), providing an analytic tool for quantifying personality change. However, to date there are few studies that have applied growth curve modeling to personality development in adolescence (e.g., Branje, Van Lieshout, & Gerris, in press). In one of them, Branje and colleagues (in press) examined developmental trajectories of the Big Five traits for a sample of Dutch adolescents aged between 11 and 17, for whom self-reports as well as parental ratings were collected three times with annual intervals. With regard to adolescents' self-reports different patterns of mean-level change were found for boys and girls: while girls increased in Openness, Agreeableness, and Conscientiousness and showed a curvilinear trajectory with an increase followed by a decrease in Extraversion, the only significant mean-level change for boys was a decrease in Extraversion. On the other hand, significant variance in growth trajectories was found more often for boys than for girls: remarkable interindividual differences in change were found for boys' Emotional Stability, Openness, and Agreeableness, whereas girls' growth trajectories appeared to be more similar to each other. Additionally, Branje et al. (in press) explored the contribution of several moderators to variance in change, but none of the studied factors (maternal support, pubertal timing, and number of positive and negative life events) significantly predicted individual differences in personality development.

In developmental research, cross-sectional and longitudinal research designs are widely used. *Cross-sectional* analyses compare two or more age groups on a single testing occasion (e.g., a sample of 20-year-olds may be compared with a sample of simultaneously assessed 30-year-olds), whereas in the *longitudinal* method, the same group of persons is followed and measured repeatedly over time (e.g., at age 20 and at age 30). In both designs, differences between the observations are interpreted as age effects, although different age samples differ also with respect to other factors. While cross-sectional design confounds age with differences between cohorts or generations, in the longitudinal design age is confounded with time of measurement as well as with the effect of repeated testing known as practice effect. These possible confounders have often been overlooked in personality development research, in spite of the evidence that they may have an important influence on personality (Baltes & Nesselrode, 1972). For example, the meta-analyses of Twenge (2000, 2001) show that over recent decades, generations of Americans have shifted toward substantially higher levels of Neuroticism and Extraversion. The presence of cohort effects not only impairs the internal validity of cross-sectional studies, but also challenges the external validity of longitudinal studies, as the results from one cohort followed over time may not be generalizable to other cohorts (Baltes, 1968).

In order to disentangle age from confounding factors, sequential designs have been proposed (see Baltes, 1968; Schaie, 1965). Baltes (1968) distinguished between ontogenetic (age-related) and generational (cohort-related) components of change, which contribution can be separated by the use of *cross-sectional* and *longitudinal sequences*. While both methods observe two or more cohorts at the same two or more age levels, the distinction between the designs lies in the use of independent (in the case of cross-sectional sequences) vs. repeated measurements (in the case of longitudinal sequences). According to Baltes (1968), longitudinal differences as well as differences between independent samples representing the same cohort measured at different ages can be interpreted as age effects, whereas time-lag differences (i.e., differences between samples of equal age but drawn from different cohorts and therefore measured at different times) refer to generational effects. As different designs rule out different confounders, it is recommended to use them in combination (Costa & McCrae, 1982). When an effect of the same direction and magnitude is found in cross-sectional, cross-sequential, and longitudinal studies, such converging evidence would make it highly likely that a real aging effect has been found (McCrae & Costa, 2003). However, there are only few studies on adolescent personality development which have confronted results from different designs (e.g., Baltes & Nesselroade, 1972; McCrae et al., 2002).

The present study builds upon our previous work on personality development in adolescence (see Allik, Laidra, Realo, & Pullmann, 2004; Pullmann, Raudsepp, & Allik, 2006), presenting data from three waves of the Longitudinal Study of Estonian Schoolchildren (LSES). Based on cross-sectional analysis of the data from the first wave, Allik et al. (2004) reported only small differences in adolescents' mean levels on the five-factor model personality dimensions at ages 12 to 18 years: Openness was found to increase, while both Agreeableness and Conscientiousness decreased. Results of the 2-year longitudinal study (Pullmann et al., 2006) also suggested little change in adolescents' personality, as according to the Reliable Change Index, more than 80% of adolescents remained at the same level on each of the five dimensions. Longitudinal data confirmed the increase in Openness, but contrary to what was found cross-sectionally, Conscientiousness demonstrated no change across the age groups, whereas results for other traits were less consistent depending on the particular age group. The mean test-retest correlation over the 2-year interval was 0.58, suggesting moderate to high levels of rank-order stability.

The main goal of this study is to go beyond the dichotomous opposition of change versus stability and try to quantitatively estimate personality change in adolescence. This paper complements our previous studies by presenting new longitudinal data from the third wave of the LSES and including analyses of interindividual variability in personality development. Moreover, using sequential analysis strategies, an attempt is made to disentangle age effect from practice and cohort effects. More particularly, the following questions are addressed: How rank-order stable are adolescents' personality traits over two-

and four-year periods? Are those adolescents who are more stable over one interval equally stable over the second interval? What is the magnitude of mean-level and individual-level changes in personality? Do the developmental trajectories of personality traits depend on adolescents' sex or being born in a certain cohort? Are there any practice and cohort effects on personality scores?

METHOD

Participants

This paper presents findings from the three waves of the Longitudinal Study of Estonian Schoolchildren (LSES). The LSES started in 2001 (Time 1), when intellectual abilities and personality traits of 2,746 adolescents, attending Grades 6, 8, 10, and 12 in 27 Estonian-speaking public secondary schools and gymnasiums were assessed (see Allik et al., 2004, for details). In 2003 (Time 2) and 2005 (Time 3), follow-up studies were conducted in 17 schools, with 1681 students participating at Time 2 and 1595 students participating at Time 3. In every school, one class of students attending Grades 6, 8, 10, and 12 was tested each time. As a result, the follow-up samples included both students who had been previously assessed (e.g., 6th graders of Time 1 who had moved to Grade 8 for Time 2) as well as new students (e.g., 6th graders at Time 2 or Time 3). Data were collected during the second half of the school year (from February to April) each time. Consent was obtained from adolescents and their parents.

Compulsory schooling in Estonia begins at the age of seven. Basic school (Grades 1–9) is obligatory for all children and is usually followed by either upper secondary school (gymnasium; Grades 10–12) or vocational secondary school. Classroom size varies considerably depending on the location of the school, ranging from less than ten pupils per class in rural areas to approximately 35 pupils in cities. The same students are together in most of their classes throughout the school years. New classes are formed only in the transition from basic to secondary school (i.e., in Grade 10). Since most schools teach both basic and secondary programs, students can continue their studies in the same school but they may also change the school. While many participants of the LSES decided to change the school after Grade 9 (and consequently dropped out from the study), those who stayed in the same school were redistributed among different parallel classes, only one of which participated in the study.

Longitudinal sequences. The longitudinal sample consisted of 323 adolescents for whom complete personality data from the three study waves were available; 191 of them studied in Grade 6 at Time 1 (66 boys and 125 girls aged 11 to 13 years, with a mean age of 12.3 ± 0.5 years; henceforth, “younger

cohort¹”), whereas 132 were studying in Grade 8 (51 boys and 81 girls aged 13 to 17 years, with a mean age of 14.3 ± 0.5 years; “older cohort”). Given that only the 6th and 8th graders at Time 1 who attended the 17 schools that participated in follow-up studies ($N = 1017$) could have been tested three times, the attrition rate of the sample was 68.2%. The main reason for attrition was changing or leaving school or class during the transition from basic to secondary school.

Comparison of the longitudinal sample of 323 students with the initial sample from the 17 schools at Time 1 revealed that attrition was not completely random but students with certain characteristics were more liable to drop out. Those who stayed in the study were more frequently girls than boys ($\chi^2(1) = 31.24, p = 0.000$), and studied more frequently in Grade 6 at Time 1 than in Grade 8 ($\chi^2(1) = 9.34, p = 0.002$). The sex difference in attrition rate was stronger in younger cohort, as was indicated by significant interaction between sex and cohort ($\chi^2(3) = 44.12, p = 0.000$). Regarding personality traits, participating girls of older cohort differed from drop-outs in being higher in Conscientiousness ($m_1 = 31.9, SD_1 = 8.1$, vs. $m_2 = 28.2, SD_2 = 6.9$; $t(248) = 3.70, p < 0.001$), while there were no significant differences between younger girls who participated in follow-ups or not. Older boys who stayed in the study scored lower than drop-outs in Neuroticism ($m_1 = 20.6, SD_1 = 7.7$, vs. $m_2 = 23.7, SD_2 = 9.1$; $t(235) = -2.22, p < 0.05$); the same tendency appeared also for younger boys ($m_1 = 21.8, SD_1 = 7.4$, vs. $m_2 = 23.9, SD_2 = 7.8$; $t(260) = 1.92, p = 0.06$). Younger participating boys scored higher than drop-outs in Conscientiousness ($m_1 = 30.7, SD_1 = 6.9$, vs. $m_2 = 27.8, SD_2 = 7.1$; $t(260) = 2.79, p < 0.01$), whereas older boys showed the same tendency ($m_1 = 31.0, SD_1 = 8.3$, vs. $m_2 = 28.5, SD_2 = 7.8$; $t(235) = 1.94, p = 0.05$). Furthermore, comparison of participants’ and drop-outs’ scores on Raven’s Progressive Matrices (Raven, 1981; see Lynn, Allik, Pullmann, & Laidra, 2002, for results for the whole sample) revealed differences in boys’ but not in girls’ intellectual abilities, with participating boys scoring significantly higher than boys who dropped out (in younger boys $m_1 = 47.8, SD_1 = 5.5$, vs. $m_2 = 44.5, SD_2 = 7.6, t(259) = 3.27, p < 0.01$; in older boys $m_1 = 50.9, SD_1 = 5.0$, vs. $m_2 = 47.9, SD_2 = 6.5, t(230) = 3.03, p < 0.01$). It can be concluded that girls who continued their studies in the same school and thus were able to be assessed three times represented the initial sample of girls quite well; on the other hand, boys who comprised the longitudinal sample were more selected with respect to their intellectual and personality characteristics.

Control sample. In order to control practice effect in longitudinal studies, the inclusion of control groups, measured at the same time and representing the same age as the longitudinal sample but without prior testing experience, is suggested (Baltes, 1968; Schaie, 1965). For this purpose, data from the

¹ Here and henceforth, the term “cohort” designates the group of pupils studying at the same grade level at the same time and therefore sharing many environmental influences.

longitudinal sample at Time 2 and Time 3 were compared with data for their classmates who had not been assessed in previous study waves. At Time 2, there were 84 such students (45 girls and 39 boys, with a mean age of 14.6 ± 0.8 years) studying in Grade 8 and 263 students (167 girls and 96 boys, with a mean age of 16.4 ± 0.6 years) studying in Grade 10, whereas at Time 3 participated 137 “new” 10th graders (87 girls and 50 boys, with a mean age of 16.5 ± 0.6 years) and 67 “new” 12th graders (48 girls and 19 boys, with a mean age of 18.4 ± 0.7 years).

Cross-sectional sequences. For each participant of the three waves, data of her or his first assessment were used in order to test cohort effect on the mean personality scores. More particularly, this analysis included all the participants from the 17 schools that participated in follow-up studies at Time 1 (i.e., adolescents who comprised the longitudinal sample as well as “dropouts”) and all “new” participants at Time 2 and Time 3 (i.e., adolescents who comprised the control sample as well as adolescents from other grade levels who had not been previously assessed). In total, data from 1793 adolescents (831 boys and 962 girls, with a mean age of 14.9 ± 2.0 years) at Time 1, 726 adolescents (312 boys and 414 girls, with a mean age of 14.6 ± 2.1 years) at Time 2, and 693 adolescents (292 boys and 401 girls, with a mean age of 14.2 ± 2.2 years) at Time 3 were included.

Measures

In each wave, participants completed the Estonian version (Allik, et al., 2004) of the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), which is a 60-item measure of the five major personality domains: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each personality dimension is measured by 12 items on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). The questionnaire was administered collectively in classrooms during school hours. Cronbach alphas for the scales ranged from .67 (for Agreeableness in Grade 8 at Time 1) to 0.88 (for Neuroticism in Grade 12 at Time 3), with mean coefficients of .83 for Neuroticism, .82 for Extraversion, .79 for Openness, .71 for Agreeableness, and .81 for Conscientiousness.

RESULTS

Rank-order stability

Rank-order stability refers to the relative placement of individuals within a group over time and is most commonly assessed through test-retest correlations (Roberts & DelVecchio, 2000). Table 1 presents test-retest correlations (uncorrected for unreliability) for the five personality factors between the three assessments. Two consistent patterns of results are evident. First, concordantly

with Roberts & DelVecchio (2000), rank-order consistency decreases as the interval between the two assessments increases: correlations for all the scales are lower over the 4-year interval (between Time 1 and Time 3) than over the 2-year interval (between Time 1 and Time 2 or Time 2 and Time 3). Second, correlations between Time 2 and Time 3 are slightly higher than correlations between Time 1 and Time 2. This may reflect the effect of age on rank-order stability (cf. Roberts & DelVecchio, 2000), but may also result from a practice effect. For example, Grade 8 – Grade 10 (i.e., Time 2 – Time 3) correlations for the younger cohort appear to be more similar to Grade 10 – Grade 12 (Time 2 – Time 3) correlations for the older cohort than to Grade 8 – Grade 10 (Time 1 – Time 2) correlations for the older cohort that would have been expected if only age mattered. It appears that by completing the same inventory several times students “learn” to respond more consistently.

Table 1. Test-retest correlations of the five personality factors

	N	E	O	A	C	<i>m</i>
All, <i>N</i> = 323						
Time 1 – Time 2	.59	.53	.49	.52	.58	.54
Time 2 – Time 3	.66	.63	.65	.57	.68	.64
Time 1 – Time 3	.51	.41	.41	.43	.49	.45
Younger cohort, <i>N</i> = 191						
Grade 6 – Grade 8	.52	.47	.54	.51	.55	.52
Grade 8 – Grade 10	.64	.61	.66	.59	.67	.63
Grade 6 – Grade 10	.46	.33	.37	.45	.48	.42
Older cohort, <i>N</i> = 132						
Grade 8 – Grade 10	.68	.59	.44	.52	.62	.57
Grade 10 – Grade 12	.68	.66	.63	.54	.71	.64
Grade 8 – Grade 12	.58	.51	.45	.39	.51	.49

Note. N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness, *m* – mean correlation.

All the correlations are significant at $p < 0.001$ level.

Asendorpf (1992) has proposed a coefficient of individual stability which shows each individual’s contribution to the overall rank-order stability and is computed as follows:

$$i_{12} = 1 - (z_1 - z_2)^2/2,$$

where z_1 and z_2 are the z -transformed scores at Time 1 and Time 2. The mean of individual indices is equal to the correlation between the two assessments for the full sample. Individual stability indices between Time 1 and Time 2 as well as between Time 2 and Time 3 for each participant were calculated according to this formula. To investigate the stability of this index in adolescence, Spearman correlations between Time 1 – Time 2 and Time 2 – Time 3 individual stability

coefficients were computed (Spearman correlations were used because the distribution of coefficients is strongly skewed to the left; see Asendorpf, 1992). Correlations across the whole sample were $r = 0.05$, $p > 0.05$ for Neuroticism; $r = 0.13$, $p < 0.05$ for Extraversion; $r = 0.07$, $p > 0.05$ for Openness; $r = 0.16$, $p < 0.01$ for Agreeableness; and $r = 0.15$, $p < 0.01$ for Conscientiousness. In younger cohort, there were statistically significant correlations for Extraversion ($r = 0.18$, $p < 0.05$) and Conscientiousness ($r = 0.20$, $p < 0.01$), whereas in older cohort, only the correlation for Agreeableness was significant ($r = 0.23$, $p < 0.01$). In general, the relationship between individual stability coefficients for different time periods was weak; suggesting that at least in adolescence individual stability is not a stable characteristic of a person but may vary from one period to another.

Mean-level change and individual differences in change

Mean-level change is another commonly used type of trait stability that reflects whether groups of people increase or decrease on trait dimensions over time (Roberts & DelVecchio, 2000). However, not everyone is characterized by the same developmental trajectory (Mroczek et al., 2006). Growth-curve modeling was used to examine both intraindividual and interindividual change in adolescents' personality. For this purpose, the multilevel module available in LISREL 8.71 (Jöreskog & Sörbom, 2004) was employed (see Jöreskog, Sörbom, du Toit, & du Toit, 2001). The Level 1 units of analysis were occasions of measurement (Time 1, 2, and 3), and the Level 2 units were persons. Grade level was used as a proxy for age, because participants were sampled by grade level and not by age. Results of the growth-curve analyses for each of the five factors are presented in Tables 2 through 6.

Model 1 is a fully unconditional model, which functions as a baseline for comparison with more sophisticated models. Intraclass correlations, defined as the ratio of between-participants variance to the total variance, were computed for the random effects. Intraclass correlation coefficients were .57, .51, .44, .50, and .57, for Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness, respectively, indicating that about half of the variance in personality scores was attributable to the differences between participants.

To control for differences in mean personality scores between boys and girls and the two cohorts, the variance decomposition model was extended by including participants' sex and cohort as fixed effects in the model. As the results for Model 2 (Tables 2 to 6) indicate, the inclusion of sex and cohort reduced the variance at level 2. However, while the effect of sex on mean scores (with girls scoring higher than boys) was significant for each of the five traits, the only significant difference between the two cohorts was found for Agreeableness, on which students belonging to the younger cohort generally scored higher.

Table 2. Growth-curve models for Neuroticism

Predictor	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Fixed effects								
Intercept	22.01***	0.43	28.18***	0.73	31.60***	1.36	31.66***	1.38
Sex			-3.77***	0.87	-3.77***	0.87	-3.80***	0.87
Cohort			-1.35	0.85	-2.83**	0.88	-2.88***	0.87
Grade					-0.74***	0.11	-0.74***	0.12
Random effects								
Level 2								
Intercept/intercept	47.51***	4.73	43.89***	4.44	44.62***	4.44	115.97**	36.54
Grade/intercept							-8.48*	3.85
Grade/grade							1.02*	0.43
Level 1								
Intercept/intercept	35.86***	2.00	35.86***	2.00	33.65***	1.87	29.43***	2.22
$-2 \times \log$ likelihood	6736.68		6716.38		6675.40		6669.55	
No of free parameters	3		5		6		8	

Note. Sex: girls – 0; boys – 1. Cohort: older – 0; younger – 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 3. Growth-curve models for Extraversion

Predictor	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Fixed effects								
Intercept	31.31***	0.37	32.40***	0.62	26.27***	1.27	26.45***	1.26
Sex			-4.19***	0.74	-4.19***	0.74	-4.21***	0.73
Cohort			0.72	0.72	1.95**	0.75	1.73*	0.75
Grade					0.61***	0.11	0.61***	0.12
Random effects								
Level 2								
Intercept/intercept	33.82***	3.57	29.57***	3.25	30.07***	3.24	81.07**	30.85
Grade/intercept							-7.64*	3.41
Grade/grade							1.09**	0.39
Level 1								
Intercept/intercept	32.83***	1.83	32.83***	1.83	31.33***	1.74	26.72***	2.00
$-2 \times \log$ likelihood	6588.02		6555.82		6525.54		6511.14	
No of free parameters	3		5		6		8	

Note. Sex: girls = 0; boys = 1. Cohort: older = 0; younger = 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 4. Growth-curve models for Openness

Predictor	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Fixed effects								
Intercept	25.88***	0.38	28.07***	0.62	13.80***	1.31	13.79***	1.30
Sex			-5.59***	0.73	-5.59***	0.73	-5.87***	0.72
Cohort			-0.27	0.72	2.58***	0.75	2.71***	0.74
Grade					1.43***	0.12	1.43***	0.13
Random effects								
Level 2								
Intercept/intercept	33.01***	3.81	25.81***	3.25	28.52***	3.22	103.48**	33.15
Grade/intercept							-11.23**	3.74
Grade/grade							1.61***	0.44
Level 1								
Intercept/intercept	42.84***	2.38	42.84***	2.38	34.69***	1.93	27.81***	2.08
$-2 \times \log$ likelihood	6777.72		6724.32		6588.04		6559.64	
No of free parameters	3		5		6		8	

Note. Sex: girls = 0; boys = 1. Cohort: older = 0; younger = 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 5. Growth-curve models for Agreeableness

Predictor	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Fixed effects								
Intercept	30.52***	0.31	30.04***	0.54	33.08***	1.08	33.02***	1.09
Sex			-1.35*	0.63	-1.35*	0.63	-1.35*	0.64
Cohort			1.62**	0.62	1.02	0.65	1.07	0.65
Grade					-0.30**	0.09	-0.30**	0.10
Random effects								
Level 2								
Intercept/intercept	23.31***	2.48	22.21***	2.39	22.33***	2.39	52.93*	23.42
Grade/intercept							-3.60	2.50
Grade/grade							0.43	0.28
Level 1								
Intercept/intercept	23.11***	1.29	23.11***	1.29	22.74***	1.27	20.90***	1.55
-2 × log likelihood	6242.75		6231.08		6220.70		6218.34	
No of free parameters	3		5		6		8	

Note. Sex: girls = 0; boys = 1. Cohort: older = 0; younger = 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 6. Growth-curve models for Conscientiousness

Predictor	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Fixed effects								
Intercept	31.23***	0.37	32.03***	0.64	36.97***	1.19	36.92***	1.20
Sex			-2.04**	0.76	-2.04**	0.76	-2.10**	0.76
Cohort			-0.11	0.74	-1.10	0.77	-1.02	0.76
Grade					-0.49***	0.10	-0.49***	0.11
Random effects								
Level 2								
Intercept/intercept	35.11***	3.50	34.15***	3.42	34.48***	3.42	93.50***	27.28
Grade/intercept							-7.91**	2.94
Grade/grade							1.05**	0.33
Level 1								
Intercept/intercept	26.75***	1.49	26.75***	1.49	25.77***	1.43	21.45***	1.62
-2 × log likelihood	6450.24		6443.10		6419.10		6404.80	
No of free parameters	3		5		6		8	

Note. Sex: girls – 0; boys – 1. Cohort: older – 0; younger – 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Next, the fixed effect of grade was added to the previous model (see Tables 2 to 6, Model 3), which reduced level-1 or error variance but slightly increased level-2 variance. The effect of grade was significant for all the traits: the biggest change by grade level was found for Openness, which was expected to increase by 1.4 raw score points per grade, whereas Extraversion increased by 0.6 points and Neuroticism, Agreeableness, and Conscientiousness decreased by 0.7, 0.3, and 0.5 points per grade, respectively. Interestingly, when grade was entered, the effect of cohort became significant for three traits – Neuroticism, Extraversion, and Openness – and stopped being significant for Agreeableness. The changes in the mean levels of the five personality traits by grade, sex, and cohort are depicted in Figure 1. It can be seen that if the data were not presented by grade level but by study wave (i.e., lines for older cohort were shifted to the left), the lines of the two cohorts would practically overlap with each other. Apparently, not all the change is due to age, but other factors, such as practice, cohort or time-of-measurement effects, may be involved.

The log-likelihood values listed at the bottom of Tables 2 to 6 can be used as indices of fit: the smaller the value, the better the model fits to the data. The significance of the difference between log-likelihood values for two models can be tested using χ^2 distribution with degrees of freedom equal to the difference in the number of free parameters in the two models. Comparison of the log-likelihood values of Models 2 and 3 revealed significant reduction ($p < 0.001$) for all of the five factors when grade level was entered. However, for N, E, A, and C this reduction was small, from 0.2 to 0.6%, whereas for O the reduction of the log-likelihood value was more pronounced (2.0%).

It is expected that the linear growth rate may vary from participant to participant, rather than be fixed. Therefore, the random component on level 2 was included in Model 4. The reduction of the log-likelihood values was even smaller than was noted previously between Models 2 and 3: allowing the growth rates to vary between participants reduced the log-likelihood values only by less than 0.1 to 0.4%. While this change was statistically significant for Extraversion, Openness, and Conscientiousness ($p < 0.001$), it was not significant for Neuroticism and Agreeableness ($p > 0.05$). Only small changes are noticeable when the fixed part of the Model 4 is compared with the corresponding part of Model 3. On level 2 of the model, all the three coefficients of random parameters were significant at $p < 0.05$ level for Neuroticism, Extraversion, Openness, and Conscientiousness, while only the intercept term was significant for Agreeableness, indicating no significant interindividual differences in the rate of change in Agreeableness. The level-1 variance was further reduced for all the traits.

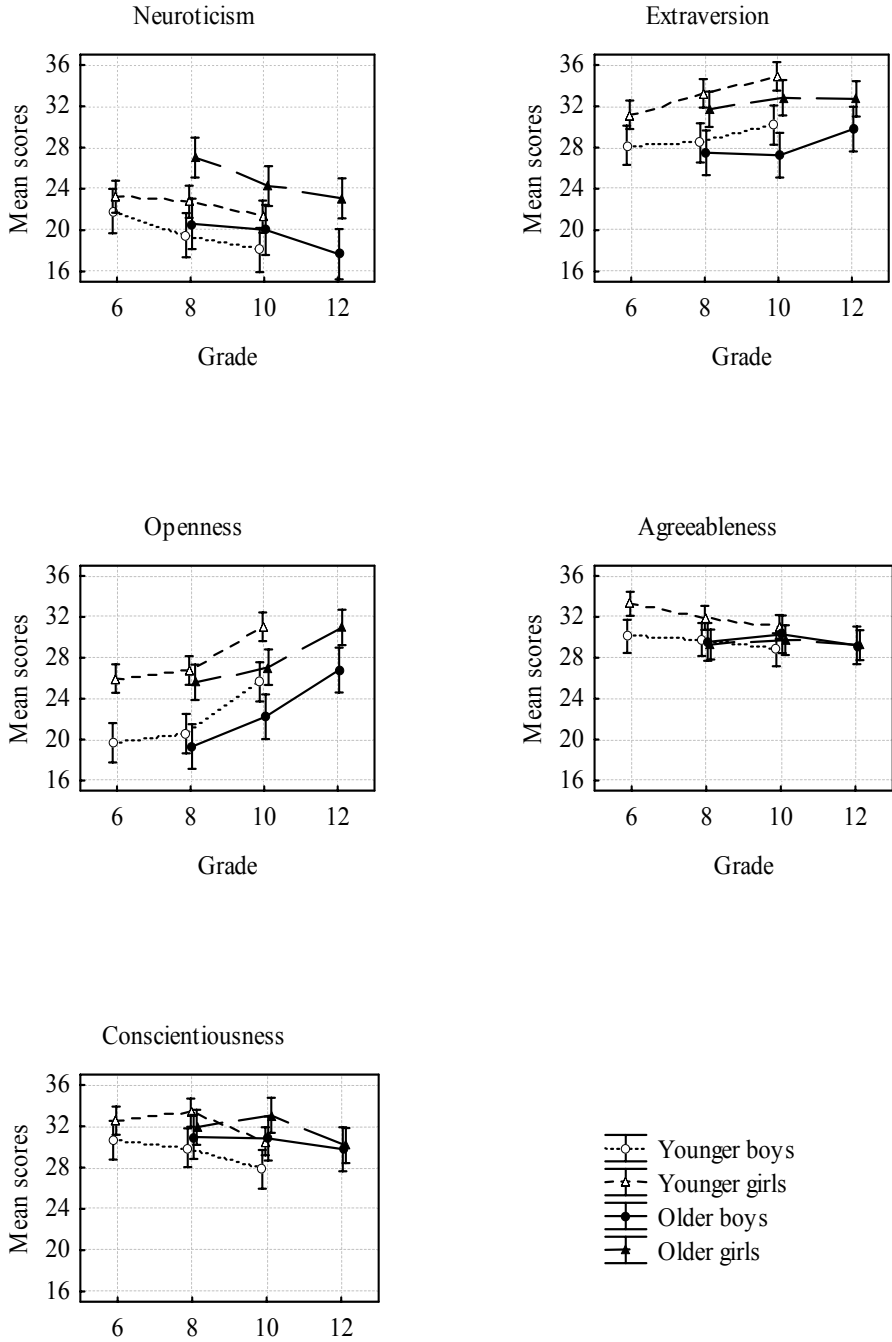


Figure 1. Mean scores by grade, sex, and cohort based on longitudinal sequences

Models which in addition to grade, sex, and cohort included sex \times grade and cohort \times grade interactions were also fitted. None of the interaction terms was significant, suggesting that there were no differences in the slopes of the growth curves between girls and boys or the two cohorts for any of the personality traits. Although the mean levels of traits were different, boys and girls and the two cohorts appeared to change in the similar way. Next, as curvilinear growth trajectories have been found for some of the traits (e.g., Branje et al., in press), models with quadratic effects of grade were tested, but none of them yielded better fit to the data than linear models.

Practice effect

When people are given the same test repeatedly, they may answer differently simply as a result of exposure to the test – this is known as a practice effect (McCrae & Costa, 2003). The design of the LSES allows to test the practice effect on the mean levels of traits by comparing the personality scores of the longitudinal sample at Time 2 and 3 with respective data from those students who were at the same grade level but had not been previously assessed. Separate *t* tests were conducted for boys and girls at different grade levels at Time 2 and 3. The following differences between “practiced” and “unpracticed” participants were significant at $p < 0.05$ level: In Wave 2, unpracticed boys in Grade 10 and unpracticed girls in Grades 8 and 10 scored lower on Conscientiousness than their practiced peers, whereas in Wave 3, the only difference was that unpracticed girls in Grade 10 scored lower on Extraversion. Regarding Conscientiousness, the difference between practiced and unpracticed participants may not reflect the effect of repeated assessment but may be caused by the bias in longitudinal sample towards higher levels of Conscientiousness. Thus, there is no clear evidence of practice effect on personality scores.

Cohort effect

A cohort effect refers to differences between generations that can be attributed to the unique experiences of each generation. The presence of a cohort effect can be inferred when there are mean-level differences between samples of equal age representing different cohorts measured at different time points (Baltes, 1968). To reveal differences in the mean trait levels measured in different occasions, analyses of variance with 3 (Wave) \times 4 (Grade) \times 2 (Sex) factorial design were performed separately for each of the five personality factors (see Table 7). Although the main effects of wave, grade, and sex were significant for almost all traits, only six of them accounted for more than 1% in the total variance. In particular, wave accounted for 1.2% of the variance in Openness, sex accounted for 1.9% of the variance in Neuroticism and 3.9% of the variance in Openness, whereas grade accounted for 1.7% of the variance in Openness and Agreeableness, as well as 1.2% of the variance in Conscientiousness. None of the interaction terms accounted for more than 0.6% of the variance, suggesting that the main effects can be generalized across waves, grades, and sexes.

Table 7. ANOVA results for cross-sectional sequences: *F* values

Source	<i>df</i>	N	E	O	A	C
Wave	2	6.93***	9.53***	21.58***	4.71**	0.43
Grade	3	0.59	6.43***	20.76***	19.27***	13.76***
Sex	1	62.25***	27.96***	143.65***	17.64***	18.56***
Wave × Grade	6	1.51	1.77	2.48*	3.52**	0.98
Wave × Sex	2	1.28	1.83	2.16	4.24*	0.80
Grade × Sex	3	5.14**	1.52	0.30	0.99	2.19
Wave × Grade × Sex	6	0.45	0.38	0.47	1.19	2.01
Error	3188					

Note. N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness.

p < 0.05; ** *p* < 0.01; *** *p* < 0.001.

These results are in line with the findings from the longitudinal analyses, confirming age-related differences in Openness, Agreeableness, and Conscientiousness. The only trait that showed a clear secular change appeared to be Openness which increased in each successive wave. Figure 2 depicts the mean levels of the five dimensions based on the time-sequential data; general similarity to Figure 1 is obvious.

DISCUSSION

This study aimed to describe quantitatively the developmental trajectories of the five major personality traits in adolescence, investigating both common and individual components of change. By using a sample of 323 adolescents who were assessed via a wide-spectrum personality measure over two time intervals long enough for true change to occur, the study followed the recommendations made by Watson (2004) for studying stability. Additionally, the presence of practice and cohort effects on personality scores was examined by the analysis of cross-sectional sequences.

The mean rank-order stability coefficients were .54 and .64 over the two 2-year intervals and .45 over the 4-year interval. The latter is very similar to the mean estimate of .43 over 6.7 years found in the meta-analysis by Roberts & DelVecchio (2000) as well as to the coefficients reported by McCrae et al. (2002) over 4 years for a sample of intellectually gifted students. Also concordantly with Roberts and DelVecchio (2000), lower levels of rank-order stability were found for younger adolescents and for longer time intervals. In addition, individual stability in personality appeared not to be stable, at least in adolescence, as adolescents' individual stability coefficients for different time intervals were only weakly related to each other.

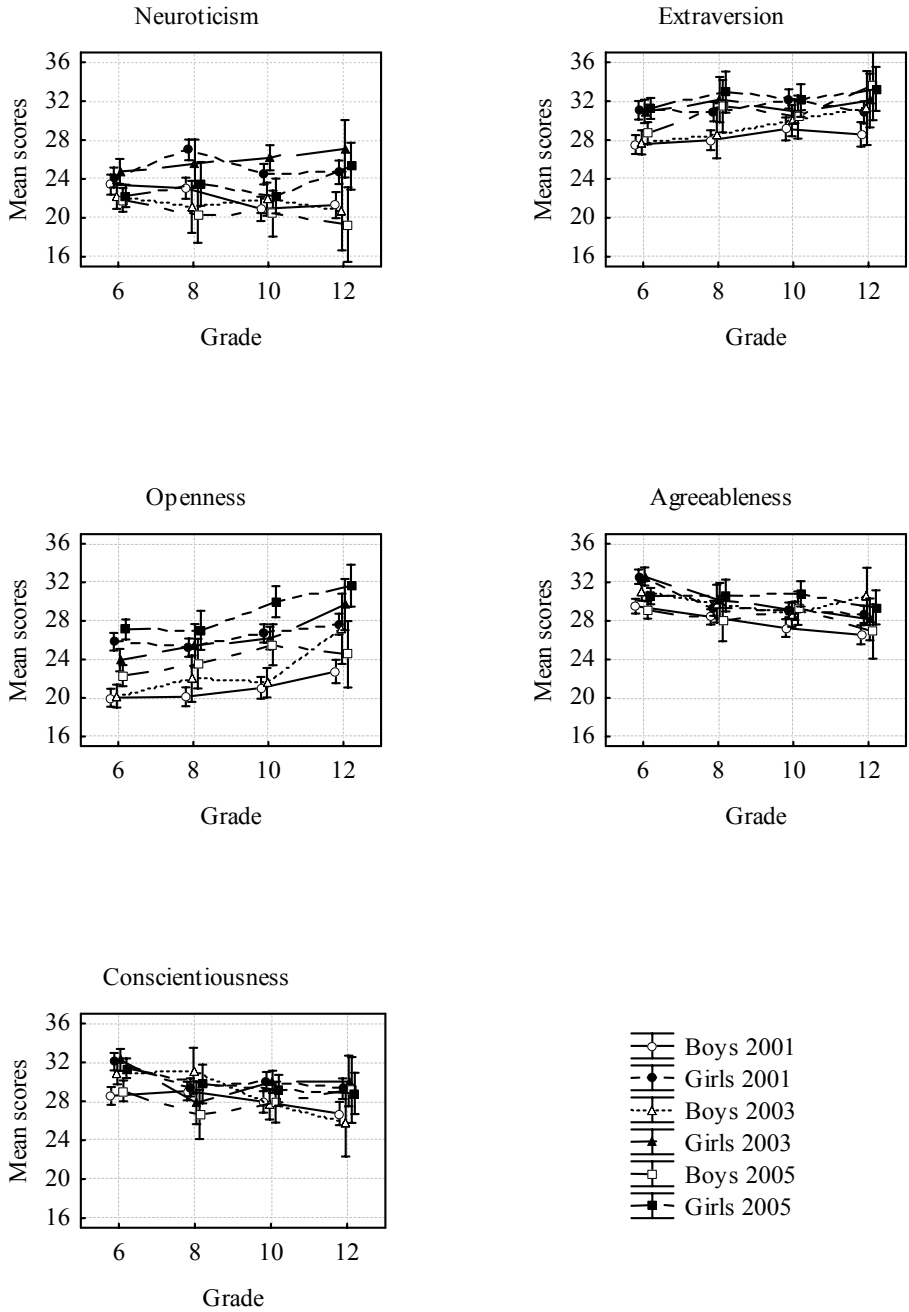


Figure 2. Mean scores by wave, grade, and sex based on cross-sectional sequences

Analyses of longitudinal and cross-sectional sequences converged in revealing a relatively big increase in Openness, as well as small declines in Agreeableness and Conscientiousness. The annual increase in Openness by 1.4 raw score points equals to the increase by 0.17 standard deviations per year, whereas annual changes in other traits remained below 0.1 standard deviations. McCrae et al. (2002) explained the rise in Openness by cognitive changes occurring during adolescence, such as an increased understanding of abstract concepts and increased reasoning abilities. Although there were sex differences and time-lag differences in the mean levels of traits, patterns of mean-level change appeared to be similar for both sexes and for different cohorts. Similarly to Baltes and Nesselroade (1972), our results suggest that sex differences have already emerged before age 12 and are not accentuated or redirected in adolescent years. The absence of sex differences in developmental trajectories agrees with the findings of Roberts et al. (2006) and McCrae et al. (2002) but is at variance with the results of Branje et al. (in press). Regarding individual differences in how personality develops, individual variance in change was significant for four traits. Nevertheless, the inclusion of the component of individual growth rate into models only slightly improved their fit.

Some evidence on the presence of practice effect on personality scores was found only for Conscientiousness, as participants with prior test experience scored generally higher on this dimension than “unpracticed” participants of the same age and measured at the same time. However, as mentioned previously, interpreting this finding as a practice effect is highly doubtful considering the fact that adolescents of the longitudinal sample scored higher on Conscientiousness than their peers already in the first measurement. Therefore, albeit small, the effects of practice and selective drop-out are probably confounded (Baltes, 1968). While the practice effect poses a problem in measuring mental abilities (e.g., Labovie, Bartsch, Nesselroade, & Baltes, 1974), it appears to be of minimal importance when personality assessment is concerned.

In many cases, adolescents at the same grade level measured in different years were less similar than adolescents at different grade levels measured in the same year, suggesting that adolescents’ average standing on personality dimensions is less dependent upon their age than upon the time (cultural moment) at which they are measured (cf. Baltes and Nesselroade, 1972). Openness was the only trait that demonstrated a clear cohort effect, with adolescents of the same age scoring higher on this dimension in each consecutive cohort. This implies that cross-sectional studies may underestimate the increase in Openness over the adolescent years because the cohort differences mask the age differences. Indeed, in Allik et al. (2004), cross-sectional differences in Openness were smaller than differences in Agreeableness and Conscientiousness. Interestingly, the cohort-related increase was found in the personality factor which is most closely related to intelligence (e.g., Harris, 2004), whereas the cohort effect on intelligence as well as the discrepancy between cross-sectional

and longitudinal results are well known (e.g., Rönnlund & Nilsson, 2006; Schaie, 1994).

One of the limitations of the study is the high attrition rate between the three waves, with less than one-third of the participants of the initial sample comprising the longitudinal sample. The high attrition was mainly caused by the design of the study, which used class as the basic unit of sampling. On the other hand, this design provided us data on previously non-participating peers, which were used for comparison. As the attrition analysis showed, girls who stayed in the study were quite similar to the girls of the initial sample, whereas boys who participated in all three waves were characterized by lower Neuroticism and higher Conscientiousness as well as higher intellectual abilities. The problem that longitudinal samples tend to be biased in a positive direction is recognized for a long time (Baltes, 1968; Schaie, 1965). The bias is caused by selective sampling and selective drop-out of participants and may challenge the generalizability of longitudinal findings as well as their comparability with cross-sectional results. Nevertheless, cross-sectional and longitudinal findings of this study generally converged with each other.

Our analyses were based exclusively on self-report data. Different developmental routes might have been found, if observer ratings were analyzed (see Branje et al., in press). For example, in an independent longitudinal study of Estonian adolescents (Laidra, Allik, Harro, Merenäkk, & Harro, 2006) where mothers, fathers, and teachers provided personality ratings of their children or pupils at age 15 and again at age 18, adolescents' personality traits were found to follow quite a different direction over the 3-year period. More particularly, teacher- and father-rated Agreeableness as well as father-rated Openness and Conscientiousness increased, while mother-rated Extraversion declined and other personality traits remained stable. Although Branje et al. (in press) also found differences between adolescents self-reports and reports by their family members in how personality changed, these differences lay rather in the magnitude of the change and not in its direction, which is the case for the two studies with Estonian adolescents. Although the use of two different personality measures in these two studies should not influence the results, the inclusion of other ratings on the same instruments would be necessary in order to explore this discrepancy.

The present study supports the general conclusion that personality change during adolescence is modest rather than dramatic for the majority of people and for most of the traits (De Fruyt et al., 2006; Donnellan et al., 2006; McCrae et al., 2002; Roberts et al., 2006). While a strong theory for predicting and interpreting mean-level changes in adolescent personality is still absent (Donnellan et al. 2006, p. 288), the accumulation of empirical studies hopefully facilitates the development of such theory.

ACKNOWLEDGMENTS

This research was supported by grants 4519 and 5677 from the Estonian Science Foundation, and by grant 0180543 from the Estonian Ministry of Science and Education to the third author. We would like to thank Anneli Veisson and Liisa Raudsepp for their help in carrying out this project. We are also grateful to Anu Realo and René Mõttus for their helpful comments on the manuscript.

REFERENCES

- Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality, 18*, 445–462.
- Asendorpf, J. B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality, 6*, 103–117.
- Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development, 11*, 145–171.
- Baltes, P. B. & Nesselroade, J. R. (1972). Cultural change and adolescent personality development: An application of longitudinal sequences. *Developmental Psychology, 7*, 244–256.
- Branje, S. J. T., Van Lieshout, C. F. M., & Gerris, J. R. M. (in press). Big Five personality development in adolescence and adulthood. *European Journal of Personality*. DOI: 10.1002/per.596
- Caspi, A. & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry, 12*, 49–66.
- Costa, P. T., Jr., & McCrae, R. R. (1982). An approach to the attribution of aging, period, and cohort effects. *Psychological Bulletin, 92*, 238–250.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Professional manual. Odessa, FL: Psychological Assessment Resources.
- De Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology, 91*, 538–522.
- Donnellan, B. M., Trzesniewski, K. H., & Robins, R. W. (2006). Personality and self-esteem development in adolescence. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 285–309). Mahwah, NJ: Lawrence Erlbaum Associates.
- Harris, J. A. (2004). Measured intelligence, achievement, openness to experience, and creativity. *Personality and Individual Differences, 36*, 913–929.
- Helson, R., Jones, C., and Kwan, V.S.Y. (2002). Personality change over 40 years of adulthood: Hierarchical linear modeling analyses of two longitudinal samples. *Journal of Personality and Social Psychology, 83*, 752–766.
- Jöreskog, K. G. and Sörbom, D. (2004). *LISREL 8.71: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

- Jöreskog, K. G., Sörbom, D., du Toit, S., & du Toit, M. (2001). *LISREL 8: New statistical features* (2nd ed.). Chicago: Scientific Software International.
- Labovie, E. W., Bartsch, T. W., Nesselroade, J. R., & Baltes, P. B. (1974). On the internal and external validity of simple longitudinal design. *Child Development, 45*, 282–290.
- Laidra, K., Allik, J., Harro, M., Merenäkk, L., & Harro, J. (2006). Agreement among adolescents, parents, and teachers on adolescent personality. *Assessment, 13*, 187–196.
- Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2002). A study of the IQ in Estonia. *Psychological Reports, 91*, 1022–1026.
- McCrae, R. R. & Costa, P. T., Jr. (2003). *Personality in Adulthood: A Five-Factor Theory Perspective* (2nd ed). New York: Guilford Press.
- 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*, 1456–1468.
- McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175–215.
- Mroczek, D. K., Almeida, D. M., Spiro, A., III, & Pafford, C. (2006). Modeling intraindividual stability and change in personality. In K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 163–180). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pullmann, H., Raudsepp, L., & Allik, J. (2006). Stability and change in adolescents' personality: A longitudinal study. *European Journal of Personality, 20*, 447–459.
- Raven, J. (1981). *Manual for Raven's Progressive Matrices and Mill Hill Vocabulary Scales*. Oxford: Oxford Psychologists Press.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3–25.
- 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*, 1–25.
- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69*, 617–640.
- Rönnlund, M. & Nilsson, L.-G. (2006). Adult life-span patterns in WAIS-R Block Design performance: Cross-sectional versus longitudinal age gradients and relations to demographic factors. *Intelligence, 34*, 63–78.
- Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64*, 92–107.
- Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49*, 304–313.
- Terracciano, A., Costa, P. T., Jr., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin, 32*, 999–1009.
- Twenge, J. (2000). The age of anxiety? Birth cohort change in anxiety and neuroticism, 1952–1993. *Journal of Personality and Social Psychology, 79*, 1007–1021.
- Twenge, J. (2001). Birth cohort changes in extraversion: a cross-temporal meta-analysis, 1966–1993. *Personality and Individual Differences, 30*, 735–748.
- Watson, D. (2004). Stability versus change, dependability versus error: issues in the assessment of personality over time. *Journal of Research in Personality, 38*, 319–350

CURRICULUM VITAE

Kaia Laidra

Citizenship: Estonian
Date of Birth: January 25, 1977
Address: Department of Psychology, University of Tartu, Tiigi 78, 50410
Tartu, Estonia
Telephone: +372 737 5916
E-mail: kaia.laidra@ut.ee

Education

1995–2000 Bachelor's studies, Department of Psychology, University of Tartu
2000–2002 Master's studies, Department of Psychology, University of Tartu
2002 *Magister scientiarum* (MSc) in psychology (*cum laude*), University of Tartu
2003–2004 Exchange student, University of Salamanca, Spain
2002–2007 Doctoral studies, Department of Psychology, University of Tartu

Professional Employment

2000–2005 Research assistant, Department of Psychology, University of Tartu
2005– Extraordinary researcher, Department of Psychology, University of Tartu

Main Research Interests

Personality development; interrater agreement on personality; academic achievement prediction; mental abilities

Membership in Professional Organizations

Union of Estonian Psychologists

CURRICULUM VITAE

Kaia Laidra

Kodakondsus: Eesti
Sünniaeg: 25. jaanuar, 1977
Aadress: Psühholoogia osakond, Tartu Ülikool, Tiigi 78, 50410 Tartu
Telefon: 737 5916
E-mail: kaia.laidra@ut.ee

Haridus

1995–2000 bakalaureuseõpe, Tartu Ülikool, psühholoogia osakond
2000–2002 magistriõpe, Tartu Ülikool, psühholoogia osakond
2002 *Magister scientiarum* (MSc) psühholoogia erialal (*cum laude*),
Tartu Ülikool
2003–2004 vahetusüliõpilane, Salamanca Ülikool, Hispaania
2002–2007 doktoriõpe, Tartu Ülikool, psühholoogia osakond

Teenistuskäik

2000–2005 laborant, Tartu Ülikool, psühholoogia osakond
2005– erakorraline teadur, Tartu Ülikool, psühholoogia osakond

Teadustöö põhisuunad

Isiksuse areng, hindajatevaheline üksmeel isiksuse hinnangutes, akadeemilise edukuse ennustamine, vaimsed võimed.

Kuulumine erialastesse organisatsioonidesse

Eesti Psühholoogide Liit