Personality Stability and Change: A Meta-Analysis of Longitudinal Studies

Wiebke Bleidorn[1]

Ted Schwaba[2]

Anqing Zheng[3]

Christopher J. Hopwood[1]

Susana S. Sosa[4]

Brent W. Roberts[3]

D. A. Briley[3]

[1]University of Zürich

[2]University of Texas at Austin

[3]University of Illinois at Urbana Champaign

[4] Loyola University Chicago

Acknowledgements: We thank Madeline Lenhausen for helping us code the personality measures.

Correspondence concerning this article should be addressed to Wiebke Bleidorn, Department of Psychology, University of Zürich, Binzmühlestrasse 14/7, 8050 Zürich, Switzerland. E-mail: wiebkebleidorn@gmail.com

This paper has been accepted for publication at ***Psychological Bulletin*** but has not undergone copyediting or formatting through the publisher. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The final article will be available, upon publication, via its DOI: 10.1037/bul0000365

**Abstract**

Past research syntheses provided evidence that personality traits are both stable and changeable throughout the lifespan. However, early meta-analytic estimates were constrained by a relatively small universe of longitudinal studies, many of which tracked personality traits in small samples over moderate time periods using measures that were only loosely related to contemporary trait models such as the Big Five. Since then, hundreds of new studies have emerged allowing for more precise estimates of personality trait stability and change across the lifespan. Here, we updated and extended previous research syntheses on personality trait development by synthesizing novel longitudinal data on rank-order stability (total $k$ = 189, total $N$ = 178,503) and mean-level change (total $k$ = 276, $N$ = 242,542) from studies published after January 1st 2005. Consistent with earlier meta-analytic findings, the rank-order stability of personality traits increased significantly throughout early life before reaching a plateau in young adulthood. These increases in stability coincide with mean-level changes in the direction of greater maturity. In contrast to previous findings, we found little evidence for increasing rank-order stabilities after age 25. Moreover, cumulative mean-level trait changes across the lifespan were slightly smaller than previously estimated. Emotional stability, however, increased consistently and more substantially across the lifespan than previously found. Moderator analyses indicated that narrow facet-level and maladaptive trait measures were less stable than broader domain and adaptive trait measures. Overall, the present findings draw a more precise picture of the lifespan development of personality traits and highlight important gaps in the personality development literature.

*Keywords:* personality development, Big Five, longitudinal, traits, meta-analysis

## Public Significance Statement

This study summarized data from hundreds of longitudinal studies to confirm that a) personality trait differences are fairly stable among adults, b) these differences tend to stabilize during adolescence and young adulthood, and c) personality tends to change in the direction of greater maturity as people age. These patterns hold across gender, nation, and ethnicity, although research from Western countries was overrepresented.

**Personality Stability and Change: A Meta-Analysis of Longitudinal Studies**

Over the past two decades, personality science has witnessed a major paradigm shift. Traditionally, traits have been viewed as highly stable and unlikely to change in adulthood (McCrae et al., 2000; James, 1950). In the 2000's, a handful of meta-analyses challenged this perspective by showing that personality traits are both enduring and open to change throughout the lifespan (Ardelt, 2000a; Ferguson, 2010; Roberts & DelVecchio, 2000; Roberts et al., 2006). The goal of the present pre-registered meta-analysis was to update and extend these works.

It would be appropriate to ask why, with such extensive prior meta-analytic work, there is a value in updating these studies. First, while prior meta-analyses reviewed a fairly large number of studies, the breadth of the existing literature, features of the data, and the analytical choices rendered many of the prior estimates to be noisier than ideal. For example, in both Roberts and DelVecchio (2000) and Roberts et al (2006), estimates were organized by age bins. When combined with the sparseness of longitudinal research in various parts of the life course (e.g., old age), this meant that estimates of stability and change were often based on only a handful of studies. Furthermore, the majority of studies included in previous meta-analyses used a broad range of measures, few of which were designed to and validated in the tradition of the Big Five taxonomy (John & Srivastava, 1999) that was used to organize measures. Finally, prior meta-analyses could not take advantage of recent advances in meta-analytic techniques that leverage information from all of the studies contained in the meta-analysis as we will explain in more detail below (Briley & Tucker-Drob, 2014; Roberts et al., 2017).

Fortunately, the rapidly growing body of research on personality trait development has led to a wealth of new and robust evidence for lifespan development of personality traits (for reviews, see Bleidorn et al., 2021; Bleidorn et al., 2020; Roberts & Yoon, 2022; Specht et al., 2014; Tucker-Drob & Briley, 2019). The availability of hundreds of new longitudinal studies provides us with the opportunity to draw a more precise picture of the development of personality traits from childhood to old age, conduct more effective tests of moderators of trait stability and change, and examine new moderators that are only now possible to test. Specifically, we aimed to answer three research questions: How rank-order stable are traits across the lifespan? How do trait levels change across the lifespan? What are moderators of rank-order stability and mean-level change in personality traits?

**How rank-order stable are personality traits?**

Traits can be defined as *relatively stable* patterns of thoughts, feelings, strivings, and behaviors that distinguish individuals from each other (Allport, 1961). Questions about the stability of traits are thus at the heart of personality science, as evidenced by multiple reviews and research syntheses on this topic (Ardelt, 2000; Bazana & Stelmack, 2004, Briley & Tucker-Drob, 2014; Ferguson, 2010; Roberts & DelVecchio, 2000; Schimmack & Anusic, 2016; Schuerger et al., 1989; Briley & Tucker-Drob, 2014). The rank-order stability of traits is typically expressed as a test-retest correlation $r$, indicating the degree to which the relative ordering of individuals on that trait is maintained across two assessments.

Virtually all longitudinal studies that have assessed personality traits more than once found that personality traits are at least somewhat stable, with rank-order stabilities typically ranging between $r$ = .40 to .60, depending on factors such as the age of the sample and the time lag between assessments. No study to date has indicated perfect stability, suggesting that personality traits remain open to rank-order change at any age across the lifespan (Bleidorn et al., 2021; Roberts & DelVecchio, 2000; Roberts & Nickel, 2017). In addition to these broad conclusions, previous research syntheses converged on three important findings while highlighting several open questions about the effects of age, time, and other moderator variables on personality rank-order stability across the lifespan. We discuss these findings and open questions next.

**Personality rank-order stability varies across the lifespan**

First, personality traits appear to increase in rank-order stability with age, particularly over the course of young adulthood. Roberts and DelVecchio (2000) found increases in stability estimates from about $r$ = .40 in early life to $r$ = .62 around age 30, and peak levels of $r$ = .75 around age 50. Ferguson (2010) reported similar results with reliability corrected estimates increasing from about $r$ = .60 in early life to $r$ = .94 by age 30, with the same level of stability in old age. This age-graded increase in rank-order stability has been often referred to as the *cumulative continuity principle of personality development* (Roberts et al., 2008). The evidence for the cumulative continuity principle appears to be robust across samples, measures, and methods (Costa et al., 2019; Ferguson, 2010; Kandler, 2010), so much so that some have referred to it as the “first law of personality development” (Roberts & Nickel, 2017, p. 161).

The finding that personality traits appear to be more prone to rank-order change early in life (especially before age 30) provides important information about the course and potential sources of personality stability and change during that life stage. In contrast, considerably less is known about the course of personality rank-order stability during middle and especially late adulthood. Past research syntheses included few participants older than 60 years (less than 5%, i.e., 6 total effects in Roberts & Del Vecchio, 2000), which made it impossible to draw conclusions about trait stability beyond age 80.

Although still a niche topic, a growing number of studies have examined the stability of personality traits in older adults over the past 20 years. These more recent studies provided mixed evidence for the late life progression of personality stability, with some studies reporting decreases in trait stability in older adulthood (Lucas & Donnellan, 2011; Wortman et al., 2012), but others indicating that stability levels remain high after age 70, at least for some trait domains (Kandler et al., 2015). With the availability of a larger number of studies that cover a wider age range, the present meta-analysis allowed us to draw a more fine-grained description of the course of personality rank-order stability, particularly for those life stages that had been only sparsely covered by previous research syntheses. The first goal of the present meta-analysis was thus to synthesize all available data on personality rank-order stability to provide a more precise description of the course of personality rank-order stability from childhood to old age.

**Personality rank-order stability decreases with increasing time intervals**

A second finding to emerge from the literature on trait stability is that rank-order correlations decrease as time intervals between assessments increase (Roberts & DelVecchio, 2000). Notably, meta-analytic evidence suggests that time-related decreases in rank-order stability are not continuous or linear. Although rank-order correlations tend to decline quickly over briefer intervals, decreases in stability appear to attenuate over longer time lags and plateau at modest values around $r = .20$ (Fraley & Roberts, 2005; Schimmack & Anusic, 2016).

This finding has important implications for the long-term stability of traits but must be considered preliminary as existing meta-analyses were constrained by the universe of available longitudinal studies, most of which tracked personality traits over moderate time periods. For example, the average lag between assessments in the Roberts and DelVecchio meta-analysis was 7 years, which was slightly inflated by a small number of studies that tracked people over more than a decade. Fraley and Roberts (2005) represented this problem as a matrix populated by meta-analytic test-retest correlations between age at baseline and age at follow-up, with nearly all of the most informative correlations for plotting the decay of stability missing. As such, we still know very little about the average rank-order stability of traits over shorter (e.g., less than 1 year) and longer time periods (e.g., 20 years).

More recent empirical studies provided novel insights into the long-term stability of traits. Following individuals over several decades, Damian et al. (2019) found rank-order stabilities around $r$ = .20 across 50 years. Covering more than 60 years, Harris and colleagues (2016) reported lower rank-order stabilities, with some approaching zero, when correlating teacher ratings of 14-year-old youths with self-reports collected when participants were 77 years old.

To refine our understanding of the association between time and stability, another goal of the present meta-analysis was to replicate and extend the findings of past meta-analyses. With the availability of a larger number of longitudinal studies that have tracked people over shorter and longer time periods, we can now probe the associations between time and trait stability to gain more precise stability estimates across varying time intervals ranging from 6 months to 51 years.

**Personality rank-order stability is robust across measures, methods, and samples**

A third finding is that little evidence exists regarding other plausible moderators of personality trait stability. Perhaps most surprising, there seem to be few differences between the different Big Five trait domains – emotional stability (versus neuroticism), extraversion, openness to experience, agreeableness, and conscientiousness. Early research on trait stability had indicated that extraversion was more stable than other trait domains (Schuerger et al., 1989). However, this effect was not replicated in more recent meta-analyses, which generally found little to no evidence for differences across trait domains (Roberts & DelVecchio, 2000). Nor was there evidence for differences across men and women or different assessment methods (e.g., self- vs. other report). Overall, these findings would suggest that rank-order stability estimates are robust and highly generalizable.

However, several issues potentially undermine this conclusion. The finding of little to no differences across trait domains relies on studies that have used a broad range of trait measures that were assigned to but not always designed to measure Big Five trait domains. In fact, few longitudinal studies in the Roberts and DelVecchio meta-analysis used instruments that were specifically designed and validated in the tradition of the Big Five taxonomy. Similarly, the use of self- vs. other-reports was confounded with the average sample age in previous meta-analyses. While other reports were typically used with children, self-report methods were more commonly used with adult samples One of the advantages of the current update to these prior meta-analyses is that many of these newly included longitudinal studies used measurement inventories explicitly designed to measure the Big Five (John, 2021). This shift in measurement practices over the last two decades will allow us to return to the test of stability of personality across Big Five domains while also examining whether the type of inventory moderated these estimates.

**Summary**

Existing meta-analytic works accumulated strong evidence that personality traits are moderately rank-order stable across the lifespan and that this stability tends to increase throughout early and middle adulthood with decreasing estimates over increasing time lags. Open questions remain about the stability of traits in middle and old age, the short-and long-term stability of traits, and the generalizability of stability findings across different trait domains, populations, and methods of assessment. With the availability of a larger number of longitudinal studies that have tracked people of different ages over varying time periods using established trait models to assess personality, we can now address these open questions and refine our understanding of the rank-order stability of traits across the lifespan.

**How do personality trait levels change across the lifespan?**

The rank-order stability of personality traits provides an important but incomplete perspective on personality trait development. Indeed, evidence for the rank-order stability of traits does not preclude the possibility that trait levels can increase or decrease over time. This possibility leads to a complementary concept in the personality development literature: the *mean-level change* of traits. Whereas rank-order stability indicates the degree to which people experience more or less change relative to one another, mean-level change reflects the degree to which trait levels decrease or increase on average in a population. Mean-level change is often expressed as standardized mean-level difference (*d*) and refers to absolute increases or decreases (gains or losses) in personality traits over a certain time.

Roberts and colleagues (2006) meta-analyzed 92 longitudinal studies of mean-level development in personality traits, covering the lifespan from age 10 to 101. They found evidence for significant mean-level change across all Big Five trait domains at some point in the life course, particularly in young adulthood but also in middle adulthood and old age. Estimates of the cumulative amount of personality mean-level change across adulthood exceeded one full standard deviation for several trait domains. This result provided evidence for the long-disputed position that personality traits continue to develop throughout adulthood, and thus had a tremendous impact on the field’s perspective on the nature and changeability of traits. In addition to probing the lifelong plasticity of personality, this meta-analysis allowed Roberts and colleagues to analyze the effects of age, time, and other moderators on mean-level change in traits across the lifespan.

**Mean-level trajectories differ across trait domains**

Compared to the seemingly universal increase in rank-order stability described above, the age-graded patterns of mean-level trait change were more complex. Although mean-level changes in traits appeared to be generally most pronounced in young adulthood, the trajectories were markedly different across different trait domains.

Specifically, Roberts and colleagues (2006) found evidence for steady and significant increases in emotional stability, conscientiousness, and, to a lesser degree, also in agreeableness throughout the adult lifespan. This pattern - now referred to as the *maturity principle of personality development* (Roberts & Nickel, 2017; Specht et al., 2014) - has since been replicated in large-scale cross-sectional (Soto et al., 2008; 2011) and longitudinal data (Lucas & Donnellan, 2011; Specht et al., 2011), across different cultures (Bleidorn et al., 2013) and trait measures (Graham et al., 2020). Although increases in emotional stability, agreeableness, and conscientiousness tend to be most pronounced during young adulthood, more recent studies found similar increases in maturity-related traits in samples of adolescents (Borghuis et al., 2017) and middle-aged adults (Schwaba et al., 2021), indicating a general age-graded trend towards greater psychological maturity (Bleidorn, 2015; Roberts & Mroczek, 2008).

In contrast to the well-established maturity principle, which applies to emotional stability, agreeableness, and conscientiousness, the lifespan trajectories of openness and extraversion, the two other Big Five traits, are less clear. Initial meta-analytic evidence indicated a curvilinear trajectory for openness with small gains in adolescence and young adulthood and similarly small decreases in older age. More recent studies replicated the age-graded gains in openness in young adults (Luedtke et al., 2011; Schwaba et al., 2019). However, findings for middle and late adulthood were more mixed, with some indicating continuous increases (Mueller et al., 2016), and others suggesting progressive decreases, especially in old age (Schwaba et al., 2018).

A possible explanation for this mixed pattern of results involves differences in the content of established openness measures, with some emphasizing intellect and others focusing more on open-mindedness and unconventionality. Similarly, there was little to no meta-analytic evidence for mean-level changes in the broad domain of extraversion in Roberts et al. (2006). However, a different picture emerged when the extraversion measures were organized according to the subdomains of social vitality (e.g., gregariousness) and social dominance (e.g., assertiveness). Again, changes in these traits – especially in social dominance – were most pronounced during young adulthood and least pronounced during middle age. Notably, the longitudinal database available at that time was limited in several important ways. Few longitudinal studies included established Big Five measures. Moreover, a disproportionate number of longitudinal studies were based on younger samples, rendering the mean-level development in middle-aged and older adults less reliable than ideal.

**Personality mean-level change varies across the lifespan**

In the present meta-analysis, we aim to provide a more fine-grained description of the age-graded mean-level changes in personality traits throughout the lifespan, particularly across older adulthood, which has been sparsely covered in previous research syntheses. Theory and some initial research suggest that the trends of personality maturation observed for young and middle adulthood may revert in older adulthood (Wagner et al., 2016).

Lifespan theories of aging (Baltes & Baltes, 1990; Freund & Baltes, 2002; Rowe & Kahn, 2015) argue that an increasing ratio of losses vs. gains provides the context for psychological changes in late adulthood. For example, decreases in health and cognitive functioning may undermine an older adult's capacity to maintain stability in their lifestyle and environment, potentially initiating decreases in personality traits that are related to psychological maturity such as emotional stability or conscientiousness.

In recent years, a growing number of longitudinal studies have examined personality mean-level changes in samples of older adults. This literature, however, has provided mixed results, with some studies reporting age-graded mean-level decreases in traits such as emotional stability or conscientiousness (Kandler et al., 2015; Mueller et al., 2016) and others indicating little to no mean-level change in personality traits in older adulthood (e.g., Kuzma et al., 2011). Mõttus et al. (2012) found evidence consistent with both trends. In a sample tracked from age 81 to 87, mean-level decreases of approximately a quarter of standard deviation were found, but in a sample tracked from age 69 to 72, almost no mean-level change was found. These findings suggest that patterns of mean-level change may be quite sensitive to age, pointing to the need for a highly powered meta-analysis and more studies of old age. In synthesizing the existing literature on lifespan personality changes, our goal was to provide more precise estimates of both rank-order stability and mean-level change in personality in old adulthood.

**Personality trait change increases with increasing time intervals**

Another important finding to emerge from the Roberts et al. (2006) meta-analysis concerns the role of time. Analogous to the findings for personality rank-order stability, longer time lags appear to be associated with more mean-level change, at least for certain trait domains. The positive link between time and change provides some evidence to suggest that mean-level trait changes may be lasting. Historically, personality traits have been often conceptualized as metabolic set points. That is, people were thought to fluctuate around their biologically predisposed trait levels in response to certain experiences or events, but eventually return to their personal set point (e.g., Ormel et al., 2017). Strict set-point models would imply a negative or null association between time and personality mean-level change, because any change would represent short-term fluctuations that disappear as people drift back to their genetically predisposed set point. The finding that time is positively associated with mean-level change speaks against such a strict set-point model and provides initial evidence for lasting trait change.

Notably, the meta-analytic evidence for the positive link between time and change must be considered against the backdrop of the longitudinal data available at the time. As mentioned above, most longitudinal studies tracked personality traits over moderate time periods. The average lag between assessments in the Roberts et al. meta-analysis was 9 years, with few studies that tracked personality over longer time periods (e.g., 20 years). Another goal of the present meta-analysis was thus to analyze the associations between time and mean-level change in a larger longitudinal database including studies with varying time intervals in order to replicate past results and refine our understanding of the association between time and mean-level personality change.

## Personality mean-level change is robust across measures, samples and methods

There is little evidence for moderators of mean-level change other than trait domain, age, and time. Roberts and colleagues (2006) tested the effects of gender, attrition, and birth cohort on change in personality traits. While there were no significant effects of gender and attrition; there were some effects of birth cohort. These effects, however, were strongly correlated with age effects and thus difficult to interpret. To address this issue, they tested and found some cohort effects within the group of young adults – the age group that demonstrated the largest mean-level change. Specifically, younger cohorts appeared to increase more in their social dominance, agreeableness, and conscientiousness than older cohorts. However, these effects need to be replicated in a larger sample of longitudinal studies to disentangle age from cohort effects. In this present meta-analysis, we aimed to provide a more comprehensive and statistically well-powered test of moderator effects on personality development.

#### Summary

Existing meta-analytic evidence indicates that personality traits continue to develop throughout the lifespan, with more pronounced mean-level changes across longer time intervals. There is a strong signal for increases in trait levels that reflect greater maturity, particularly during young adulthood. However, comparatively less is known about the normative trajectories of openness and extraversion, about mean-level development in middle and old adulthood, and the generalizability of findings across samples from different populations and methods of assessment.

### Additional moderators of personality rank-order stability and mean-level change

In addition to moderators that have been tested in previous research syntheses, the availability of a larger number of studies allowed us to explore the effects of novel and hitherto untested moderator variables. These novel moderator tests are crucial for evaluating the robustness and generalizability of the meta-analytic findings.

#### Publication year

Older and more recently published studies may differ in important ways. For example, modern standards for analyzing and reporting data have introduced important changes in how researchers treat missing data and report results (Aczel et al., 2019). Newer studies may thus be more likely to report all available data in greater transparency. We examined the effects of publication year and contrasted studies published before and after Roberts and colleagues' most recent meta-analysis (2006).

**Sample characteristics**

The increased recognition of personality traits as dynamic variables has led to a noticeable increase of longitudinal research, including large-scale and nationally representative samples from different cultures. The availability of a larger number of samples allowed us to explore the potential effects of additional sample characteristics. Specifically, we tested whether there are differences between nationally representative panels such as the German Socioeconomic Panel (GSOEP, Wagner et al., 2007) and convenience samples.

In addition, we aimed to explore the potential effects of country and ethnicity. Like most psychological research, the majority of research on personality development has focused on White samples from Western, educated, industrial, rich, and democratic (WEIRD, Henrich et al., 2010). As such, very little is known about ethnic or cultural differences in personality trait stability or mean-level change. To begin to address this issue, we aimed to examine differences in the rank-order stability and mean-level change of personality traits across different ethnic groups and countries.

**Measurement properties**

Although researchers seem to prefer certain popular self-report instruments, there still is tremendous variety in the types, content, and quality of measures used to assess personality traits. Corresponding differences in measurement properties may have introduced systematic variability in the literature on personality rank-order stability and mean-level change. Here we tested four potential moderators and additionally explored the role of measurement invariance testing.

First, we tested differences between measures that were specifically designed and validated in the tradition of the Big Five or Five-Factor taxonomy vs. other measures. Second, we explored differences between traits measured as broad domains and traits measured as narrow facets. Third, we explored differences between adaptive and maladaptive trait measures. Longitudinal research in clinical samples suggests that maladaptive traits may be less stable than normal range traits (Hopwood & Bleidorn, 2018; Schuerger et al., 1989). Here, we examined whether these differences generalize to normal and maladaptive traits as measured in non-clinical samples. In addition to the pre-registered moderators, we tested whether differences in measurement unreliability (as indicated by a scale's internal consistency) were associated with estimates of rank-order stability and mean-level change. Finally, we coded whether studies included measurement invariance tests and recorded the results of these tests. Measurement invariance across assessment wave is a necessary condition to meaningfully interpret estimates of rank-order stability and mean-level change (Vandenberg & Lance, 2000).

**Overview and hypotheses**

Past research provided strong evidence that personality traits are both stable and changeable throughout the lifespan. Rank-order stability appears to be highest during middle adulthood and lowest during young adulthood. Mean-level change, on the other hand, appears to be most pronounced during young adulthood. The average direction of change is clearly positive, as most people increase in trait levels that reflect greater psychological maturity. While time is positively related to change and negatively related to stability in personality traits, there is little evidence for systematic influences of other moderators on either rank-order stability or mean-level change. These findings would appear to provide a solid foundation for scholars to build their understanding of the nature and mechanisms of personality development upon (e.g., Roberts & Nickel, 2017; Bleidorn et al., 2020; Tucker-Drob & Briley, 2019). However, several important answers to questions about the effects of age, time, and other moderators on rank-order stability and mean-level change in personality traits remain provisional at best, given the small number of longitudinal studies informing estimates in the past meta-analyses.

The purpose of the present pre-registered meta-analysis was to synthesize all available data to provide more conclusive answers to these questions and identify remaining gaps in current research on personality trait development. As mentioned above, we organized our review in reference to the Big Five taxonomy. Although some of the trait measures studied here were not originally conceptualized within the framework of the Big Five, synthesizing these traits into the dominant paradigmatic model for personality psychology allowed us to communicate findings across numerous personality trait measures and facilitated comparisons with other research.

Consistent with theory (Fraley & Roberts, 2005) and previous research syntheses (Roberts & DelVecchio, 2000; Ferguson, 2010), we expected the average rank-order stability of traits to range between .50-.60, with considerable heterogeneity across studies (H1). We expected the rank-order stability of traits to increase with age [holding time constant] (H2a), to peak in late adulthood (after age 65, H2b), to decrease in old age (after age 80, H2c), but to never reach unity at any age. We expected that the rank-order stability of traits decreases with increasing time between assessments (H3a) but never =/<.20 regardless of the length of time lag (H3b). We further explored whether there were interaction effects between age and time lag on rank-order stability. We expected no meaningful differences in rank-order stability across genders (H4a), Big Five domains (H4b), or between self- vs. other report instruments (H4c).

For mean-level trait changes, we predicted these to be most pronounced in young adulthood (~age 18-40 years, H5a) and least pronounced in middle adulthood (~41-65 years, H5b). We expected that rates of mean-level change increase with increasing time lags between assessment waves (H6a) and explored interactions between effects of age and time lag on rates of mean-level change. We expected rates of mean-level change to vary across trait domains (H7), with more pronounced changes in emotional stability, agreeableness, and conscientiousness than in openness and extraversion. We expected no differences in mean-level change across genders (H8a) or self- vs. other reports (H8b), and birth cohorts (H8c). Finally, we explored the moderating effects of study features (publication year), sample characteristics (nationally representative vs. other, ethnicity, country), and measurement properties (Big Five vs. other, narrow vs. broad, maladaptive vs. normal, internal consistency).

## Method

We first reanalyzed the data of two previous meta-analyses on rank-order stability (Roberts & DelVecchio, 2000, $k$ = 152, $N$ = 55,180) and mean-level change (Roberts et al., 2006, $k$ = 92, $N$ = 50,120) using contemporary meta-analytic techniques. Details about the literature searches and data aggregation procedures for these databases can be found in the original publications. Importantly, we applied the current inclusion and exclusion criteria (see below) to the archival data which resulted in fewer studies and smaller sample sizes for both stability ($k$ = 67, $N$ = 29,651) and change ($k$ = 84, $N$ = 47,235). We then synthesized the longitudinal data on personality rank-order stability and mean-level change from studies published after the completion of these meta-analyses (i.e., after January 1st 2005), and finally merged and meta-analyzed all available data. Here, we report the search terms and databases used for identifying individual studies published after January 1, 2005, inclusion and exclusion criteria, the extraction of data and coding of effect sizes, and our statistical approaches to meta-analyze the data.

The meta-analytic strategy for this review was pre-registered at https://osf.io/ucqwd. We followed the PRISMA-P checklist when preparing the protocol, and we followed PRISMA reporting guidelines for the final report. The meta-analytic analysis code is available at https://osf.io/6wjnf/. The meta-analytic data are shared at https://osf.io/gfbjs/.

**Literature searches**

We performed an abstract search of PsycINFO for studies that included any combination of terms from two categories: personality (personality, trait, temperament) and longitudinal (test-retest, longitudinal). We restricted our search to quantitative studies (including dissertations) of human populations published in English after January 1st 2005. This approach produced a total of 4,905 potentially relevant articles (3,829 articles in a first search in January 2017 and 1,076 articles in a search update in January 2020).

We included studies if they fulfilled the following criteria. First, the study used a longitudinal design (i.e., at least two assessments of the same sample). Second, the test-retest intervals were greater than or equal to 6 months. Third, the study included a trait measure (i.e., enduring, cross-situational consistency). Consistent with Roberts et al. (2006), we excluded measures of attitudes, values, self-esteem, affect, mood, intelligence, cognitive functioning, sex role, and validity scales; however, we included measures of temperament. Fourth, the trait measure was identical across assessment waves (in terms of number of items, wording, item content, and response scale). Fifth, the sample was nonclinical and was not the focus of an intervention. Sixth, the sample was sufficiently homogeneous with regard to age, as operationalized by a cutoff value of $SD \leq 3$ years for age at baseline. Seventh, the study contained sufficient information to compute effect sizes. When relevant effect size information was missing, we contacted the authors of the original studies by email and requested the data. In 16% of these cases ($k = 23$), authors provided usable additional data that was included in the meta-analysis.

Applying these criteria, we identified 205 studies. We summarize this process in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (Figure 1) and describe it in detail in the pre-registration (https://osf.io/ucqwd).[1] We identified multiple articles that used the same data or similar, updated data. When this was the case, we removed redundant articles and kept the publication with the most time points or the most measures (see, Briley & Tucker-Drob, 2014). To test whether effect sizes differed for broad vs. narrow trait measures, we included studies that used the same data but reported the results at different levels of trait generality (e.g., Prinzie & Deković, 2008 reported both domain and facets of the *Hierarchical Personality Inventory for Children*). Because many of these studies reported data from several samples, the number of samples, 250, was greater than the total number of studies. The final dataset included 3,598 rows corresponding to unique pairs of time points and measures. We divided this new dataset into three subsets for analytic purposes. These subsets were 1) self-report test-retest stability effect sizes (2,213 effect sizes from 122 studies representing a total sample size of $N$ = 148,922 participants), 2) other-report test-retest stability effect sizes (689 effect sizes from 61 studies representing a total sample size of $N$ = 51,485 participants), and 3) mean-level change effect sizes (3,442 effect sizes from 192 studies representing a total of sample size of $N$ = 233,510 participants). A full list of these studies is provided in the supplementary online materials (SOM, Table S1).

[1] Please note that we initially identified 207 studies but had to exclude two studies after the pre-registration that did not meet our study inclusion criteria (see Figure 1).

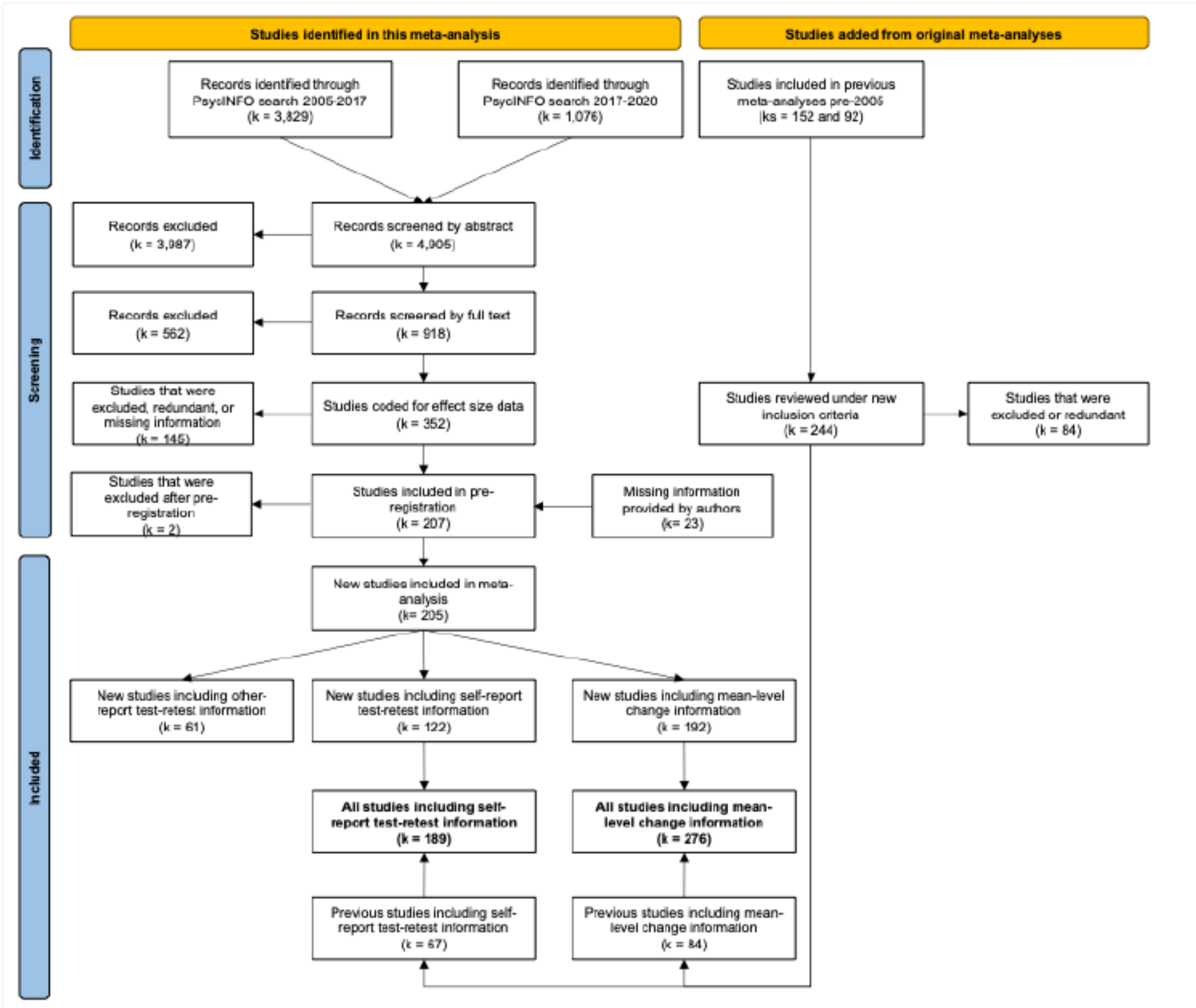


Figure 1. PRISMA flow diagram.

**Coding of study variables and effect size information**

We developed a detailed codebook (https://osf.io/j3x54/) for recording the relevant study and sample characteristics, the personality variables, and effect size information. Each sample from every usable study was coded for several study variables (e.g., publication year, sample size, percentage of females in the sample) and effect size information (*M, SD,* test-retest correlation). If studies provided information that allowed us to code independent subsamples (e.g., different cohorts or age groups), we coded subsamples rather than the full sample to increase the precision of analyses. Of the initially 352 included studies, 222 were coded by two coders and 25 by three coders. Initial estimates of interrater agreement for 25 randomly selected studies indicated perfect interrater agreement (ICC = 1) for all study variables and effect size information, except for ratings of Big Five trait category. To address this problem, we installed an evidence-based classification strategy and assigned traits to Big Five domains using published correlations (for a similar approach, see Stanek & Ones, 2018, see below for more details about the assignment of measures to Big Five trait domains).

We recorded the average *age* in years of participants at each assessment wave. A few studies reported a range of ages (e.g., 20 to 30, 30 to 40). For these studies, the midpoints of the reported age ranges were used as estimates of age. When studies did not report age directly but valid indicators of age were given, we used this information to estimate age (see Orth et al., 2018). For example, if a study reported that participants were children in kindergarten, we estimated the mean age of participants as 5 years (= mean of 4 to 6 years). When studies reported age only for time 1 but not for time 2 or later assessments, we estimated age (time 2) = age (time 1) + time lag between time 1 and time 2. We computed the *lag* between assessments by subtracting age in years at time 1 from age in years at time 2.

We empirically assigned each scale to a *Big Five domain*. Personality scales that did not directly assess the Big Five (e.g., the 18 scales in the California Personality Inventory) were sorted into corresponding Big Five categories based on studies that examined the correlations between these personality scales and established Big Five measures. When this was not possible, we sorted scales into corresponding Big Five categories on theoretical grounds or using information from similar scales (for details and scale correlations, see codebook in the SOM). We coded personality scales that were most strongly correlated with multiple Big Five traits (within $r = .05$ of the strongest correlation) as "blended" to reflect the scale's association with multiple "mature" (i.e., positively evaluated) personality scores, such as a scale blending agreeableness and conscientiousness. To ensure that the direction of blended traits was consistent, we coded all effect sizes such that positive values would reflect the expected maturation trends. For example, we would code a trait that was a blend of high conscientiousness and agreeableness as a positive blend and would code a trait that was a blend of low conscientiousness and disagreeableness as a negative blend. The negative blends were reflected such that higher scores would reflect lower levels of the positive pole, similar to coding for the other Big Five (e.g., scores on neuroticism were reflected to indicate emotional stability). In situations in which a trait reflected a blend of mature and immature traits (e.g., low neuroticism and conscientiousness, such as callous-unemotional behaviors), we coded these traits as contrasts. This final catch-all category is difficult to interpret because of the wide variety of measures included, but assigning relevant effect sizes to this category assured that the primary Big Five codes were not contaminated by potentially mismatched scales.

We coded the *country* in which the data was collected as well as the *percentage of female*, *ethnicity* and *national representativeness* of the sample. To record ethnicity, we coded the percentage of Asian, Black, Latino/Hispanic, Native American, and other ethnicities in studies that reported the ethnic composition of the sample. As very few studies reported this information (approximately 20%), we did not analyze effects of ethnicity. A failure to report the ethnic compositions of samples continues to be a problem for meta-analyses. We coded effect sizes from nationally representative samples such as GSOEP or HILDA as representative and all other effect sizes as based on convenience samples.

We further coded several properties of the measures. Effect sizes based on target reports were coded as *self-report*, effect sizes based on data from observers, parents, informants, or anyone other than the target were coded as *other-report*. A total of 31 effect sizes were drawn from a combination of self-report and other-report. These effect sizes were treated as other-report. We classified effect sizes based on Big Five domain scales or broader as *broad measures* and effect sizes based on aspect, facet, or more specific scales as *narrow measures*. For example, we coded the five domain scales of the Big Five Inventory-2 (Soto & John, 2017) as broad measures and its 15 facet scales (e.g., achievement, control, harm avoidance, etc.) as narrow measures. Trait measures that were reported independent of a general taxonomy (e.g., shyness) were coded as narrow measures. We classified effect sizes from measures designed in the tradition of the Five Factor Model or the Big Five (e.g., BFI-2; NEO-PI-R, TIPI) as *Big Five measures*. Trait measures that were developed independent of the Big Five taxonomy were coded as *non-Big Five* measures. Measures were coded as *maladaptive* if they contained content that is indicative of personality problems (e.g., internalizing / externalizing behavior in children; conduct problems, mood problems); all other scales were coded as *adaptive*.

We further coded two indicators of measurement quality that were requested in review. First, we coded and tested the moderating effects of the measures' internal consistency (i.e., Cronbach's alpha) as an indicator of measurement reliability. Second, we coded whether studies tested for measurement invariance and the results of these tests (no measurement invariance, metric, or scalar measurement invariance). Metric invariance (i.e., the invariance of factor loadings across assessments) is a necessary precondition for the interpretation of rank-order stability coefficients, and scalar invariance (i.e., the invariance of item intercepts) is a necessary condition for interpreting estimates of mean-level change across assessments (Vandenberg & Lance, 2000).

**Calculation of effect sizes and standard errors**

We meta-analyzed correlation coefficients *r* and Cohen's *d* using the single-group, pre-test-post-test raw score metric (Morris & DeShon, 2002). To calculate the sampling variance for the effect sizes, we used standard formulas and used the smallest sample size for the pair of time points. The sampling variance for the correlation coefficient is estimated by

$$Var = \frac{(1-r\ 2)2}{n-1}$$

The sampling variance for Cohen's *d* is more complex and reported clearly in Morris and DeShon's (2002). Importantly, this formula requires information about test-retest stability. When available, we used the reported test-retest correlation. In order to minimize the impact of missing data, we used a model-based imputation approach. Specifically, we used the best fitting model predicting test-retest stability from age at baseline and time lag between assessments to estimate the expected test-retest stability when missing. Previous work found minimal differences in results when assuming different values of test-retest stability for this purpose (i.e., assuming .3, .5, or .7; Roberts et al., 2006). Therefore, we chose to implement a straightforward imputation strategy.

**Analytic strategy**

Prior to our main analyses, we evaluated the potential for publication bias in two ways. First, we qualitatively inspected funnel plots. Funnel plots display the association between effect size and precision. When there is no publication bias, effect sizes should form a symmetrical funnel around the true population effect size. Large sample size studies will form a tight distribution around the true effect size, and low sample size studies will form a wide, but importantly, symmetrical distribution around the true effect size. An asymmetric funnel would emerge if only studies with significant results were published.

Second, we included the sampling variance as a predictor in our meta-regression models. This approach is called the precision-effect estimate with standard error (PEESE, Stanley & Doucouliagos, 2014). By including the sampling variance in the model, we tested whether less precise studies tended to have larger effect sizes. If publication bias is a problem, then less precise studies would have larger effect sizes because those are the only effects they are powered to detect. If publication bias is a problem, then less precise studies would have larger effect sizes because those are the only effects they are powered to detect. In addition to the regression coefficient, the intercept of these models takes on special meaning. Conceptually, the intercept reflects the effect size estimate when the sampling variance is zero, meaning a study with infinite sample size. Of course, this estimate extrapolates from the data as no such study exists, but it provides a less biased estimate of the true effect size after taking into account the potential effects of publication bias.

To test our main hypotheses, we used random-effect meta-analytic structural equation modeling (Cheung, 2008). All models were estimated using M*plus* version 8 (Muthen & Muthen, 2011-2020). We weighted each model by the inverse of the sampling variance and the number of effect sizes included per sample. The first component of our weighting variable, the sampling variance, weights for the precision of the estimate with more precise (i.e., larger sample size) studies carrying more weight in the analysis. The second component of our weighting variable, the number of effect sizes included per sample, ensures that samples that contribute many effect sizes (e.g., 30 facets vs. 5 scale scores) do not receive undue influence on the results. In effect, this weighting scheme ensures that each sample contributes equal weight in the analysis, holding sample size constant. Additionally, we used cluster robust standard errors at the level of the sample due to the non-independence of effect sizes derived from the same sample (McNeish et al., 2017), a technique that is similar to robust variance estimation (e.g., Hedges et al., 2010).

In each meta-regression model, we estimated at least two parameters: the weighted average effect size and $\tau$, an estimate of between-study heterogeneity. Conceptually, $\tau$ reflects the spread around the weighted average effect size after taking into account sampling variability, much like the standard deviation reflects spread around the mean of a variable. Next, we added moderators to statistically account for between-study heterogeneity using meta-regression models. To evaluate the importance of the moderators, we used the meta-regression coefficient and the change in between-study heterogeneity ($\tau$). If a moderator is important, we should estimate a meaningful regression coefficient (i.e., practically important when plotted across the lifespan or in context of other moderators), and we should find that between-study heterogeneity has decreased. We interpret the change in $\tau$ across models similar to $R^2$. For example, if in a baseline model the between-study heterogeneity is $\tau = .50$ and in a model that includes a moderator the between-study heterogeneity is $\tau = .30$, then this moderator has statistically accounted for 40% of the between-study heterogeneity.

Because a major goal of the project was to better document lifespan trends, we evaluated many functional forms connecting age and time lag with test-retest stability and mean-level change. To flexibly model age-trends, we used a computationally intensive specification search using spline regression (for a similar strategy, see Briley & Tucker-Drob, 2014). We created 5-year age bins to restructure the continuous age variable. For example, an age of 7 would result in a score of 5 for the first age bin and a score of 2 for the second age bin. If one were to sum the age bins, the original continuous age variable would be found. The last age bin was for values greater than 75 as data coverage for additional bins was sparse. To identify the best fitting model, we tested all possible combinations of fixing and freeing the 16 age bins. We selected the model that had the lowest BIC as the best fitting model, and we additionally inspected the top 10 best fitting models to evaluate how the best fitting model deviates from nearly-equivalent models. Centered time lag was included in all models as a predictor.

As we expected age and time lag to account for a substantial portion of between-study heterogeneity, we tested the other moderators in the context of the best fitting spline model. This approach ensured that we did not mistakenly identify a moderator which was simply correlated with the age of the sample. One exception was trait category, which was included in three specifications – in isolation, including age and lag as covariates, and including interactions between age bins. The final specification allowed us to plot trait-specific lifespan trends for each trait domain.

To complement the spline approach, we also estimated models specifying an exponential functional form connecting age and lag to stability and change. These models are particularly useful for time lag as it is unlikely that the effect of lag follows a linear function (Fraley & Roberts, 2005; see also, Card, 2019; Kuiper & Ryan, 2019)). Instead, the exponential specification allows the decay of stability to reach some lower asymptote. We additionally explored the dependency between age and lag with the expectation that stability decays more rapidly at younger ages, a trend observed for cognitive ability (Tucker-Drob & Briley, 2014).

We followed our pre-registered analytic plan with two exceptions. First, we ran additional moderator analyses including internal consistency as an indicator of measurement reliability. Second, we intended to test the method of report (self- or other-report) as a moderator of changes in rank-order stability across the life span. However, other-report effect sizes were heavily concentrated at younger ages (max age at baseline = 17 years), making it impossible to clearly disentangle age-trends from report effects. Stability estimates from other-reports tend to be much larger relative to self-reports at younger ages, likely for a combination of reasons (e.g., implicit models of personality; lack of access to internal thoughts and feelings). For this reason, we split the test-retest stability dataset into a self-report dataset and an other-report dataset. Here, we focus on self-report test-retest stability estimates as age was strongly related to report format. Analytically, this decision forced us to collapse across the age 0-5 and age 5-10 bins for the self-report stability analyses given the lack of variability in the age 0-5 bin. For similar analyses conducted on the limited informant-report data, please see the online supplement. Generally, stability was relatively high for informant-reports (average effect size = .545), increased with age in very early life at a rate of .045 per year from ages 0 to 5 years, plateaued in stability following age 5, and was lower at longer time lags (b = -.026).

## Results

All analytic code, model output, and data are available on the OSF project page https://osf.io/j3x54/. Because the project spans many analyses and robustness tests, we focus here on the most precise estimates from the best-fitting models based on all available data.

Specifically, for the lifespan trends in rank-order stability and mean-level change across trait domains, the results are based on the merged dataset containing all available effect sizes. For the remaining moderator analyses, we report analyses based only on the newly coded effect sizes from studies published after 2005 as the coding of these variables was not consistent with the previous meta-analyses. Funnel plots and models incorporating sampling variance as a predictor are available online. The OSF page includes a report documenting each step of the analyses across all datasets.

**Publication bias**

The qualitative inspection of funnel plots provided no evidence for publication bias in the dataset All funnel plots are depicted in Figures S1-S4, S7, and S13. By including the sampling variance in the model (PEESE correction), we also provided a quantitative test of publication bias. Using this quantitative approach, we again found no evidence for publication bias. The regression coefficient for the sampling variance tended to be negative, indicating that less precise studies tended to have smaller effect size estimates, and all interpretations of the trends are unaffected by using the adjusted intercept. The results of these PEESE-corrected models are presented in Chapter 3.2 of the SOM. The present findings are consistent with previous meta-analyses in the area, suggesting that there is no strong press for significant effects in either direction. In other words, the results suggest that publication does not depend on the size or significance of a personality rank-order stability or mean-level change effect.

**Descriptive statistics**

Table 1 reports the descriptive statistics for continuously coded variables for stability and change studies in the novel data; Table 2 for the merged novel and original data. A comparison of the effect sizes listed in Table 2 and 3 indicates that of the 3,578 effect sizes in the self-report rank-order stability meta-analysis, 38% (NES=1,365) were derived from studies analyzed by Roberts and Del Vecchio (2000) and 62% from papers published after 2005. Of the 5,034 effect sizes in the mean-level change meta-analysis, 32% (NES=1,592) were drawn from Roberts et al. (2006) and 68% from papers published after 2005. Table 3 presents frequencies for the categorical coded variables in the novel data; Table 4 for the merged databases (for a complete list of all variables for all possible subsets of the data, see the dataset description in Chapter 1 of the SOM). Roughly the full lifespan was represented in the data (age at baseline range = 0.25 years to 100.5 years). Considerable variation was also observed for time lag (range = 0.5 to 51 years).

To explore the role of measurement invariance testing, we further recorded the number of novel studies that tested for metric or scalar invariance and the results of these tests. Overall, 29% of the novel studies (k = 59 of 205) tested for measurement invariance. Of those, k = 18 studies examined metric invariance, and k = 39 examined scalar invariance. All studies but two established scalar or metric measurement invariance (see Table S1). The results further indicated that, while MI testing was rarely reported in earlier studies, it has become common practice in recent years. Notably, a possible explanation for the relatively small number of studies that did report measurement invariance results is that several of the studies included in this meta-analysis assessed personality traits as auxiliary rather than main variables.

Table 1. Descriptive statistics for continuously coded variables in novel data published after 2005.

| | Rank-order stability | | | | | Mean-level change | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | *SD* | Min | Max | NES | Mean | *SD* | Min | Max | NES |
| Age at Baseline | 22.851 | 15.286 | 5.85 | 80 | 2213 | 16.591 | 14.849 | 0.25 | 81 | 3442 |
| Age at Follow-up | 27.393 | 18.088 | 6.63 | 84 | 2213 | 20.175 | 17.037 | 1 | 86.61 | 3442 |
| Time lag between assessment | 4.541 | 6.419 | 0.5 | 51 | 2213 | 3.584 | 5.213 | 0.5 | 51 | 3442 |
| %White | 50.476 | 33.367 | 0 | 100 | 439 | 68.109 | 32.286 | 0 | 100 | 908 |
| %Black | 15.179 | 14.208 | 0 | 58 | 362 | 10.885 | 13.607 | 0 | 100 | 743 |
| %Asian | 25.026 | 40.128 | 0 | 100 | 349 | 13.586 | 29.47 | 0 | 100 | 728 |
| %Hispanic | 13.375 | 15.145 | 0 | 100 | 361 | 13.102 | 18.281 | 0 | 100 | 731 |
| %Native American | 0.145 | 0.313 | 0 | 2.1 | 317 | 0.159 | 0.67 | 0 | 7 | 682 |
| %Other | 2.471 | 3.58 | 0 | 24 | 340 | 7.135 | 17.417 | 0 | 92 | 710 |
| %Female | 52.905 | 14.428 | 0 | 100 | 2063 | 52.875 | 16.817 | 0 | 100 | 3138 |
| Publication Year | 2014.696 | 3.418 | 2005 | 2020 | 2213 | 2014.229 | 3.646 | 2005 | 2020 | 3442 |
| Cronbach’s Alpha | .726 | .128 | .25 | .95 | 1734 | .756 | .115 | .22 | .96 | 2730 |

*Notes*. NES indicates the number of effect sizes for which the information is available. Reported mean and standard deviations are weighted by sample size and the inverse of the number of effect sizes included per dataset.

Table 2. Descriptive statistics for continuously coded variables in merged data.

| | Rank-order stability | | | | | Mean-level change | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | *SD* | Min | Max | NES | Mean | *SD* | Min | Max | NES |
| Age at Baseline | 22.661 | 14.814 | 5.85 | 80 | 3578 | 20.697 | 15.528 | 0.25 | 100.5 | 5034 |
| Age at Follow-up | 27.535 | 17.578 | 6.63 | 85 | 3578 | 24.924 | 17.726 | 1 | 102 | 5034 |
| Time lag between assessment | 4.875 | 6.566 | 0.5 | 51 | 3578 | 4.228 | 5.392 | 0.5 | 51 | 5034 |

*Notes*. NES indicates the number of effect sizes for which the information is available. Reported mean and standard deviations are weighted by sample size and the inverse of the number of effect sizes included per dataset.

Table 3. Frequencies for categorical coded variables in novel data published after 2005.

| | Rank-order stability | | | Mean-level change | | |
|---|---|---|---|---|---|---|
| Variable | NES | Nk | N-Sample | NES | Nk | N-Sample |
| Total | 2213 | 122 | 148931 | 3442 | 192 | 233522 |
| Big Five | | | | | | |
| Extraversion | 353 | 53 | 18564 | 487 | 70 | 26037 |
| Agreeableness | 385 | 52 | 14479 | 527 | 74 | 23343 |
| Conscientiousness | 407 | 63 | 40045 | 619 | 89 | 47613 |
| Emotional Stability | 501 | 80 | 42826 | 761 | 114 | 73008 |
| Openness | 318 | 43 | 12042 | 378 | 47 | 12819 |
| Blend | 166 | 26 | 15137 | 479 | 71 | 36850 |
| Contrast | 77 | 16 | 5735 | 185 | 34 | 13747 |
| Region | | | | | | |
| Global South | 122 | 5 | 22308 | 209 | 9 | 30977 |
| Northern Europe | 1531 | 41 | 39600 | 2014 | 59 | 83788 |
| Southern Europe | 63 | 10 | 4883 | 125 | 18 | 10351 |
| North America | 419 | 51 | 69948 | 1004 | 87 | 99921 |
| Asia | 73 | 14 | 11976 | 75 | 16 | 7938 |
| Facet | | | | | | |
| Broad | 1548 | 69 | 70667 | 2021 | 106 | 98330 |
| Narrow | 665 | 70 | 78264 | 1421 | 117 | 135192 |
| Maladaptive | | | | | | |
| Normal range | 1951 | 95 | 107411 | 2878 | 141 | 173373 |
| Maladaptive | 262 | 32 | 41521 | 564 | 67 | 60149 |
| Representative | | | | | | |
| Convenience | 1201 | 104 | 91881 | 2287 | 161 | 141917 |
| Representative | 1012 | 18 | 57050 | 1155 | 31 | 91605 |
| Big Five Taxonomy | | | | | | |
| No | 834 | 87 | 96209 | 1746 | 151 | 175107 |
| Yes | 1379 | 38 | 52722 | 1696 | 46 | 58415 |

*Notes*. NES indicates the number of effect sizes. Nk indicates the number of studies. N-Sample indicates the effective sample size in terms of the number of participants in the studies, taking into account overlapping measures and datasets.

Table 4. Frequencies for categorical variables in merged data.

| | Rank-order stability | | | Mean-level change | | |
|---|---|---|---|---|---|---|
| Variable | NES | Nk | N-Sample | NES | Nk | N-Sample |
| Total | 3578 | 189 | 178503 | 5034 | 276 | 242542 |
| Big Five | | | | | | |
| Extraversion | 695 | 100 | 28270 | 846 | 127 | 41150 |
| Agreeableness | 531 | 85 | 16275 | 713 | 110 | 22060 |
| Conscientiousness | 507 | 91 | 41312 | 751 | 124 | 46882 |
| Emotional Stability | 826 | 129 | 49164 | 1226 | 185 | 79073 |
| Openness | 526 | 81 | 14850 | 591 | 92 | 14173 |
| Blend | 300 | 53 | 20786 | 655 | 104 | 28759 |
| Contrast | 187 | 37 | 7742 | 246 | 52 | 10341 |

*Notes*. NES indicates the number of effect sizes. Nk indicates the number of studies. N-Sample indicates the effective sample size in terms of the number of participants in the studies, taking into account overlapping measures and datasets.

**Age and time effects on rank-order stability**

Table 5 reports the results for all primary analytic models involving rank-order stability in the full meta-analytic dataset of all available effect sizes (the results for only the novel data are presented in Chapter 7 of the SOM). Self-reported stability was high on average ($r$ = .608, 95% CI = .579, .637), but with considerable between-study heterogeneity ($\tau$ = .132). Including linear age and time lag terms into the model reduced between-study heterogeneity ($\tau$ = .098) by 25.76% (i.e., interpreted similarly to an $R^2$ metric). A positive age coefficient would indicate that stability increased across the lifespan, and a negative age coefficient would indicate that stability decreased across the lifespan. The linear model indicated that stability was .006 correlation units higher for each year of age, and that stability was .009 correlation units lower for each additional year of time lag. The intercept of this model was .445, which represents the expected stability for a theoretical study that recruited participants at birth and tracked them for 4.875 years (i.e., the average time lag across all studies).

The full piecewise spline model included 15 age bins and time lags. This model accounted for 41.67% of the between-study heterogeneity, a sizable improvement over the linear term. However, this model was very likely to be overparameterized. After testing all possible combinations of fixed age parameters, the best-fitting model in terms of BIC implied that only three age terms were needed. According to this model, stability appears to increase rapidly from age 0 to 20 (b = .029, 95% CI = .023, .035), it increases more slowly from age 20 to 25 (b = .018, 95% CI = .010, .026), and very slowly increases across the remainder of the lifespan (b = .001, 95% CI = .001, .001). This simplified model accounted for the same amount of between-study heterogeneity as the full spline model. Figure 2 plots these model-implied trends superimposed on the underlying effect sizes. In this plot, each effect size is represented as bubbles with the size of the bubble scaled to reflect the weight the effect size carried in the analysis. Larger bubbles reflect effect sizes that were measured with more precision.

The best-fitting model in terms of AIC indicated slightly more complex trends. The parameter estimates were essentially identical until age 25. Following age 25, the best fitting model in terms of AIC indicated slightly faster increases in stability in midlife ($b$ = .002, 95% CI = 0, .004), followed by decreasing stability after age 60 ($b$ = -.001, 95% CI = -.003, .001). This model accounted for the same between-study heterogeneity as the more parsimonious model, and the parameter estimates indicate trivially small differences. Figure S17 displays age-trends for the ten best fitting models.

**Trait specific age-trends for rank-order stability**

Trait category statistically accounted for a relatively small portion of between-study heterogeneity when included without age and time lag (5.30%) or when included with age and time lag (increase of 2.27% relative to the best spline model). The largest trait effects occurred for extraversion ($b$=.080, 95% CI = .055, .105), and openness ($b$=.059, 95% CI = .032, .086), which appeared to be more stable, and contrast traits (blended traits that combine an adaptive and maladaptive Big Five domain), which appeared to be less stable ($b$=-.055, 95% CI = -.116, .006).

Next, we tested whether allowing for trait-specific lifespan trends by including interaction terms would improve model fit. This model included an additional 18 parameters (6 trait category variables × 3 age bins), but only statistically accounted for an additional 1.52% of between-study heterogeneity relative to the model that did not include interaction effects. Overall, including trait domain led to a fairly trivial improvement in explained variance given the number of additional parameters.

Table 5. Meta-regression results for rank-order stability as a function of age, time lag, and the Big Five in the merged dataset.

| Parameter | Estimate | 95% LB | 95% UB | τ | $R^2$ |
|---|---|---|---|---|---|
| 1. Mean Effect Size | | | | .132 | |
| Intercept | .608 | .579 | .637 | | |
| 2. Linear Age & Lag | | | | .098 | 25.758 |
| Intercept | .445 | .404 | .486 | | |
| Age | .006 | .006 | .006 | | |
| Lag | -.009 | -.011 | -.007 | | |
| 3. Best Fitting Piecewise | | | | .077 | 41.667 |
| Intercept | .069 | -.045 | .183 | | |
| Age 0-20 | .029 | .023 | .035 | | |
| Age 20-25 | .018 | .01 | .026 | | |
| Age 25+ | .001 | .001 | .001 | | |
| Lag | -.011 | -.015 | -.007 | | |
| 4. Trait Category | | | | .125 | 5.303 |
| Intercept | .594 | .565 | .623 | | |
| Extraversion | .08 | .055 | .105 | | |
| Agreeableness | .016 | -.011 | .043 | | |
| Conscientiousness | -.008 | -.053 | .037 | | |
| Emotional Stability | -.003 | -.034 | .028 | | |
| Openness | .059 | .032 | .086 | | |
| Blend | -.055 | -.116 | .006 | | |
| Contrast | -.089 | -.163 | -.015 | | |
| 5. Trait Category, Age, & Lag | | | | .074 | 43.939 |
| Intercept | .079 | -.027 | .185 | | |
| Age 0-20 | .028 | .022 | .034 | | |
| Age 20-25 | .017 | .009 | .025 | | |
| Age 25+ | .001 | .001 | .001 | | |
| Lag | -.011 | -.015 | -.007 | | |
| Extraversion | .045 | .025 | .065 | | |

| | | | | | |
|---|---|---|---|---|---|
| Agreeableness | .003 | -.019 | .025 | | |
| Conscientiousness | -.003 | -.04 | .034 | | |
| Emotional Stability | -.011 | -.035 | .013 | | |
| Openness | .032 | .01 | .054 | | |
| Blend | -.011 | -.064 | .042 | | |
| Contrast | -.055 | -.116 | .006 | | |
| 6. Trait-specific age-trends | | | | .072 | 45.455 |
| 7. Full piecewise | | | | .077 | 41.667 |

*Notes*. Intercepts can be interpreted as rank-order stability estimates (*r*) when the moderators are 0 (e.g., age and time lag are 0); moderator effects as unstandardized regression effects (*b*) from meta-regressions. Positive moderator coefficients indicate higher stability, and negative moderator coefficients indicate lower stability. For parameter estimates for the full piecewise model and the model used to produce trait-specific trends, see the online supplement. 95% LB and UB refer to the lower bound and upper bound of the 95% confidence interval around the parameter estimate.

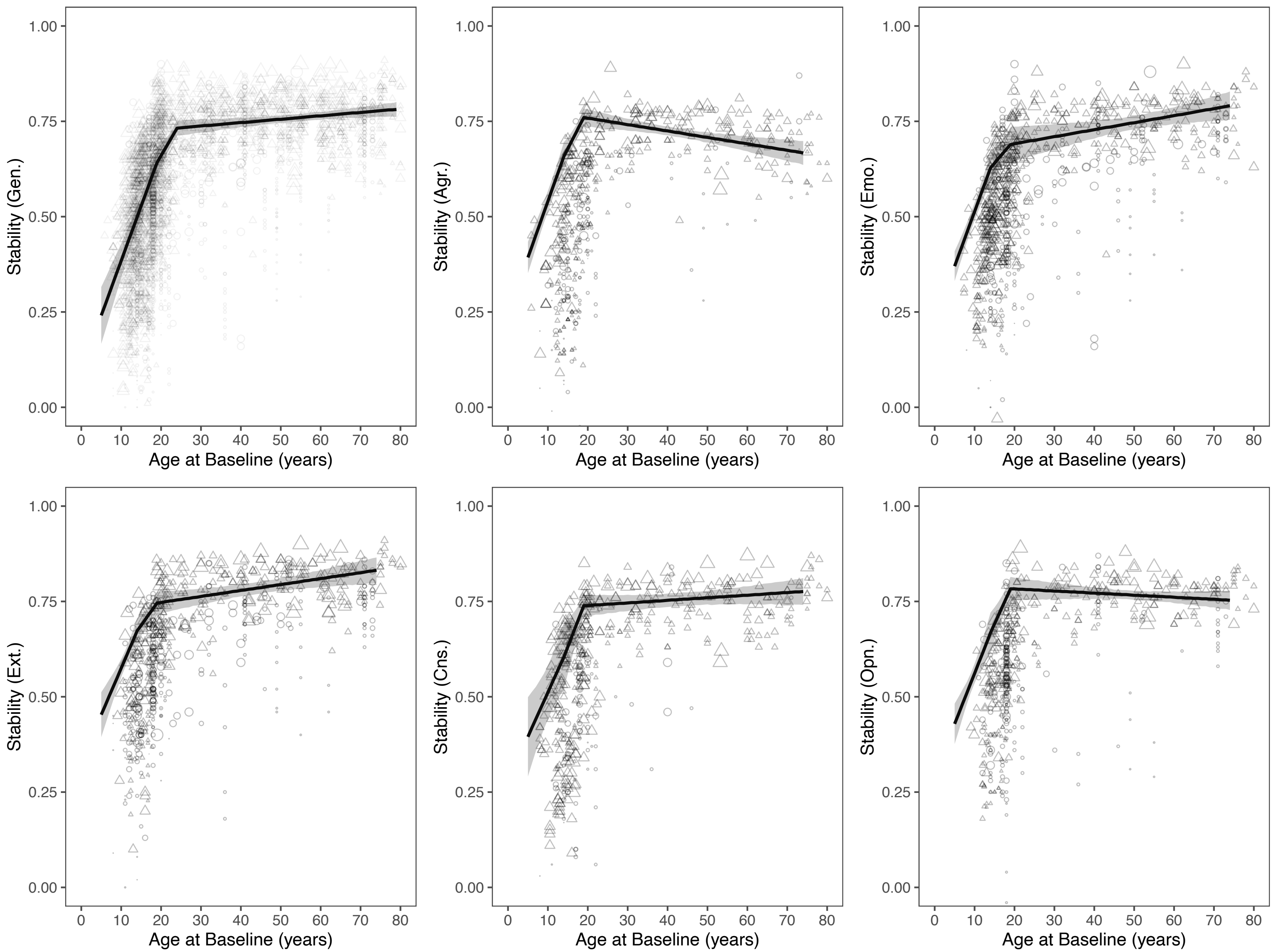
Stability (Gen.)
Stability (Agr.)
Stability (Emo.)
Stability (Ext.)
Stability (Cns.)
Stability (Opn.)
Age at Baseline (years)
0.00
0.25
0.50
0.75
1.00
0 10 20 30 40 50 60 70 80

Figure 2. Lifespan trends for rank-order stability estimates (*r*) for all traits and the Big Five separately. The first panel plots results for the full dataset, and the subsequent panels plot results for extraversion, agreeableness, conscientiousness, emotional stability, and openness, in that order. Effect sizes are plotted in addition to the best fitting spline model and scaled relative to the weight the effect size carried in the analysis, with larger plotting characters carrying more weight. Effect sizes represented as a circle are from previous meta-analyses, and effect sizes represented as a triangle are from the newly coded data. Shading around the trend line reflects the 95% confidence interval. Gen. = General personality effect size. Ext. = Extraversion. Agr. = Agreeableness. Cns. = Conscientiousness. Emo. = Emotional Stability. Opn. = Openness.

Table 6. Meta-regression results from specifying exponential age and time trends for stability in merged data.

| Model | *b*0 | *b*1 | *b*2 | *b*3 | *b*4 | *b*5 | τ | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential Age, Linear Lag | 0.826 | 1.250 | -0.105 | -0.010 | | | 0.079 | 40.152 |
| Exponential Age and Lag | 0.590 | 1.290 | -0.099 | -0.320 | -0.121 | | 0.074 | 43.939 |
| Differential Decay | 0.562 | 1.301 | -0.104 | -0.331 | -0.122 | 0 | 0.074 | 43.939 |

*Notes*. *b*0-*b*5= Unstandardized regression coefficients from meta-regressions. All parameters are statistically significant, except for *b*5 ($p$ = .44). See the online supplement for confidence intervals.

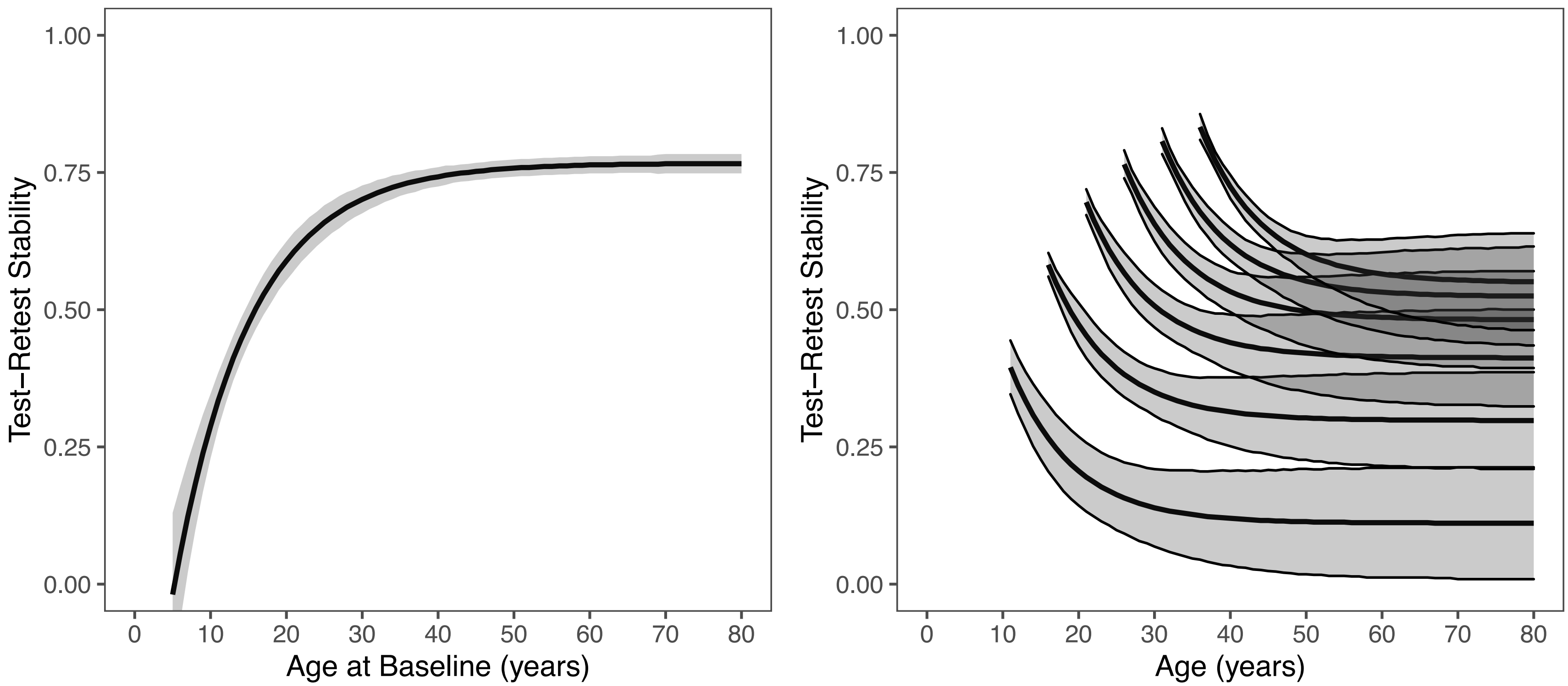


Figure 3. Lifespan trends for estimates of rank-order stability (*r*) and time-based decay of rank-order stability (*r*) based on exponential modeling. The figure on the left displays the trend of increasing stability across the lifespan. The figure on the right displays the time-based decay of stability from the best fitting exponential model. Trend lines reflect expected decay of stability for varying time intervals with ages at baseline of 10, 15, 20, 25, 30, and 35 years.

**Exponential age-trends for rank-order stability**

A benefit of piecewise spline models is that they can uncover complex functional forms due to independent slope estimates. A more parsimonious, but less flexible, model could specify that the age-trends follow an exponential function. Table 6 reports results of exponential specifications of the form:

$$Y = b0 + b1eb2(age) + b3(lag)$$

In this specification, b0 represents the horizontal asymptote (i.e., the maximum stability estimate), $b1$ represents a scaling factor, $b2$ represents the rate of growth, and $b3$ represents a linear time lag term. This exponential model accounted for 40.152% of between-study heterogeneity, nearly identical to the best fitting spline model (Figure 3). The exponential model may be preferred as it uses fewer parameters to describe the age trend. The parameters imply a rapid increase in stability in early life, rising from .087 at age 5, up to .567 by age 15, and to .735 by age 25. Put in terms of the percentage of increase from zero to the asymptote, stability increased by 10%, 69%, and 89% of the total increase at these ages. By age 35, 96% of the total increase has occurred. Following age 35, stability slowly increases for the remainder of the lifespan. At age 50, model-implied stability was .820, very close to the asymptote of .826.

Another benefit of the exponential model over the spline model is that time lag can also be modeled along an exponential function, rather than a linear function. Including an exponential function for lag accounts for an additional 3.79% of between-study heterogeneity. Results imply that studies with longer time intervals tend to have lower estimates of stability, holding age at baseline constant. The expected decrease in stability for hypothetical studies tracking participants across 1, 5, 10, and 50 years would be .036 correlation units (11.4% of total decrease), .145

correlation units (45.4% of total decrease), .225 correlation units (70.1% of total decrease), and .319 correlation units (99.8% of total decrease), respectively.

We attempted to test whether the time-based decay of stability was dependent on age at baseline. Results indicated that the decay or test-retest stability was not dependent on age at baseline. The key parameter was estimated at 0, and including the term did not account for any additional between-study heterogeneity.

**Age and time effects on mean-level change**

For the next set of analyses, we transition from test-retest stability effect sizes to mean-level change. The analyses reported here make use of both self- and other report data.

Table 7 reports results of the primary meta-regression models involving mean-level change in the full meta-analytic dataset of all available effect sizes (the results for only the novel data are presented in Chapter 7 of the SOM). On average, personality traits increased modestly across time ($d$ = .040, 95% CI = .024, .056) with considerable heterogeneity ($\tau$ = .157). Next, we included linear age and time lag terms. Mean-level change was less pronounced at older ages ($b$ = -.001, 95% CI = -.001, -.001), and change was more apparent for studies that used longer time intervals ($b$ = .008, 95% CI = .004, .012). That longer time intervals were related to more evidence of change is consistent with true trait change, rather than potential practice effect confounds. However, age and time lag statistically accounted for a minor amount of between-study heterogeneity (3.82%), particularly relative to the results for stability (25.76%). This discrepancy was not due to a more complex functional form for mean-level change. The best fitting spline model only statistically accounted for 5.73% of between-study heterogeneity, a far lower amount than for stability (41.67%). Figure S20 displays age-trends for the ten best fitting spline models.

Assuming a time lag of 4.24 years (the average lag across studies), the best fitting spline model implied a large amount of mean-level change for a hypothetical study following a birth cohort (*d* = .126, 95% CI = .052, .200). The rate of change per 4.24-year lag decreased for ages at baseline from 0 to 10 years (*b* = -.015, 95% CI = -.025, -.005), increased to a peak rate of change from age 10 to 20 (*b* = .017, 95% CI = .009, .025), declined rapidly from age 20 to 25 (*b* = -.027, 95% CI = -.037, -.017), and plateaued across the remainder of the lifespan at a level of small decreases (*b* = -.001, 95% CI = -.003, .001). Figure 4 displays this trend and shows a clear spike in the rate of mean-level change around the transition from adolescence to adulthood[2]. For the majority of the lifespan, the rate of change in personality traits was negative.

Importantly, these meta-regression parameters reflect the expected rate of change for a study initiated at a certain age at baseline that tracked participants for 4.24 years (each additional year of time lag was associated with an additional *d* of .008, 95% CI = .004, .012). Alternatively, we can describe the results with respect to the cumulative mean-level change that would be expected of a hypothetical sample tracked across the lifespan (cf. Figure 2 in Roberts et al., 2006). Figure 5 displays these descriptive trends for all effect sizes and each of the five trait domains. Here, we assumed that a hypothetical cohort was assessed from birth to age 80.56 years every 4.24 years, and that the cohort experienced the expected rate of change for each age of assessment. For example, the sample would be expected to increase 0.13 standard deviation from age 0 years to age 4.24 years, and then the sample would be expected to increase .06 standard

[2]Given that the lifespan trend did not follow an exponential or any other common functional form, we did not explore alternative specifications as the linear spline model efficiently captured the identified spike in the rate of change.

deviation from age 4.24 years to 8.48 years as implied by the spline results. The lifespan trend for cumulative change is relatively flat with modest increases for the first 15 years of life, followed by a rapid increase up to age 20 with a slight downward trajectory across the remainder of the lifespan. At the peak, traits are implied to increase by about half a standard deviation.

**Trait specific age-trends for mean-level change**

The lifespan trends identified in the previous section reflect the aggregated trends across all traits. In this section, we describe trends for each trait domain. Similar to the results for stability, trait category accounted for relatively little between-study heterogeneity either by itself (1.91%) or in addition to age and time lag (7.64% together vs. 5.73% for age and time lag alone). Including trait-specific age trends accounted for 10.19% of between-study heterogeneity.

As shown in Figure 4, most traits follow the general trend of a peak in the rate of change around age 20, with relatively small rates of change for the remainder of the lifespan. The largest exceptions occurred for emotional stability and conscientiousness. Emotional stability maintained a positive rate of change for the entire lifespan with only a small drop in the rate of change after age 20. Conscientiousness is the trait with the largest shifts across the lifespan. In the early adolescent years, the model-implied rate of change was approximately -.10, followed by a peak of positive change of approximately .20 at age 20. The rate of change slowly approached zero in midlife, and by age 70, the rate of change was -.10 again. Blended traits display a similar trend, but data coverage in adulthood limited a clear interpretation. Similarly, contrast traits were not well-represented in adulthood.

Figure 5 also plots a descriptive representation of cumulative mean-level change separately for each of the Big Five, assuming a hypothetical cohort assessed every 4.24 years from birth to age 80.56 that follows the expected age-specific rates of change (Table 4).

Extraversion displayed little cumulative change for the first portion of the lifespan, followed by slow declines in midlife and old age. Agreeableness showed little cumulative change until late adolescence and early adulthood, followed by increases of about .50 standard deviations early and middle adulthood and slight decreases thereafter. Conscientiousness followed a similar pattern in early life, but the increases were larger (~1 *SD*) and faded out across the lifespan such that mean-levels at age 80 returned to those observed in adolescence. Cumulative mean-level change for emotional stability followed a relatively linear, monotonically increasing trend amounting to almost 1.50 standard deviations across the lifespan. Openness increased in early adulthood, but decreased through the remainder of adulthood by about 0.50 standard deviation. Given the low data coverage for blends and contrasts in adulthood, we include their plots in Figure S19.

Table 7. Meta-regression results for mean-level change as a function of age, time lag, and the Big Five in the merged dataset.

| Parameter | Estimate | 95% LB | 95% UB | τ | $R^2$ |
|---|---|---|---|---|---|
| 1. Mean Effect Size | | | | .157 | |
| Intercept | .04 | .024 | .056 | | |
| 2. Linear Age & Lag | | | | .151 | 3.822 |
| Intercept | .072 | .047 | .097 | | |
| Age | -.001 | -.001 | -.001 | | |
| Lag | .008 | .004 | .012 | | |
| 3. Best Fitting Piecewise | | | | .148 | 5.732 |
| Intercept | .126 | .052 | .2 | | |
| Age 0-10 | -.015 | -.025 | -.005 | | |
| Age 10-20 | .017 | .009 | .025 | | |
| Age 25-30 | -.027 | -.037 | -.017 | | |
| Age 30+ | -.001 | -.003 | .001 | | |
| Lag | .008 | .004 | .012 | | |
| 4. Trait Category | | | | .154 | 1.911 |
| Intercept | .034 | .018 | .05 | | |
| Extraversion | -.048 | -.066 | -.03 | | |
| Agreeableness | .003 | -.026 | .032 | | |
| Conscientiousness | -.002 | -.037 | .033 | | |
| Emotional Stability | .041 | .016 | .066 | | |
| Openness | -.007 | -.042 | .028 | | |
| Blend | .044 | .013 | .075 | | |
| Contrast | -.032 | -.081 | .017 | | |
| 5. Trait Category, Age, & Lag | | | | .145 | 7.643 |
| Intercept | .104 | .035 | .173 | | |
| Age 0-10 | -.014 | -.024 | -.004 | | |
| Age 10-20 | .018 | .01 | .026 | | |
| Age 25-30 | -.027 | -.037 | -.017 | | |
| Age 30+ | 0 | -.002 | .002 | | |

| | | | | | |
|---|---|---|---|---|---|
| Lag | .008 | .004 | .012 | | |
| Extraversion | -.052 | -.07 | -.034 | | |
| Agreeableness | .005 | -.024 | .034 | | |
| Conscientiousness | .003 | -.028 | .034 | | |
| Emotional Stability | .037 | .012 | .062 | | |
| Openness | -.014 | -.045 | .017 | | |
| Blend | .052 | .021 | .083 | | |
| Contrast | -.030 | -.081 | .021 | | |
| 6. Trait-specific age-trends | | | | .141 | 1.191 |
| 7. Full piecewise | | | | .148 | 5.732 |

Notes. Intercepts can be interpreted as Cohen's *d* when the moderators are 0 (e.g., age and time lag are 0); moderator effects as unstandardized regression effects (*b*) from meta-regressions. Positive moderator coefficients indicate more positive (or less negative) rates of change, and negative moderator coefficients indicate more negative (or less positive) rates of change. For parameter estimates for the full piecewise model and the model used to produce trait-specific trends, see the online supplement. 95% LB and UB refer to the lower bound and upper bound of the 95% confidence interval around the parameter estimate.

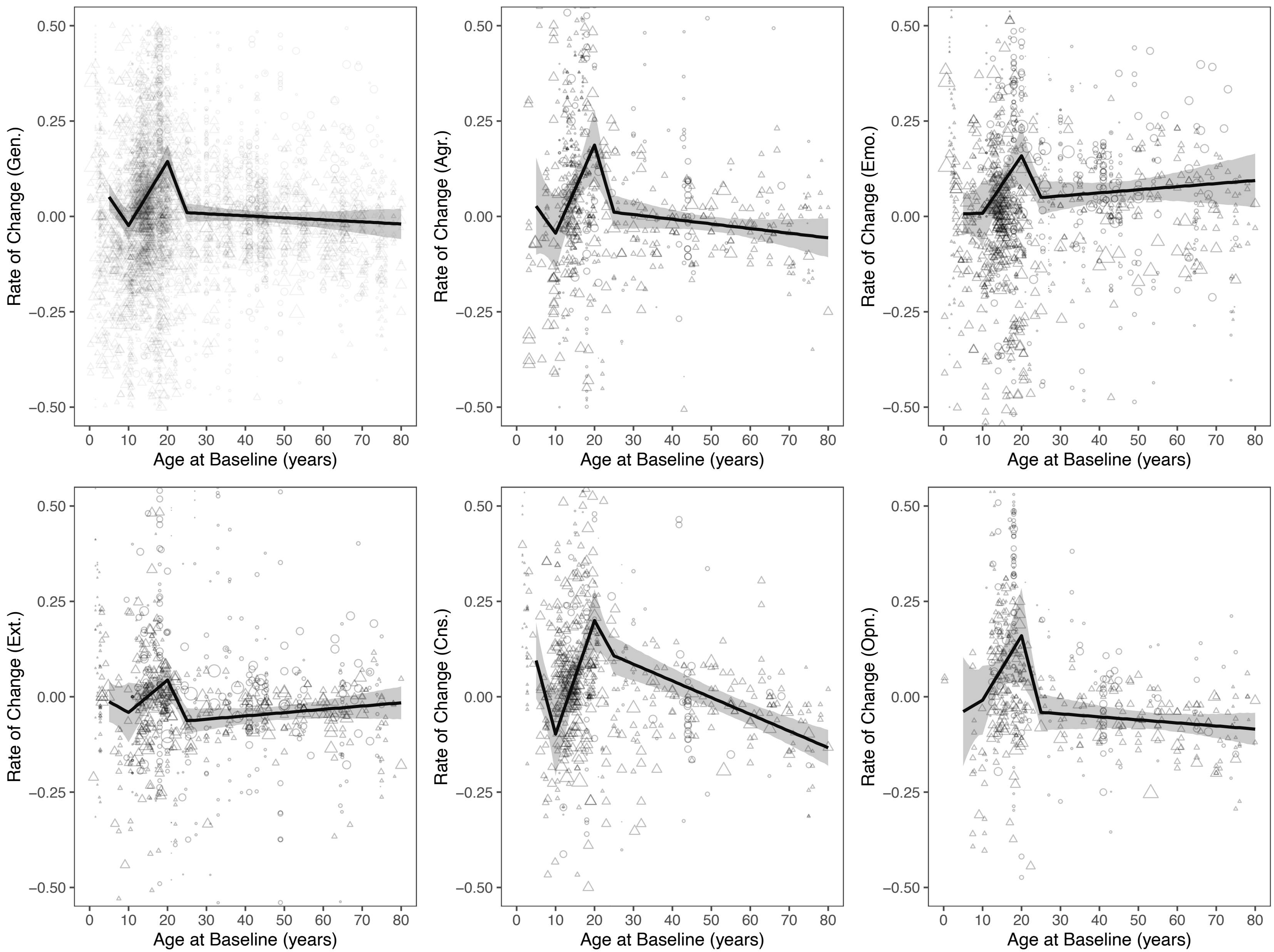
Rate of Change (Gen.)
Rate of Change (Agr.)
Rate of Change (Emo.)
Rate of Change (Ext.)
Rate of Change (Cns.)
Rate of Change (Opn.)
Age at Baseline (years)
0.50
0.25
0.00
−0.25
−0.50
0 10 20 30 40 50 60 70 80

Figure 4. Lifespan trends for rates of mean-level change (Cohen's *d*) for all effect sizes and the Big Five separately. The first panel plots results for the full dataset, and the subsequent panels plot results for extraversion, agreeableness, conscientiousness, emotional stability, and openness, in that order. Effect sizes are plotted in addition to the best fitting spline model and scaled relative to the weight the effect size carried in the analysis, with larger plotting characters carrying more weight. Effect sizes represented as a circle are from previous meta-analyses, and effect sizes represented as a triangle are from the newly coded data. Shading around the trend line reflects the 95% confidence interval. Gen. = General personality effect size. Ext. = Extraversion. Agr. = Agreeableness. Cns. = Conscientiousness. Emo. = Emotional Stability. Opn. = Openness.

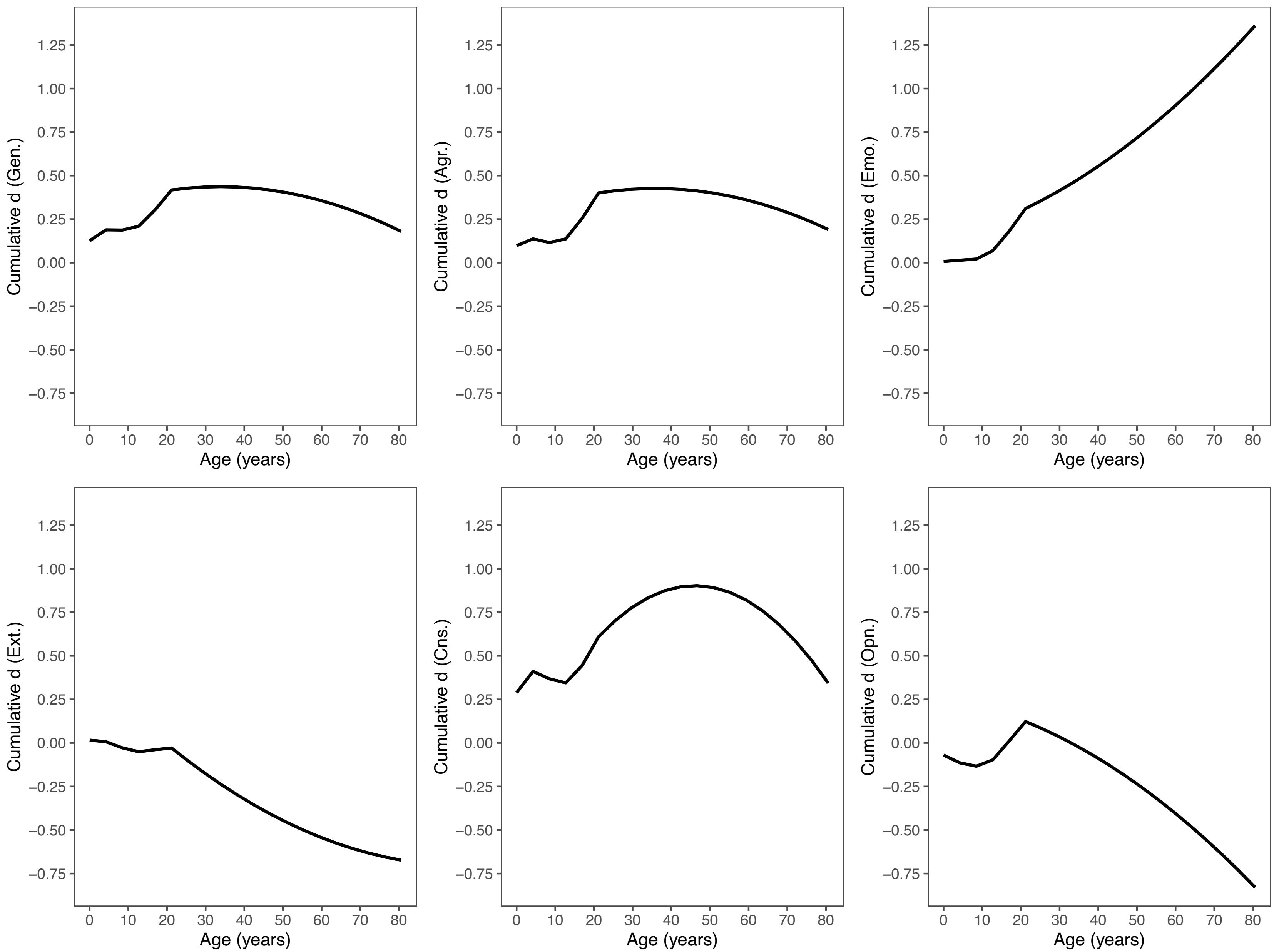
Cumulative d (Gen.)
Cumulative d (Agr.)
Cumulative d (Emo.)
Cumulative d (Ext.)
Cumulative d (Cns.)
Cumulative d (Opn.)
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0
10
20
30
40
50
60
70
80
Age (years)

Figure 5. Description of expected cumulative mean-level change (Cohen's *d*) across the lifespan for all effect sizes and the Big Five separately assuming a hypothetical cohort subject to the age-specific rates of change (Figure 4) and tracked from birth to age 80.56 years every 4.24 years. The first panel plots results for the full dataset, and the subsequent panels plot results for extraversion, agreeableness, conscientiousness, emotional stability, and openness, in that order. Gen. = General personality effect size. Ext. = Extraversion. Agr. = Agreeableness. Cns. = Conscientiousness. Emo. = Emotional Stability. Opn. = Openness.

**Additional moderators of test-retest stability and mean-level change**

Table 8 reports meta-regression results for the remaining moderators. The results are based on the newly coded effect sizes from studies published after 2005. We focus on self-report rank-order stability effect sizes and both self- and other-report mean-level change effect sizes (for results for other-report rank-order stability estimates, see Table S7 in the online supplement).

For rank-order stability, the moderators accounted for a small amount of between-study heterogeneity beyond the age-trends (mean = 2.89%, range = 0% – 7.60%). Effect sizes associated with facet-level ($b$ = -.058, 95% CI = -.107, -.009) or maladaptive (b = -.101, 95% CI = -.172, -.030) measurement tended to be less stable whereas effect sizes from studies with Big Five measures tended to be more stable ($b$ = .055, 95% CI = .018, .092). The effect sizes derived from representative studies also appeared to be less stable (b = -.052, 95% CI = -.134, .030) according to reduction in between-study heterogeneity; however, the confidence interval included zero so we did not interpret this effect. Effect sizes derived from measures with higher Cronbach's alpha tended to display larger estimates of test-retest stability (b = .366, 95%C CI = -.121, .611). This result implies that a measure with a reliability estimate of .90, rather than .80, would be expected to display .037 higher test-retest stability.

For mean-level change, none of the coded moderators substantially reduced between-study heterogeneity. Mean-level change was not more pronounced for domains or facets, for maladaptive or normal-range traits, or for representative or convenience samples. More reliable measures tended to display greater mean-level change, with an expected increase in mean-level change of .051 with .90 rather than .80 reliability.

Reliability could also differ between other moderator categories which could induce spurious associations. For example, if maladaptive measures tend to be measured with less reliability, then lower reliability would be the most likely explanation of our moderator findings, rather than anything concerning the content domain. To evaluate this possibility, we re-ran all of our primary meta-regression models including reliability as a covariate. This inclusion did not alter any of our conclusions. For example, maladaptive measures of personality were less stable than normal trait measures with alpha in the model ($b$ = -.092) or without alpha ($b$ = -.101). Full tabulated results can be found in the supplemental materials.

In summary, we evaluated a range of potential moderators of test-retest stability and mean-level change of personality dimensions. Approximately 50% of between-study heterogeneity was able to be statistically accounted for in stability effect sizes, but only approximately 5% for mean-level change.

Table 8. Additional moderators tested in novel data published after 2005.

| | Stability | | | | | Change | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | Estimate | 95% LB | 95% UB | τ | delta $R^2$ | Estimate | 95% LB | 95% UB | τ | delta $R^2$ |
| Facet | -.057 | -.106 | -.008 | .077 | 2.837 | -.003 | -.038 | .032 | .164 | 0 |
| Maladaptive | -.101 | -.172 | -.03 | .074 | 4.965 | .02 | -.031 | .071 | .164 | 0 |
| Representative | -.052 | -.134 | .03 | .078 | 2.128 | -.013 | -.058 | .032 | .164 | 0 |
| Publication Year | -.001 | -.005 | .003 | .081 | 0 | -.004 | -.008 | 0 | .164 | 0 |
| %Female | 0 | 0 | 0 | .079 | 0 | 0 | 0 | 0 | .163 | 0 |
| North America | -.041 | -.094 | .012 | .076 | 3.546 | .032 | -.001 | .065 | .164 | 0 |
| Global South | .035 | .008 | .062 | | | -.012 | -.057 | .033 | | |
| Northern Europe | .037 | .012 | .062 | | | .005 | -.02 | .03 | | |
| Southern Europe | -.038 | -.107 | .031 | | | -.026 | -.071 | .019 | | |
| Asia | .006 | -.043 | .055 | | | .001 | -.05 | .052 | | |
| Big Five Taxonomy | .055 | .018 | .092 | .077 | 2.837 | -.033 | -.064 | -.002 | .164 | 0 |
| Self-Report | | | | | | .023 | -.085 | .131 | .164 | 0 |
| Other-Report | | | | | | .038 | -.074 | .15 | | |
| Combined-Report | | | | | | -.06 | -.27 | .15 | | |
| Cronbach's alpha | .366 | .121 | .611 | .073 | 7.595 | .514 | .057 | .971 | .152 | 0 |

*Notes*. Moderator effects as unstandardized regression effects ($b$) from meta-regressions. Positive coefficients indicate higher stability and more positive (or less negative) rates of change. Negative coefficients indicate lower stability and more negative (or less positive) rates of change. Each block of moderators was tested in a separate model which included the best fitting piecewise spline model for age and lag as covariates. For comparison in the full dataset, the best fitting spline model had a τ and $R^2$ of .081 and 42.553 for stability, and similar coefficients for change were .164 and 6.818. Because there was missing data for %Female and Cronbach's alpha, τ differs slightly. For test-retest stability, the reference τ was .079 for complete %Female and Cronbach's alpha data. For mean-level change, the reference τ was .163 for complete %Female data and .152 for complete Cronbach's alpha data. For full parameter estimates, see the supplemental materials. Coefficients reflect deviations from the intercept for the identified category. Delta $R^2$ reflects the increase in $R^2$ from including the moderators relative to the best fitting spline model.

## Discussion

In this meta-analysis, we synthesized all available longitudinal data on personality rank-order stability and mean-level change to examine the lifespan development of personality traits from infancy to old age. The combined analyses included over 5,000 effect sizes from more than 300 samples, and data from over 280,000 participants aged 3 months to 100 years. Together, these analyses provide a refined picture of personality trait development across the lifespan and point to important gaps in the literature.

We found an average rank-order stability of $r = .60$, with considerable heterogeneity across studies (H1). We found support for the expected course of trait stability across the lifespan, as indicated by age-graded increases throughout childhood, young, and middle adulthood (H2a). Across the adult years, stability increases glacially, a total increase of only .035 correlation units from age 25 to 60. This result is consistent with our hypothesis of age-graded increases in stability throughout adulthood (H2b), but at a slower pace than previous work has implied. We found partial support for decreases in stability in late adulthood (after age 60, H2c) in some of the specifications, but our primary specification did not include a late-life decrease. We further replicated the negative effect of time on rank-order stability (H3a) with minimum stability estimates >.20, even for time lags of multiple decades, for studies which track adolescent or older participants (H3b). The asymptotic decay of stability in our data was approximately .320 correlation units. Like previous research syntheses, we found little evidence for moderator effects in addition to age and time (H4).

For mean-level change, we predicted and found the highest rates of change in young adulthood (H5a), with relatively small rates of change for the remainder of the lifespan (H5b). We found evidence for increasing rates of mean-level change with increasing time lags between assessment waves (H6a) but little evidence for interaction effects of age and time on rates of change. Consistent with the maturity principle, we found higher rates of mean-level change in emotional stability, agreeableness, and conscientiousness than in openness and extraversion (H7) and no evidence for any other moderator effects (H8). Notably, while we were able to account for 50% of the between-study heterogeneity in rank-order stability effect sizes, only 5% of the between-study heterogeneity for mean-level change was accounted for by including age, time, and other moderator variables.

**The role of age and time in personality trait rank-order stability and mean-level change**

Will you recognize your college friend's personality when you meet her again at age 50? How stable are personality traits from infancy to old age? What is the relationship between time and mean-level change in traits? The present meta-analysis provides answers to these questions by estimating the effects of age and time on rates of trait stability and change. Compared to previous research syntheses, a distinct feature of the present meta-analysis involves the inclusion of longitudinal studies that followed a wide range of age groups over short, long, and very long time periods of up to 5 decades, providing us with the unique opportunity to scrutinize the effects of age and time on personality stability and change across the lifespan. Four findings stand out.

First, consistent with theory and previous research syntheses, we found young adulthood to be the most critical life stage for personality development (Arnett, 2000; Roberts & Mroczek, 2008; Roberts & DelVecchio, 2000; Roberts & Davis, 2016). It is during young adulthood that trait differences crystalize and most traits undergo pronounced mean-level changes. Specifically, throughout childhood and adolescence, but particularly during the transition to young adulthood, traits become increasingly stable with peak levels around age 25. At the same time, this stage is marked by substantial mean-level increases in traits, especially in emotional stability, agreeableness, and conscientiousness. These two patterns may very well be linked and reflect an overall trend towards increased stability and psychological maturity during young adulthood (Bleidorn, 2015; Roberts & Mroczek, 2008; Specht et al., 2014). What drives these pervasive trends in young adulthood? Recent behavioral genetic research provided some answers to this question, indicating that both genetic and environmental influences contribute to both stability and change in personality traits (Bleidorn et al., 2014; Bleidorn et al., 2009; Briley & Tucker-Drob, 2014; Hopwood et al., 2011; Tucker-Drob & Briley, 2018). Identifying the specific genetic and environmental pathways to personality stability and change, however, has turned out to be a more challenging task. Theory and research have emphasized the role of life experiences in personality trait development in young adulthood, indicating that different experiences may be differentially related to stability or change in specific trait domains (e.g., Denissen et al., 2019; Jackson et al., 2012; Lüdtke et al., 2011; Lodi-Smith & Roberts, 2007; Roberts et al., 2005; Schwaba et al., 2019). However, as we will outline in more detail below, research on the sources of personality trait development has yet to account for the complex ways in which persons and environments interact in producing the observed patterns of personality stability and change (Wagner et al., 2020).

A second important finding to emerge from this meta-analysis involves the patterns of rank-order stability and mean-level change in middle adulthood. In contrast to Roberts and DelVecchio (2000) but consistent with other research syntheses (Briley & Tucker-Drob, 2014; Ferguson, 2010), we found minimal evidence for continued increases in rank-order stability throughout the adult lifespan. Instead, the present meta-analytic findings indicate that stability estimates peak around age 25, plateau in middle adulthood, and remain stable or possibly decrease slightly in old age. This discrepancy between Roberts and DelVecchio (2000) and the current results appears to primarily be driven by the relatively lower stability estimates of midlife personality traits compared to the new data (*r* = .57 in original for the 20s vs. a model implied estimate of *r* = .76 at age 25). There are several possible explanations for this difference. The previous meta-analysis was based on a much smaller sample size, and therefore the estimates were less precise. The previous meta-analysis also included non-Likert measures, such as Rorschach tests and behavioral assessments, which tend to be less stable over time. Assessment instruments may have also improved in reliability across time as the field reached a paradigmatic consensus on the Big Five. Finally, it may be the case that individuals' personality traits are indeed increasing in stability more quickly in recent years than previously, although research on emerging adulthood as a historically recent developmental stage would suggest the opposite pattern (Bleidorn & Schwaba, 2017).

The high and stable levels of rank-order stability in middle adulthood go hand in hand with decreasing rates of mean-level change for most traits, except emotional stability, which maintains positive rates of change and continues to increase throughout late adulthood. Together, these findings suggest that the *cumulative continuity principle* (Roberts & Nickel, 2017) is a fitting concept to describe the course of stability in early life but not in middle or late adulthood. The full meta-analytic picture calls for a revision of this principle and provides novel insights into the stability and change of personality traits in middle and late adulthood. Lifespan theories of aging and existing research characterize middle adulthood as a period of maintenance, mastery, and control (Freund & Baltes, 2002; Huttemann et al., 2014). According to Neugarten (1968), enhanced levels of self-awareness, competence, and a wide array of coping strategies prepare middle-aged adults to cope with stressors and maintain established lifestyles. High levels of rank-order stability and decreasing rates of mean-level change are consistent with this depiction of middle adulthood as a period of control, consistency, and maintenance.

These findings also reinforce some but not all aspects of the *maturity principle* of personality development (Roberts & Nickel, 2017; Schwaba et al., 2021). Specifically, the robust increases in conscientiousness and emotional stability in early and middle adulthood are consistent with a trend of psychological maturation and lend further credence to this principle as one of the most replicable “laws” of personality development. However, agreeableness did not show continued increases in midlife, which necessitates a revision of the proposition that this trait would show similar increases throughout this life stage. Furthermore, the three aforementioned traits seem to follow more distinct age-graded trajectories than previously assumed (e.g., Klimstra et al., 2013). Agreeableness does not show continued increases past age 20, conscientiousness increases asymptotically from ages 20-50, and emotional stability shows continuous increases throughout the adult lifespan. Future research and theory may benefit from considering development in these traits separately.

Third, the present meta-analytic findings provide novel insights into the course of personality trait development in late adulthood. While our most parsimonious models indicated stable rates of rank-order stability after age 60, nearly all competing models with roughly equivalent fit tended to include decreases in older age. The mean-levels of most traits decreased in old age. This pattern could be interpreted in the context of theory and research that emphasize the role of losses and resources for late life development (Sutin et al., 2013; Wagner et al., 2016). Old adulthood comes with appreciable challenges such as health problems, loss of loved ones, and a general disengagement from social roles. To the degree that these changes affect older adults' patterns of thoughts, feelings, and behavior, they may explain the observed mean-level decreases in traits such as conscientiousness, extraversion, and agreeableness. Physical and cognitive declines may further limit older individuals' capacity to engage in intellectually demanding activities or to seek out novel experiences, which might contribute to decreases in older adults' levels of openness to experiences (Mueller et al., 2016; Schwaba & Bleidorn, 2021). It can be expected that such effects are stronger for individuals who lack the social, mental, and financial resources to cope with late-life challenges, potentially leading to the observed decreases in trait stability during old age.

A notable exception to the pattern of decreasing mean-levels is the trajectory of emotional stability. Unlike the other traits, emotional stability continues undergoes significant mean-level increases up until old age. This finding adds to the broad evidence for a phenomenon sometimes referred to as the "paradox of aging" (Kunzmann et al., 2000; Kunzmann et al., 2014; Carstensen et al., 2006) describing the finding that, despite the challenges associated with aging, older people tend to be as happy (or even happier) than younger people. According to socioemotional selectivity theory (Carstensen et al., 1999) and the theory of strength and vulnerability integration (Charles & Luong, 2013), the generally high levels of well-being and emotional stability in old age may be explained by age-graded motivational and behavioral changes that lead older adults to prioritize goals that involve emotional meaning and engage in activities that promise immediate gratification and satisfaction. However, conclusive evidence for the sources that underlie personality stability and change in late adulthood has yet to be provided, as we will discuss in more detail below.

A fourth important finding of this meta-analysis involves the role of time. Consistent with theory and previous research syntheses, we found the stability estimates of all traits to decrease as time intervals between assessments increase (Roberts & DelVecchio, 2000). Specifically, exponential models provided evidence for the hypothesis that time-related decreases in rank-order stability decline quickly over shorter intervals, attenuate over longer time intervals, and plateau at modest values around $r = .50$ in adulthood (Fraley & Roberts, 2005; Schimmack & Anusic, 2016). In other words, despite time-related decreases in traits, it is possible to predict individual differences in personality even over extended periods of time, suggesting that there is an enduring or “core” quality to personality traits that remains stable across the entirety of the lifespan (Damian et al., 2019; Lilgendahl et al., 2013). Replicating Roberts et al. (2006), we found more pronounced mean-level changes in traits for studies with longer time intervals between assessment waves. That longer time intervals were related to more evidence of change is consistent with true trait change and speaks against potential practice effect confounds.

In summary, the present findings highlight the role of age and time in personality development. Both personality trait stability and change are closely connected to people’s life stage, implying the effects of age-graded sources that promote stability and drive normative changes in traits. Overall, however, age and time lag accounted for a minor amount of between-study heterogeneity of studies of mean-level change (~6%), particularly in comparison to the results for stability (~41%). A natural question to arise from this finding is what then -- if not time and age -- can explain the large between-study heterogeneity in studies of personality mean-level change? To address this question, we tested additional moderators of personality trait stability or change.

**Additional moderators of personality rank-order stability and mean-level change**

We considered the effects of several moderator variables beyond age and time, including publication year, sample characteristics (nationally representative vs. other, country), and measurement properties (Big Five vs. other, narrow vs. broad, maladaptive vs. normal). We identified three moderators, beyond the effects of age and time.

First, we found evidence for the hypothesis that maladaptive traits are less rank-order stable than adaptive traits. Recent longitudinal research in clinical samples has suggested that maladaptive traits are less stable than normal range traits (for a review, see Hopwood & Bleidorn, 2018). For example, longitudinal research indicated that personality disorders exhibit lower stabilities (Hopwood et al., 2013) than normal-range traits (Roberts & DelVecchio, 2000). Notably, this research has been mostly focused on clinical populations (Morey & Hopwood, 2013). Here, we found evidence for the hypothesis that these differences generalize to maladaptive traits as measured in non-clinical samples.

Second, we found differences between traits measured at a broad level, such as the Big Five vs. narrow level, such as facet traits. Previous meta-analyses that included this distinction found few and rather small differences suggesting that broad traits are slightly less rank-order stable than narrow traits (Briley & Tucker-Drob, 2014; Ferguson, 2010). In contrast to these previous studies, we found broad trait domains to be more rank-order stable than narrow facet traits. Similar to the previous meta-analysis, the difference was of a fairly small magnitude. At first, it may seem surprising that facet and narrow measures are only slightly less stable than domains. The principle of aggregation played an important role in the history of personality psychology. Rather than a specific behavior, personality is reflected in aggregations of behavior which tended to be more stable across time and context. Intuitively, one might expect that broader aggregation of thoughts, feelings, and behaviors should lead to a more stable construct. However, personality nuances (i.e., variance at the item-level that is not shared with the domain) tend to display similar psychometric properties as the domains (Mõttus et al., 2019). By aggregating items to facets to domains, specific variance is reduced in favor of variance that is common across items or facets. The current results imply that the specific variance at the facet level is only modestly less stable than the common variance found at the domain level.

Third, measures that were developed in the tradition of the Big Five taxonomy appear to be more stable than non-Big Five measures. The past two decades has seen a significant increase in the usage of Big Five measures in personality psychology and other subdisciplines. One possible explanation for the higher stability estimates of Big Five measures is that these measures are more reliable and thus less affected by measurement error which may dampen estimates of stability in other measures more strongly.

These three significant effects must be considered in the context of the more general pattern of small moderator effects. Together, these variables explained a trivial amount of the between-study variance in personality stability and change estimates beyond age and time. As such, we still do not have a good account of the large between-study heterogeneity in studies of personality development, especially for studies of mean-level change. A clear goal for future research on personality development should be to account for the substantial between-study heterogeneity unaccounted for here.

It is possible that there are other relevant moderator variables that were not included in this meta-analysis. For example, few studies reported information about the ethnic composition of the sample, so we were not able to test for the effects of ethnicity. Another possibility is that we still lack the statistical power to detect significant effects for some of the moderator variables. For instance, the vast majority of studies included in the present meta-analysis were conducted in samples from Northern European or North American countries. Very few effect sizes were derived from Asian or African samples, precluding a rigorous test of cultural differences in personality stability and change.

**What accounts for personality rank-order stability and mean-level change?**

Evidence that personality traits are dynamic characteristics naturally leads to questions about the sources that underlie the patterns of rank-order stability and mean-level change across the lifespan. The past two decades have seen a surge of studies that were aimed at identifying the factors that can explain personality stability or predict change in personality traits (Bleidorn et al., 2021). As described above, there is clear evidence that both genetic and environmental influences contribute to stability and change in personality traits (Briley & Tucker-Drob, 2014; Hopwood et al., 2011). However, there still is little evidence for replicable effects of any specific set of genetic or environmental factors that can explain these trends at the population level (cf. Roberts & Yoon, 2022).

Few studies have examined the particular genetic and biological sources of personality rank-order stability (Lo et al., 2017; Penke & Jokela, 2016) and little is known about the specific environmental factors that contribute to the patterns of personality rank-order stability across the lifespan. Another and even less researched source of personality rank-order stability involves person-environment transactions (Fraley & Roberts, 2005, Hopwood et al., 2022, Roberts & Nickel, 2017). Person-environment transactions can manifest in three general was. First, people can select into environments that are consistent with their personality; second, people may evoke certain reactions from the environment that reinforce their personality, and third, people may actively shape environments in ways that make them more consistent with their personality. There is good evidence that people do select into certain kinds of environments, evoke reactions from their social environment, and can shape the environments they are in (Rauthmann &

Sherman, 2020). However, few studies have tested whether these mechanisms indeed contribute to the observed patters of personality rank-order stability across the lifespan.

Evidence for replicable sources of mean-level changes in personality traits is also sparse. Theory and some research have focused on the role of age-graded life transitions in explaining mean-level change in personality traits (Bleidorn, 2015, Roberts et al., 2005). These works have tested the idea that age-graded life events, such as graduating from college, entering the first job, or becoming a parent, can trigger personality-trait change because they force people to change their patterns of thoughts, feelings, and behaviors (Bleidorn, 2012; Luedtke et al., 2011; Jackson et al., 2012; Jokela et al., 2014; Kornadt et al. 2018; Specht et al., 2011; Wagner et al., 2016). For example, young adults who entered their first romantic relationship have been found to increase in levels of emotional stability and conscientiousness (e.g., Wagner et al., 2015). Similarly, there is some evidence that graduation from school or college is associated with increases in emotional stability, openness, and conscientiousness (e.g., Bleidorn, 2012; Luedtke et al., 2011, Schwaba et al., 2018). However, these findings must be evaluated within the broader context of research that yielded more mixed and sometimes conflicting results about the links between life events and personality development (Asendorpf & Wilpers, 1998; Denissen et al., 2019; Specht et al., 2011; van Scheppingen et al., 2016; for a review, see Bleidorn et al., 2018). The implicit assumption that single life events, such as parenthood, unemployment, or divorce, would elicit the same trait changes in most people – independent of their particular context and life circumstances – may be too simplified (Luhmann et al., 2020). Life events are not random, do not occur in isolation, and elicit different changes in different people's personality traits. For example, selecting into college is predicted by the very personality traits that appear to change during college (Noftle & Robins, 2007). Moreover, the transition to college is associated with a host of other potentially meaningful experiences such as moving out of one's parents' home, meeting new friends and romantic partners, or exploring new identities and worldviews (Bleidorn et al., 2020); and graduation from college may open the door to a whole new set of life and career events such as graduate school or paid employment, which, again, entail exposure to new environments and relationships. Finally, the same life events may elicit different responses in different people depending on their psychological background and life situation (Denissen et al., 2019). An isolated focus on the main effects of single and discrete life events is thus difficult to achieve and possibly misleading (Luhmann et al., 2020).

A more promising avenue to the sources of mean-level change involves the inclusion of subjective experiences of age-graded life events (Lodi-Smith & Roberts, 2007; Luhmann et al., 2020). For example, only those individuals who consider a life event as important and are truly impacted in their patterns of thoughts, feelings, strivings, and behavior, may also experience changes in their personality traits (Schwaba et al., 2022). The findings of the present meta-analysis point to critical developmental periods during which personality trait levels appear to be more malleable and presumably also more amenable to influences of environmental experiences.

**A meta-perspective on the empirical landscape of personality psychology**

By combining effect sizes from earlier meta-analyses, we were able to detect shifts in the empirical landscape of personality psychology. In particular, a large shift has taken place from a reliance on boutique samples tracked by individual research groups toward large-scale panel studies. This shift has positive and negative elements. For example, the new studies tended to have much larger sample sizes, were nationally representative, and covered longer time spans than a few years. These features are possible when conducting research at that scale. As awareness that personality is consequential for numerous life outcomes increases (e.g., Bleidorn et al., 2019), personality inventories have been included in several databases such as GSOEP, the Household, Income and Labour Dynamics in Australia (HILDA), or the Health and Retirement Study (HRS). These high-quality databases are treasure troves for personality researchers. With sample sizes of tens of thousands, even the most elaborate statistical tests may be sufficiently powered.

However, this trend is not without costs. Panel studies tend to use short measures of personality traits and other constructs. The studies may also not include all the relevant variables necessary to test one's hypotheses or measure variables at timescales that are disconnected from the theoretical process underlying change (Hopwood et al., 2021). Individual studies designed to test individual hypotheses will always be necessary at the cutting edge; panel studies stick to standard and short measures, typically. For the current meta-analysis, the biggest cost was a lack of any informant-report effect sizes in samples of participants older than 17 years. This gap was not present to nearly the same extent in the Roberts and DelVecchio (2000) data. Longitudinal work using multi-method approaches to assess traits to triangulate sources of stability and change are sorely needed.

In our view, there are two good options for personality psychology going forward. First, researchers will never be able to fully rely on panel studies to test narrow and targeted hypotheses of interest. Therefore, some effort in the field should go to collecting data. We would propose collaborative consortia as an effective way to maximize sample size and diversity and minimize burden on any one investigator. Second, and perhaps more outside the typical personality psychologist's comfort zone, researchers should find ways to lobby stakeholders to include more psychometrically sound and interesting personality development study designs in future or ongoing panel studies. As personality is policy relevant (Bleidorn et al., 2019) due to the many replicable associations with meaningful life outcomes (Soto, 2019), we have a powerful argument that our expertise will be useful and effective.

**Strengths and limitations**

The present study is a comprehensive meta-analysis of rank-order stability and mean-level change in personality traits across the lifespan. We integrated data from over 300 studies that sampled more than 280,000 individuals who ranged in age from infancy to old age and used state-of-the-art meta regression techniques to model both linear and non-linear effects of age and time on stability and change. However, our approach is not without limitations.

First, while our random-effects spline models allowed us to detect discontinuities in trends, the interpretation of these models is complicated when data is sparse. For example, our best-fitting model indicated stable rank-order estimates throughout late adulthood while alternative models suggested decreasing rates of rank-order stability in old age. Given that there are still relatively few studies including ages greater than 60 years old at baseline (323 effect sizes derived from 11 studies representing 6820 participants, only 2.5% of the total sample), such subtle model differences are difficult to detect. To address the possibility that decreasing rank-order stabilities provide a more accurate description of the data, we have provided results from alternative modeling approaches. With the most complex piecewise spline model, age trends can be examined with the greatest flexibility. Alternatively, the continuous exponential models provide a more parsimonious general impression of the data that is potentially less influenced by noise (Briley & Tucker-Drob, 2014). Notably, visual inspection of age trends across traits indicates that each model provides essentially the same results with only slight deviations.

Second, several moderator tests may have been underpowered if there was not sufficient data density across levels of the moderator for the entire life span. This limitation was particularly relevant for tests of self- vs. other report format. The predominant use of other (i.e. parent and teacher) report assessments in personality stability studies of children and adolescents precluded a rigorous test of moderation given the strong confound between report format and age. Although we found consistent results for stability estimates when analyses were restricted to effect sizes from other report data (see supplemental materials), more studies are needed to test potential effects in the full data set. A lack of data also prevented us from estimating the potential effects of ethnicity and country on personality stability and change. The vast majority of studies included in this meta-analysis utilized data collected in Western, educated, industrialized, rich, and democratic (WEIRD, Henrich et al., 2010) countries in Northern Europe and North America. Longitudinal studies in samples from non-WEIRD cultures remain sparse. Moreover, existing studies often fail to report the ethnic composition of the sample, precluding researchers from examining the role of ethnic background in personality development. More research on cultural and ethnic differences in personality development is needed to probe the generalizability of the current findings and examine both universal and specific mechanisms that might underlie stability and change in personality traits. A related concern could be that we restricted our search to English speaking articles on PsycINFO. These restrictions might have contributed to the lack of data from non-WEIRD cultures. However, additional searches of non-English works on other databases did not increase the number of hits.

A third concern is that several of the moderator variables examined here are not independent but correlated, which complicates the interpretation of the moderator effects. For example, most nationally representative panel studies also used broad Big Five measures. Our goal was to describe the expected effect sizes across different measures, samples, and methods.

By presenting the data with respect to all moderators, we aim to provide researchers with more ways to gauge heterogeneity in the literature.

Fourth, in this meta-analysis we considered both rank-order stability and mean-level change. However, neither this meta-analysis nor previous research syntheses provided information about individual differences in trait change. Individual differences in change can be expressed as variance or standard deviation and reflect the degree to which individuals conform to vs. deviate from the overall trends of mean-level change (e.g., Mroczek & Spiro, 2003). Although there is a growing body of evidence for individual differences in personality trait change throughout the lifespan (Graham et al., 2020; Schwaba & Bleidorn, 2018), there were not enough studies that focused on this type of change to include it in this meta-analysis. A reliable assessment of individual differences in personality change is critical for examinations of environmental and experiential influences on personality change. Future meta-analyses should include this type of change to provide a more comprehensive account of these effects and their links with time and age.

Fifth, as with previous meta-analyses on this topic, a remaining limitation to the present study was the necessity of categorizing various trait measures into the Big Five domains. This approach allowed us to synthesize and communicate findings from a broad range of domains and research areas. However, despite the fact that our classification approach was rooted in theory, based on empirical correlations, and executed by multiple raters, the act of categorizing personality measures into broad trait domains inherently leads to some loss of information (Roberts et al., 2006). Perhaps most apparently, we still do not have a good account of personality stability and change at lower levels of the trait hierarchy. Breaking down the broader trait domains into lower-order facets may reveal more nuanced developmental trends that may be obscured by focusing on the broader domain level (Schwaba et al., 2021; Soto & John, 2012). Unfortunately, very few longitudinal studies have examined personality stability and change in lower-order traits, pointing to an important direction for future research in this area. Similarly, the interstitial space between the Big Five dimensions is not well-documented in terms of test-retest stability and mean-level change. Rather than code pairwise (or higher order) combinations of the Big Five, we chose to organize the effect sizes in terms of blended and contrast categories. We chose this approach to maximize power and minimize the risk of false positives from odd combinations of dimensions that may only be represented by a small number of studies. Of course, the psychological or behavioral meaning of blended or contrast traits is not obvious and may be more heterogeneous than for the Big Five. We evaluated this possibility and found that there were similar amounts of between-study heterogeneity in test-retest stability for emotional stability ($\tau = .185$), blends ($\tau = .150$), and contrasts ($\tau = .213$).

**Conclusion**

The results of the present meta-analysis update and extend previous research syntheses on personality trait rank-order stability and mean-level change. Together, these findings provide a compelling picture of personality trait development across the lifespan. Providing further evidence for the relevance of young adulthood as a formative period of personality development and maturation, we find this life stage to be characterized by rapid increases in trait stability and high rates of mean-level change, all of which are in the direction of greater maturity. While middle adulthood appears to be a period of stability and continued increases in traits, we find late adulthood to be characterized by mean-level decreases in all traits except emotional stability. Across age, rank-order stability estimates decrease while rates of mean-level change increase with increasing time intervals. Lingering questions remain about the genetic and environmental sources of personality trait stability and change. Previous research has highlighted the need to account for person-environment transactions in explaining both stability and change in personality traits. A crucial next step for personality theory and research will be to document how such effects unfold over time to result in personality development.

**References**

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., ... & Ioannidis, J. P. (2020). A consensus-based transparency checklist. Nature human Behaviour, 4, 4-6.

*Allemand, M., Schaffhuser, K., & Martin, M. (2015). Long-Term Correlated Change Between Personality Traits and Perceived Social Support in Middle Adulthood. Personality and Social Psychology Bulletin, 41, 420–432.

*Allemand, M., Zimprich, D., & Hertzog, C. (2007). Cross-Sectional Age Differences and Longitudinal Age Changes of Personality in Middle Adulthood and Old Age. Journal of Personality, 75, 323–358.

Allport, G. W. (1961). Pattern and growth in personality. Oxford, England: Holt, Reinhart & Winston.

Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. Journal of Personality and Social Psychology, 110, 766.

Ardelt, M. (2000). Still stable after all these years? Personality stability theory revisited. Social Psychology Quarterly, 392-405.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. American Psychologist, 55, 469–480

*Asendorpf, J. B., & Motti–Stefanidi, F. (2018). Mediated Disposition–Environment Transactions: The Dae Model. European Journal of Personality, 32, 167–185.

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. Journal of Personality and Social Psychology, 74, 1531–1544.

*Ashenhurst, J. R., Harden, K. P., Corbin, W. R., & Fromme, K. (2015). Trajectories of binge drinking and personality change across emerging adulthood. Psychology of Addictive Behaviors, 29, 978–991.

*Atherton, O. E., Zheng, L. R., Bleidorn, W., & Robins, R. W. (2019). The codevelopment of effortful control and school behavioral problems. Journal of Personality and Social Psychology, 117, 659–673.

*Baardstu, S., Karevold, E. B., & von Soest, T. (2017). Childhood antecedents of Agreeableness: A longitudinal study from preschool to late adolescence. Journal of Research in Personality, 67, 202–214.

Baltes, P. B., & Baltes, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. Successful aging: Perspectives from the Behavioral Sciences, 1, 1-34. New York: Cambridge University Press.

*Banny, A. M., Tseng, W.-L., Murray-Close, D., Pitula, C. E., & Crick, N. R. (2014). Borderline personality features as a predictor of forms and functions of aggression during middle childhood: Examining the roles of gender and physiological reactivity. Development and Psychopathology, 26, 789–804.

*Barker, E. D., & Salekin, R. T. (2012). Irritable oppositional defiance and callous unemotional traits: Is the association partially explained by peer victimization?: Irritable oppposition. Journal of Child Psychology and Psychiatry, 53, 1167–1175.

Bleidorn, W. (2012). Hitting the road to adulthood: Short-term personality development during a major life transition. Personality and Social Psychology Bulletin, 38, 1594-1608.

Bleidorn, W. (2015). What accounts for personality maturation in early adulthood? Current Directions in Psychological Science, 24, 245-252.

Bleidorn, W., Hill, P. L., Back, M.D., Denissen J.J.A., Hennecke, M., Hopwood, C.J., … & Roberts, B. W. (2019). The policy relevance of personality traits. *American Psychologist.*

Bleidorn, W., Hopwood, C. J., Back, M. D., Denissen, J. J., Hennecke, M., Hill, P. L., ... & Zimmermann, J. (2021). Personality trait stability and change. *Personality Science,* 2, 1-20.

Bleidorn, W., Hopwood, C. J., Back, M.D., Denissen J.J.A., Hennecke, M.., … & Zimmermann, J. (2020). Longitudinal Experience-Wide Association Studies—A framework for studying personality change. European Journal of Personality, 34, 285-300.

Bleidorn, W., Hopwood, C. J., Lucas, R. E. (2018). Life events and personality trait change. Journal of Personality, 86, 83-96.

Bleidorn, W., Kandler, C., & Caspi, A. (2014). The behavioral genetics of personality development in adulthood - Classic, contemporary, and future trends. European Journal of Personality, 28, 244-255.

Bleidorn, W., Kandler, C., Riemann, R., Angleitner, A., & Spinath, F.M. (2009). Patterns and sources of adult personality development: Growth curve analyses of the NEO-PI-R scales in a longitudinal twin study. Journal of Personality and Social Psychology, 97, 142-155.

Bleidorn, W., Klimstra, T.A., Denissen, J.J.A., Rentfrow, P.J., Potter, J., & Gosling, S.D. (2013). Personality maturation around the world: A cross-cultural examination of Social Investment Theory. Psychological Science, 24, 2530-2540.

*Blonigen, Daniel M., Carlson, M. D., Hicks, B. M., Krueger, R. F., & Iacono, W. G. (2008). Stability and Change in Personality Traits From Late Adolescence to Early Adulthood: A Longitudinal Twin Study. Journal of Personality, 76, 229–266.

*Blonigen, Daniel Michael. (2008). Extending the developmental and behavior genetic literature of psychopathy from a personality-based approach: Continuity and change from adolescence to adulthood and gene-environment interplay. (Doctoral dissertation). University of Minnesota Twin Cities, Minneapolis, MN.

*Borghuis, J., Bleidorn, W., Sijtsma, K., Branje, S., Meeus, W. H. J., & Denissen, J. J. A. (2020). Longitudinal associations between trait neuroticism and negative daily experiences in adolescence. Journal of Personality and Social Psychology, 118, 348–363.

Borghuis, J., Denissen, J. J. A., Oberski, D., Sijtsma, K., Meeus, W. H. J., Branje, S., . . . Bleidorn, W. (2017). Big Five personality stability, change, and co-development across adolescence and early adulthood. Journal of Personality and Social Psychology, 113, 641-657.

*Bornovalova, M. A., Verhulst, B., Webber, T., McGue, M., Iacono, W. G., & Hicks, B. M. (2018). Genetic and environmental influences on the codevelopment among borderline personality disorder traits, major depression symptoms, and substance use disorder symptoms from adolescence to young adulthood. Development and Psychopathology, 30, 49–65.

*Bornstein, M. H., Hahn, C.-S., Putnick, D. L., & Pearson, R. (2019). Stability of child temperament: Multiple moderation by child and mother characteristics. British Journal of Developmental Psychology, 37, 51–67.

*Bould, H., Joinson, C., Sterne, J., & Araya, R. (2013). The Emotionality Activity Sociability Temperament Survey: Factor analysis and temporal stability in a longitudinal cohort. Personality and Individual Differences, 54, 628–633.

*Brandt, N. D., Mike, A., & Jackson, J. J. (2019). Do school-related experiences impact personality? Selection and socialization effects of impulse control. Developmental Psychology, 55, 2561–2574.

*Bratko, D., & Butkovic, A. (2007). Stability of Genetic and Environmental Effects from Adolescence to Young Adulthood: Results of Croatian Longitudinal Twin Study of Personality. Twin Research and Human Genetics, 10, 151–157.

*Brendgen, M., Wanner, B., Morin, A. J. S., & Vitaro, F. (2005). Relations with Parents and with Peers, Temperament, and Trajectories of Depressed Mood During Early Adolescence. Journal of Abnormal Child Psychology, 33, 579–594.

*Brenning, K., Soenens, B., Braet, C., & Beyers, W. (2013). Longitudinal Dynamics of Depressogenic Personality and Attachment Dimensions in Adolescence: An Examination of Associations with Changes in Depressive Symptoms. Journal of Youth and Adolescence, 42, 1128–1144.

Briley, D. A., & Tucker-Drob, E. M. (2014). Genetic and environmental continuity in personality development: A meta-analysis. Psychological Bulletin, 140, 1303.

*Burris, J. L., Riley, E., Puleo, G. E., & Smith, G. T. (2017). A longitudinal study of the reciprocal relationship between ever smoking and urgency in early adolescence. Drug and Alcohol Dependence, 178, 519–526.

Carstensen, L. L. (2006). The influence of a sense of time on human development. Science, 312, 1913–1915.

Card, N. A. (2019). Lag as moderator meta-analysis: A methodological approach for synthesizing longitudinal data. *International Journal of Behavioral Development, 43,* 80-89.

Charles, S. T., & Luong, G. (2013). Emotional experience across adulthood: The theoretical model of strength and vulnerability integration. Current Directions in Psychological Science, 22, 443-448.

*Chen, S.-K., Lo, M.-T., & Lin, S. S. J. (2017a). Impulsivity as a precedent factor for problematic Internet use: How can we be sure?: IMPULSIVITY AS A PRECEDENT FACTOR. International Journal of Psychology, 52, 389–397.

*Chen, S.-K., Lo, M.-T., & Lin, S. S. J. (2017b). Impulsivity as a precedent factor for problematic Internet use: How can we be sure?: IMPULSIVITY AS A PRECEDENT FACTOR. International Journal of Psychology, 52, 389–397.

*Childs, A. W., Fite, P. J., Moore, T. M., Lochman, J. E., & Pardini, D. A. (2014). Bidirectional Associations Between Parenting Behavior and Child Callous-Unemotional Traits: Does Parental Depression Moderate this Link? Journal of Abnormal Child Psychology, 42, 1141–1151.

*Christensen, D., Fahey, M. T., Giallo, R., & Hancock, K. J. (2017). Longitudinal trajectories of mental health in Australian children aged 4-5 to 14-15 years. PLOS ONE, 12, e0187974.

*Clark, D. A., Donnellan, M. B., Robins, R. W., & Conger, R. D. (2015). Early adolescent temperament, parental monitoring, and substance use in Mexican-origin adolescents. Journal of Adolescence, 41, 121–130.

Costa, P. T., McCrae, R. R., & Löckenhoff, C. E. (2019). Personality across the life span. Annual Review of Psychology, 70, 423-448.

*Damian, L. E., Stoeber, J., Negru-Subtirica, O., & Băban, A. (2017). On the Development of Perfectionism: The Longitudinal Role of Academic Achievement and Academic Efficacy: Development of Perfectionism. Journal of Personality, 85, 565–577.

*Damian, R. I., Spengler, M., Sutu, A., & Roberts, B. W. (2019). Sixteen going on sixty-six: A longitudinal study of personality stability and change across 50 years. Journal of Personality and Social Psychology, 117, 674.

*Dancho, L. D. (2005). Cognitive and personality correlates of adolescent depressed mood. (Doctoral dissertation). Fordham University, Bronx, NY.

*Davies, P. T., Sturge-Apple, M. L., Cicchetti, D., Manning, L. G., & Vonhold, S. E. (2012). Pathways and processes of risk in associations among maternal antisocial personality symptoms, interparental aggression, and preschooler's psychopathology. Development and Psychopathology, 24, 807–832.

*de Leeuw, R. N. H., Scholte, R. H. J., Sargent, J. D., Vermulst, A. A., & Engels, R. C. M. E. (2010). Do interactions between personality and social-environmental factors explain smoking development in adolescence? Journal of Family Psychology, 24, 68–77.

*Dekker, L. P., Hartman, C. A., van der Vegt, E. J., Verhulst, F. C., van Oort, F. V., & Greaves-Lord, K. (2015). The longitudinal relation between childhood autistic traits and psychosexual problems in early adolescence: The Tracking Adolescents' Individual Lives Survey study. Autism, 19, 684–693.

*den Boer, L., Klimstra, T. A., Branje, S. J. T., Meeus, W. H. J., & Denissen, J. J. A. (2019). Personality Maturation during the Transition to Working Life: Associations with Commitment as A Possible Indicator of Social Investment. European Journal of Personality, 33, 456–467.

Denissen, J.J.A., Luhmann, M. Chung, J.M. Bleidorn, W. Transactions between life events and personality traits across the adult lifespan (2019). Journal of Personality and Social Psychology, 116, 612-633.

*Dennis, D. J. (2010). The development of boys' aggressive behaviour: A Process-Person-Context-Time model. (Doctoral Dissertation). University of Alberta, AB, Canada.

*Deventer, J., Lüdtke, O., Nagy, G., Retelsdorf, J., & Wagner, J. (2019). Against all odds - is a more differentiated view of personality development in emerging adulthood needed? The case of young apprentices. British Journal of Psychology, 110, 60–86.

*Duckworth, A. L., & Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (Grit–S). Journal of Personality Assessment, 91, 166–174.

*Duriez, B., Klimstra, T. A., Luyckx, K., Beyers, W., & Soenens, B. (2012). Right–Wing Authoritarianism: Protective Factor against Or Risk Factor for Depression? European Journal of Personality, 26, 536–549.

*Edelstein, R. S., Newton, N. J., & Stewart, A. J. (2012). Narcissism in Midlife: Longitudinal Changes in and Correlates of Women's Narcissistic Personality Traits: Narcissism in Midlife. Journal of Personality, 80, 1179–1204.

*Eisenbarth, H., Demetriou, C. A., Kyranides, M. N., & Fanti, K. A. (2016). Stability Subtypes of Callous–Unemotional Traits and Conduct Disorder Symptoms and Their Correlates. Journal of Youth and Adolescence, 45, 1889–1901.

*Eklund, J. M., Kerr, M., & Stattin, H. (2010). Romantic relationships and delinquent behaviour in adolescence: The moderating role of delinquency propensity. Journal of Adolescence, 33, 377–386.

*Ellis, R. E. R., Seal, M. L., Simmons, J. G., Whittle, S., Schwartz, O. S., Byrne, M. L., & Allen, N. B. (2017). Longitudinal Trajectories of Depression Symptoms in Adolescence: Psychosocial Risk Factors and Outcomes. Child Psychiatry & Human Development, 48, 554–571.

*Ericson, M. (2011). Relationships between latent executive function ability and antisocial behavior throughout development: A common pathways cross-lagged analysis. (Doctoral Dissertation). University of Southern California, Los Angeles, CA.

*Fan, C. K. (2011). A Longitudinal Examination of Children's Emotion Regulation Problems, Negative Parenting Behaviors, and the Development of Internalizing Behavior Problems. (Doctoral dissertation). University of Michigan, Ann Arbor, MI.

*Fanti, K. A., & Munoz Centifanti, L. C. (2014). Childhood Callous-Unemotional Traits Moderate the Relation Between Parenting Distress and Conduct Problems Over Time. Child Psychiatry & Human Development, 45, 173–184.

*Farrell, A. H., Volk, A. A., & Vaillancourt, T. (2020). Empathy, Exploitation, and Adolescent Bullying Perpetration: A Longitudinal Social-Ecological Investigation. Journal of Psychopathology and Behavioral Assessment, 42, 436–449.

Ferguson, C. J. (2010). A meta-analysis of normal and disordered personality across the life span. Journal of Personality and Social psychology, 98, 659.

*Fisher, J. H., & Brown, J. L. (2018). A Prospective, Longitudinal Examination of the Influence of Childhood Home and School Contexts on Psychopathic Characteristics in Adolescence. Journal of Youth and Adolescence, 47, 2041–2059.

*Fontaine, N. M. G., McCrory, E. J. P., Boivin, M., Moffitt, T. E., & Viding, E. (2011). Predictors and outcomes of joint trajectories of callous–unemotional traits and conduct problems in childhood. Journal of Abnormal Psychology, 120, 730–742.

*Fortman, T. L. (2011). A longitudinal study of the stability of hope in late adolescence. (Doctoral Dissertation). The Ohio State University, Columbus, OH.

*Fosco, W. D., Hawk, L. W., Colder, C. R., Meisel, S. N., & Lengua, L. J. (2019). The development of inhibitory control in adolescence and prospective relations with delinquency. Journal of Adolescence, 76, 37–47.

*Fossati, A., Sharp, C., Borroni, S., & Somma, A. (2019). Psychometric Properties of the Borderline Personality Features Scale for Children-11 (BPFSC-11) in a Sample of Community Dwelling Italian Adolescents. European Journal of Psychological Assessment, 35, 70–77.

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: a dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course, Psychological Review, 112, 60-74.

Freund, A. M., & Baltes, P. B. (2002). Life-management strategies of selection, optimization and compensation: Measurement by self-report and construct validity. Journal of Personality and Social Psychology, 82, 642.

*Frogner, L., Gibson, C. L., Andershed, A.-K., & Andershed, H. (2018). Childhood psychopathic personality and callous–unemotional traits in the prediction of conduct problems. American Journal of Orthopsychiatry, 88, 211–225.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: properties of persons, situations, and behaviors. Journal of personality and social psychology, 60, 773.

*Galla, B. M., Wood, J. J., Tsukayama, E., Har, K., Chiu, A. W., & Langer, D. A. (2014). A longitudinal multilevel model analysis of the within-person and between-person effect of effortful engagement and academic self-efficacy on academic performance. Journal of School Psychology, 52, 295–308.

*Garcia, S. E., Tully, E. C., Tarantino, N., South, S., Iacono, W. G., & McGue, M. (2013). Changes in genetic and environmental influences on trait anxiety from middle adolescence to early adulthood. Journal of Affective Disorders, 151, 46–53.

*Geng, F., Xu, T., Wang, Y., Shi, H., Yan, C., Neumann, D. L., … Chan, R. C. (2013). Developmental trajectories of schizotypal personality disorder-like behavioural manifestations: A two-year longitudinal prospective study of college students. BMC Psychiatry, 13, 323.

*Golle, J., Rose, N., Göllner, R., Spengler, M., Stoll, G., Hübner, N., … Nagengast, B. (2019). School or Work? The Choice May Change Your Personality. Psychological Science, 30, 32–42.

*Gómez-Baya, D., & Mendoza, R. (2018). Trait Emotional Intelligence as a Predictor of Adaptive Responses to Positive and Negative Affect During Adolescence. Frontiers in Psychology, 9, 2525.

Graham, E. K., Weston, S. J., Gerstorf, D., Yoneda, T. B., Booth, T., Beam, C. R., ... & Estabrook, R. (2020). Trajectories of Big Five Personality Traits: A Coordinated Analysis of 16 Longitudinal Samples. European Journal of Personality. Advance online publication, doi: 10.1002/per.2259

*Grant, V. V., & Bagnell, A. L. (2009). Early Temperament Prospectively Predicts Anxiety in Later Childhood. The Canadian Journal of Psychiatry, 54, 11.

*Greischel, H., Noack, P., & Neyer, F. J. (2016). Sailing Uncharted Waters: Adolescent Personality Development and Social Relationship Experiences During a Year Abroad. Journal of Youth and Adolescence, 45, 2307–2320.

*Grevenstein, D., & Bluemke, M. (2017). Longitudinal Factor Analysis and Measurement Invariance of Sense of Coherence and General Self-Efficacy in Adolescence. European Journal of Psychological Assessment, 33, 377–387.

*Guarneri-White, M. E. (2017). Making a Bad Situation Worse: Co-rumination and Peer Victimization in Two Adolescent Samples. (Doctoral Dissertation). University of Texas at Arlington, Arlington, TX.

*Hallquist, M. N., Hipwell, A. E., & Stepp, S. D. (2015). Poor self-control and harsh punishment in childhood prospectively predict borderline personality symptoms in adolescent girls. Journal of Abnormal Psychology, 124, 549–564.

*Hanington, L., Ramchandani, P., & Stein, A. (2010). Parental depression and child temperament: Assessing child to parent effects in a longitudinal population study. Infant Behavior and Development, 33, 88–95.

*Harakeh, Z., Scholte, R. H. J., de Vries, H., & Engels, R. C. M. E. (2006). Association between personality and adolescent smoking. Addictive Behaviors, 31, 232–245.

*Harden, K. P., Quinn, P. D., & Tucker-Drob, E. M. (2012). Genetically influenced change in sensation seeking drives the rise of delinquent behavior during adolescence: Sensation seeking and delinquency. Developmental Science, 15, 150–163.

*Hayden, E. P., Olino, T. M., Mackrell, S. V. M., Jordan, P. L., Desjardins, J., & Katsiroumbas, P. (2013). Cognitive vulnerability to depression during middle childhood: Stability and associations with maternal affective styles and parental depression. Personality and Individual Differences, 55, 892–897.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39-65.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. Nature, 466, 29.

*Hicks, B. M., Durbin, C. E., Blonigen, D. M., Iacono, W. G., & McGue, M. (2012). Relationship between personality change and the onset and course of alcohol dependence in young adulthood: Alcohol dependence and personality change. Addiction, 107, 540–548.

*Hicks, B. M., Johnson, W., Durbin, C. E., Blonigen, D. M., Iacono, W. G., & McGue, M. (2014). Delineating Selection and Mediation Effects Among Childhood Personality and

Environmental Risk Factors in the Development of Adolescent Substance Abuse. Journal of Abnormal Child Psychology, 42, 845–859.

*Hill, P. L., & Burrow, A. L. (2012). Viewing Purpose Through an Eriksonian Lens. Identity, 12, 74–91.

*Hoff, K. A., Song, Q. C., Einarsdóttir, S., Briley, D. A., & Rounds, J. (2020). Developmental structure of personality and interests: A four-wave, 8-year longitudinal study. Journal of Personality and Social Psychology, 118, 1044–1064.

*Holzman, J. B. (2018). Examining cross-lagged relations between behavioral inhibition and inhibitory control during early childhood: Predicting subsequent internalizing and externalizing problems. (Doctoral Dissertation). Northern Illinois University, DeKalb, IL.

*Hong, R. Y., Lee, S. S. M., Chng, R. Y., Zhou, Y., Tsai, F.-F., & Tan, S. H. (2017a). Developmental Trajectories of Maladaptive Perfectionism in Middle Childhood: Maladaptive Perfectionism. Journal of Personality, 85, 409–422.

*Hong, R. Y., Lee, S. S. M., Chng, R. Y., Zhou, Y., Tsai, F.-F., & Tan, S. H. (2017b). Developmental Trajectories of Maladaptive Perfectionism in Middle Childhood: Maladaptive Perfectionism. Journal of Personality, 85, 409–422.

Hopwood, C. J., & Bleidorn, W. (2018). Stability and change in personality and personality disorders. Current Opinion in Psychology, 21, 6-10.

Hopwood, C. J., Bleidorn, W., Wright, A. G. C. (2021). Connecting theory to methods in longitudinal research. Perspectives on Psychological Science.

Hopwood, C. J., Donnellan, M. B., Blonigen, D. M., Krueger, R. F., McGue, M., Iacono, W. G., & Burt, S. A. (2011). Genetic and environmental influences on personality trait stability and growth during the transition to adulthood: A three-wave longitudinal study. Journal of Personality and Social Psychology, 100, 545-556.

Hopwood, C.J., Morey, L.C., Donnellan, M.B., Samuel, D.B., Grilo, C.M., McGlashan, T.H., Shea, M.T., Zanarini, M.C., Gunderson, J.G., & Skodol, A.E. (2013). Ten year rank-order stability of personality traits and disorders in a clinical sample. Journal of Personality, 81, 335-344.

Hopwood, C. J., Wright, A.G.C., Bleidorn, W. (2022). Distinguishing personality and psychopathology. *Nature Reviews Psychology.*

Hudson, N. W., Roberts, B. W., & Lodi-Smith, J. (2012). Personality trait development and social investment in work. Journal of research in personality, 46, 334-344.

*Hull, J. G., Brunelle, T. J., Prescott, A. T., & Sargent, J. D. (2014). A longitudinal study of risk-glorifying video games and behavioral deviance. Journal of Personality and Social Psychology, 107, 300–325.

*Hutteman, R., Denissen, J. J. A., Asendorpf, J. B., & van Aken, M. A. G. (2009). Changing dynamics in problematic personality: A multiwave longitudinal study of the relationship between shyness and aggressiveness from childhood to early adulthood. Development and Psychopathology, 21, 1083–1094.

Hutteman, R., Hennecke, M., Orth, U., Reitz, A. K., & Specht, J. (2014). Developmental tasks as a framework to study personality development in adulthood and old age. European Journal of Personality, 28, 267-278.

*Ilmarinen, V., Vainikainen, M., Verkasalo, M., & Lönnqvist, J. (2019). Peer Sociometric Status and Personality Development from Middle Childhood to Preadolescence. European Journal of Personality, 33, 606–626.

*Israel, A., Lüdtke, O., & Wagner, J. (2019). The longitudinal association between personality and achievement in adolescence: Differential effects across all Big Five traits and four achievement indicators. Learning and Individual Differences, 72, 80–91.

Jackson, J. J., Thoemmes, F., Jonkmann, K., Lüdtke, O., & Trautwein, U. (2012). Military training and personality trait development: Does the military make the man, or does the man make the military?. Psychological Science, 23, 270-277.

*Jambon, M., & Smetana, J. G. (2018). Callous–unemotional traits moderate the association between children's early moral understanding and aggression: A short-term longitudinal study. Developmental Psychology, 54, 903–915.

James, W. (1950). *The principles of psychology.* New York: Dover. (Original work published 1890)

John, O. P. (2021). History, measurement, and conceptual elaboration of the Big Five trait taxonomy: The paradigm matures. In O. P. John & R. W. Robins (Eds.), Handbook of personality: Theory and research (p. 35–82). The Guilford Press.

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.

Jokela, M., Hakulinen, C., Singh-Manoux, A., & Kivimäki, M. (2014). Personality change associated with chronic diseases: pooled analysis of four prospective cohort studies. Psychological Medicine, 44, 2629–2640.

*Josefsson, K., Jokela, M., Cloninger, C. R., Hintsanen, M., Salo, J., Hintsa, T., … Keltikangas-Järvinen, L. (2013). Maturity and change in personality: Developmental trends of temperament and character in adulthood. Development and Psychopathology, 25, 713–727.

*Joyce, A. W., Kraybill, J. H., Chen, N., Cuevas, K., Deater-Deckard, K., & Bell, M. A. (2016). A Longitudinal Investigation of Conflict and Delay Inhibitory Control in Toddlers and Preschoolers. Early Education and Development, 27, 788–804.

*Kaiser, A., Bonsu, J. A., Charnigo, R. J., Milich, R., & Lynam, D. R. (2016). Impulsive Personality and Alcohol Use: Bidirectional Relations Over One Year. Journal of Studies on Alcohol and Drugs, 77, 473–482.

Kandler, C., & Ostendorf, F. (2016). Additive and synergetic contributions of neuroticism and life events to depression and anxiety in women. European Journal of Personality, 30, 390-405.

Kandler, C., Bleidorn, W., Riemann, R., Spinath, F.M., Thiel, W., & Angleitner, A. (2010). Sources of cumulative continuity in personality: A longitudinal multiple-rater twin study. Journal of Personality and Social Psychology, 98, 995-1008.

Kandler, C., Kornadt, A. E., Hagemeyer, B., & Neyer, F. J. (2015). Patterns and sources of personality development in old age. Journal of Personality and Social Psychology, 109, 175.

*Karukivi, M., Pölönen, T., Vahlberg, T., Saikkonen, S., & Saarijärvi, S. (2014). Stability of alexithymia in late adolescence: Results of a 4-year follow-up study. Psychiatry Research, 219, 386–390.

*Kawamoto, T., & Endo, T. (2015). Genetic and Environmental Contributions to Personality Trait Stability and Change Across Adolescence: Results From a Japanese Twin Sample. Twin Research and Human Genetics, 18, 545–556.

*Kechter, A., & Leventhal, A. M. (2019). Longitudinal Association of Sleep Problems and Distress Tolerance During Adolescence. Behavioral Medicine, 45, 240–248.

*Keefer, K. V., Holden, R. R., & Parker, J. D. A. (2013). Longitudinal assessment of trait emotional intelligence: Measurement invariance and construct continuity from late childhood to adolescence. Psychological Assessment, 25, 1255–1272.

*Kienbaum, J., Zorzi, M., & Kunina-Habenicht, O. (2019). The development of interindividual differences in sympathy: The role of child personality and adults' responsiveness to distress. Social Development, 28, 398–413.

*Kiive, E., Laas, K., Akkermann, K., Comasco, E., Oreland, L., Veidebaum, T., & Harro, J. (2014). Mitigating aggressiveness through education? The monoamine oxidase A genotype and mental health in general population. Acta Neuropsychiatrica, 26, 19–28.

*Kim, J., Deater–Deckard, K., Mullineaux, P. Y., & Beekman, C. R. (2010). Context Specificity in Stability of Hyperactivity–Impulsivity. European Journal of Personality, 24, 656–674.

*Klimstra, T. A., Akse, J., Hale, W. W., Raaijmakers, Q. A. W., & Meeus, W. H. J. (2010). Longitudinal associations between personality traits and problem behavior symptoms in adolescence. Journal of Research in Personality, 44, 273–284.

Klimstra, T. A., Bleidorn, W., Asendorpf, J. B., Van Aken, M. A., & Denissen, J. J. (2013). Correlated change of Big Five personality traits across the lifespan: A search for determinants. Journal of Research in Personality, 47, 768-777.

*Klimstra, T. A., Luyckx, K., Hale III, W. W., & Goossens, L. (2014). Personality and externalizing behavior in the transition to young adulthood: The additive value of personality facets. Social Psychiatry and Psychiatric Epidemiology, 49, 1319–1333.

*Kokko, K., & Pulkkinen, L. (2005). Stability of aggressive behavior from childhood to middle age in women and men. Aggressive Behavior, 31, 485–497.

Kunzmann, U., Kappes, C., & Wrosch, C. (2014). Emotional aging: a discrete emotions perspective. Frontiers in Psychology, 5, 380.

Kunzmann, U., Little, T. D., & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study. Psychology and Aging, 15, 511.

*Kuzma, E., Sattler, C., Toro, P., Schönknecht, P., & Schröder, J. (2011). Premorbid Personality Traits and Their Course in Mild Cognitive Impairment: Results from a Prospective Population-Based Study in Germany. Dementia and Geriatric Cognitive Disorders, 32, 171–177.

Kuiper, R. M., & Ryan, O. (2020). Meta-analysis of lagged regression models: A continuous-time approach. *Structural Equation Modeling: A Multidisciplinary Journal, 27,* 396-413

*Laceulle, O. M., van Aken, M. A. G., Ormel, J., & Nederhof, E. (2015). Stress-sensitivity and reciprocal associations between stressful events and adolescent temperament. Personality and Individual Differences, 81, 76–83.

*Lee, E. H., Zhou, Q., Eisenberg, N., & Wang, Y. (2013). Bidirectional relations between temperament and parenting styles in Chinese children. International Journal of Behavioral Development, 37, 57–67.

*Lehnart, J., Neyer, F. J., & Eccles, J. (2010). Long-Term Effects of Social Investment: The Case of Partnering in Young Adulthood. Journal of Personality, 78, 639–670.

*Leikas, S., & Salmela-Aro, K. (2015). Personality Trait Changes Among Young Finns: The Role of Life Events and Transitions: Personality Change and Life Events. Journal of Personality, 83, 117–126.

*Lengua, L. J., & Kovacs, E. A. (2005). Bidirectional associations between temperament and parenting and the prediction of adjustment problems in middle childhood. Journal of Applied Developmental Psychology, 26, 21–38.

*LeRoy, M. (2013). Predictors of coparenting: Infant temperament, infant gender, and hostile-reactive parenting. (Doctoral Dissertation). Bowling Green State University, Bowling Green, OH.

*Lilgendahl, J. P., Helson, R., & John, O. P. (2013). Does Ego Development Increase During Midlife? The Effects of Openness and Accommodative Processing of Difficult Events: Ego Development in Midlife. Journal of Personality, 81, 403–416.

*Lin, K., Twisk, J. W. R., & Rong, J. (2011). Longitudinal interrelationships between frequent geographic relocation and personality development: Results from the Amsterdam Growth and Health Longitudinal Study. American Journal of Orthopsychiatry, 81, 285–292.

*Liu, E. T.-H., Chen, W.-L., Tsai, L.-T., Wu, M.-S., & Hong, C.-L. (2012). Interpersonal Mechanisms in the Relationships between Dependency/Self-Criticism and Depressive

Symptoms in Taiwanese Undergraduates: A Prospective Study. Journal of Social and Clinical Psychology, 31, 972–1006.

*Liu, J., & Xia, L.-X. (2018). The Direct and Indirect Relationship between Interpersonal Self-Support Traits and Perceived Social Support: A Longitudinal Study. Current Psychology, 37, 73–81.

Lo, M.-T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C.-C., Wang, Y., . . . Chen, C.-H. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. Nature Genetics, 49, 152-156.

Lodi-Smith, J., & Roberts, B. W. (2007). Social investment and personality: A meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism. Personality and Social Psychology Review, 11, 68-86.

*Lodi-Smith, J., Geise, A. C., Roberts, B. W., & Robins, R. W. (2009). Narrating personality change. Journal of Personality and Social Psychology, 96, 679–689.

*Luan, Z., & Bleidorn, W. (2020). Self–other personality agreement and internalizing problems in adolescence. Journal of Personality, 88, 568–583.

*Luan, Z., Hutteman, R., Denissen, J. J. A., Asendorpf, J. B., & van Aken, M. A. G. (2017). Do you see my growth? Two longitudinal studies on personality development from childhood to young adulthood from multiple perspectives. Journal of Research in Personality, 67, 44–60.

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. Journal of Personality and Social Psychology, 101, 847–861.

*Lüdtke, O., Roberts, B. W., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: Life paths, life events, and personality trait change at the transition to university life. Journal of Personality and Social Psychology, 101, 620–637.

Lüdtke, O., Roberts, B.W., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: Life paths, life events, and personality trait change at the transition to university life. Journal of Personality and Social Psychology, 101, 620-637.

*Lüdtke, O., Trautwein, U., & Husemann, N. (2009). Goal and Personality Trait Development in a Transitional Period: Assessing Change and Stability in Personality Development. Personality and Social Psychology Bulletin, 35, 428–441.

*Luengo Kanacri, B. P., Pastorelli, C., Eisenberg, N., Zuffianò, A., Castellani, V., & Caprara, G. V. (2014). Trajectories of prosocial behavior from adolescence to early adulthood: Associations with personality change. Journal of Adolescence, 37, 701–713.

Luhmann, M., Fassbender, I., Alcock, M., & Haehner, P. (2020). A dimensional taxonomy of perceived characteristics of major life events. Journal of Personality and Social Psychology.

*MacDonell, E. T., & Willoughby, T. (2020). Investigating honesty-humility and impulsivity as predictors of aggression in children and youth. Aggressive Behavior, 46, 97–106.

*Malmberg, M., Kleinjan, M., Overbeek, G., Vermulst, A. A., Lammers, J., & Engels, R. C. M. E. (2013). Are there reciprocal relationships between substance use risk personality profiles and alcohol or tobacco use in early adolescence? Addictive Behaviors, 38, 2851–2859.

*Martin, A. J., Nejad, H. G., Colmar, S., & Liem, G. A. D. (2013). Adaptability: How students' responses to uncertainty and novelty predict their academic and non-academic outcomes. Journal of Educational Psychology, 105, 728–746.

*Mayzer, R. (2004). First alcohol use and the development of antisocial behavior problems from preschool through early adolescence. (Doctoral Dissertation). Michigan State University, East Lansing, EI.

McCrae, R. R., & Costa, P. (2008). The Five-Factor Theory of Personality. In O. P. John, R. W. Robins & L. A. Pervin (Eds.), Handbook of personality: Theory and research (3rd ed.) (pp. 1-58). New York, NY, US: Guilford Press.

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22,* 114.

McCrae, R. R., Costa, P. T. J., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M. D., . . . Smith, P. B. (2000). Nature over nurture: Temperament, personality, and life span development. Journal of Personality and Social Psychology, 78, 173-186.

*Meldrum, R. C., Petkovsek, M. A., Boutwell, B. B., & Young, J. T. N. (2017). Reassessing the relationship between general intelligence and self-control in childhood. Intelligence, 60, 1–9.

*Miranda, D., & Claes, M. (2008). Personality Traits, Music Preferences and Depression in Adolescence. International Journal of Adolescence and Youth, 14, 277–298.

Mischel W. Personality and assessment. New York: Wiley; 1968.

Morey, L. C., & Hopwood, C. J. (2013). Stability and change in personality disorders. Annual Review of Clinical Psychology, 9, 499-528.

*Morizot, J., & Le Blanc, M. (2005). Searching for a Developmental Typology of Personality and Its Relations to Antisocial Behavior: A Longitudinal Study of a Representative Sample of Men. Journal of Personality, 73, 139–182.

Mõttus, R., Johnson, W., & Deary, I. J. (2012). Personality traits in old age: Measurement and rank-order stability and some mean-level change. Psychology and Aging, 27, 243-249.

*Mõttus, R., Johnson, W., & Deary, I. J. (2012). Personality traits in old age: Measurement and rank-order stability and some mean-level change. Psychology and Aging, 27, 243–249.

Mroczek, D. K., & Spiro III, A. (2003). Modeling intraindividual change in personality traits: Findings from the Normative Aging Study. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58, 153-165.

Mueller, S., Wagner, J., Drewelies, J., Duezel, S., Eibich, P., Specht, J., … Gerstorf, D. (2016). Personality development in old age relates to physical health and cognitive performance: Evidence from the Berlin Aging Study II. Journal of Research in Personality, 65, 94–108.

*Muñoz, L. C., Pakalniskiene, V., & Frick, P. J. (2011). Parental monitoring and youth behavior problems: Moderation by callous-unemotional traits over time. European Child & Adolescent Psychiatry, 20, 261–269.

*Nauta, M. M. (2012). Temporal Stability, Correlates, and Longitudinal Outcomes of Career Indecision Factors. Journal of Career Development, 39, 540–558.

*Negru-Subtirica, O., Pop, E. I., Crocetti, E., & Meeus, W. (2020). Social comparison at school: Can GPA and personality mutually influence each other across time? Journal of Personality, 88, 555–567.

*Nelson, B. W., Byrne, M. L., Simmons, J. G., Whittle, S., Schwartz, O. S., O'Brien-Simpson, N. M., … Allen, N. B. (2018). Adolescent temperament dimensions as stable prospective risk and protective factors for salivary C-reactive protein. British Journal of Health Psychology, 23, 186–207.

Neugarten, B. L. (1968). Middle age and aging (Vol. 10). University of Chicago Press.

*Nichols, L. R., Samek, D. R., & McConnell, L. (2019). Key personality traits and alcohol use disorder symptoms in first and second year college students: Detangling antecedent from consequence. Addictive Behaviors, 89, 178–187.

*Nielsen, J. D., Olino, T. M., Dyson, M. W., & Klein, D. N. (2019). Reactive and Regulatory Temperament: Longitudinal Associations with Internalizing and Externalizing Symptoms through Childhood. Journal of Abnormal Child Psychology, 47, 1771–1784.

*O'Neill, S. E. (2004). Personality processes in the development of alcohol problems during the college years and beyond. (Doctoral Dissertation). University of Missouri-Columbia, Columbia, MO.

*Ogle, C. M., Rubin, D. C., & Siegler, I. C. (2014). Changes in Neuroticism Following Trauma Exposure: Changes in Neuroticism After Trauma. Journal of Personality, 82, 93–102.

Oltmanns, J. R., Jackson, J. J., & Oltmanns, T. F. (2020). Personality change: Longitudinal self-other agreement and convergence with retrospective-reports. Journal of Personality and Social Psychology, 118, 1065–1079.

Ormel, J., Von Korff, M., Jeronimus, B. F., & Riese, H. (2017). Set-Point Theory and personality development: Reconciliation of a paradox. In J. Specht (Ed.), Personality development across the lifespan (p. 117–137). Elsevier Academic Press.

*Orth, U., & Luciano, E. C. (2015). Self-esteem, narcissism, and stressful life events: Testing for selection and socialization. Journal of Personality and Social Psychology, 109, 707–721.

Orth, U., Erol, R. Y., & Luciano, E. C. (2018). Development of self-esteem from age 4 to 94 years: A meta-analysis of longitudinal studies. Psychological Bulletin, 144, 1045-1080.

*Osafo Hounkpatin, H., Wood, A. M., Boyce, C. J., & Dunn, G. (2015). An Existential-Humanistic View of Personality Change: Co-Occurring Changes with Psychological Well-Being in a 10 Year Cohort Study. Social Indicators Research, 121, 455–470.

*Oshri, A., Kogan, S. M., Kwon, J. A., Wickrama, K. A. S., Vanderbroek, L., Palmer, A. A., & MacKillop, J. (2018). Impulsivity as a mechanism linking child abuse and neglect with substance use in adolescence and adulthood. Development and Psychopathology, 30, 417–435.

*Partridge, T., & Lerner, J. V. (2007). A latent growth-curve approach to difficult temperament. Infant and Child Development, 16, 255–265.

Penke, L., & Jokela, M. (2016). The evolutionary genetics of personality revisited. Current Opinion in Psychology, 7, 104-109.

*Peterson, S. J., & Smith, G. T. (2019). Impulsigenic personality: Is urgency an example of the jangle fallacy? Psychological Assessment, 31, 1135–1144.

*Prinzie, P., & Deković, M. (2008). Continuity and change of childhood personality characteristics through the lens of teachers. Personality and Individual Differences, 45, 82–88.

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

*Putnam, S. P., & Rothbart, M. K. (2006). Development of Short and Very Short Forms of the Children's Behavior Questionnaire. Journal of Personality Assessment, 87, 102–112.

*Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. Infant Behavior and Development, 29, 386–401.

*Ramos, M. C., Guerin, D. W., Gottfried, A. W., Bathurst, K., & Oliver, P. H. (2005). Family Conflict and Children's Behavior Problems: The Moderating Role of Child Temperament. Structural Equation Modeling: A Multidisciplinary Journal, 12, 278–298.

*Rantanen, J., Tillemann, K., Metsäpelto, R.-L., Kokko, K., & Pulkkinen, L. (2015). Longitudinal study on reciprocity between personality traits and parenting stress. International Journal of Behavioral Development, 39, 65–76.

*Rawlings, A. M., Tapola, A., & Niemivirta, M. (2020). Longitudinal predictions between temperamental sensitivities and achievement goal orientations in the early school years. European Journal of Psychology of Education, 35, 451–475.

*Rikoon, S. H. (2013). Toward an omnibus assessment of noncognitive skills: A longitudinal multitrait-multimethod application in middle schools. (Doctoral Dissertation). University of Pennsylvania, Philadelphia, PA.

Rauthmann, J. F., & Sherman, R. A. (2020). The situation of situation research: Knowns and unknowns, *Current Directions in Psychological Science, 29*, 473-480.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. Journal of Research in Personality, 43, 137-145.

Roberts, B. W., & Davis, J. P. (2016). Young adulthood is the crucible of personality development. Emerging Adulthood, 4, 318-326.

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., & Mroczek, D. (2008). Personality trait change in adulthood. Current directions in Psychological Science, 17, 31-35.

Roberts, B. W., & Nickel, L. B. (2021). Personality development across the life course: A neo socioanalytic perspective*. In O.P. John & R.W. Robins (Eds). Handbook of personality theory and research (Chapter 11).* Guilford.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. Psychological Bulletin, 143, 117-141.

Roberts, B. W., Walton, K. & 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.

*Roberts, B. W., Walton, K., Bogg, T., & Caspi, A. (2006). De-investment in work and non-normative personality trait change in young adulthood. European Journal of Personality, 20, 461–474.

Roberts, B. W., Wood, D., & Smith, J. L. (2005). Evaluating five factor theory and social investment perspectives on personality trait development. *Journal of Research in Personality, 39,* 166-184.

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.

Roberts, B. W., & Yoon, L. B. (2022). Personality Psychology. *Annual Review of Psychology.*

*Romero, E., & Alonso, C. (2015). Hyperactive Behaviors from Childhood to Adolescence: Prospective Outcomes in a Sample of Spanish Children. International Journal of Psychological Studies, 7, p67.

*Rosander, P., & Bäckström, M. (2014). Personality traits measured at baseline can predict academic performance in upper secondary school three years late. Scandinavian Journal of Psychology, 55, 611–618.

Rowe, J. W., & Kahn, R. L. (2015). Successful aging 2.0: Conceptual expansions for the 21st century. The Journals of Gerontology: Series B, 70, 593-596.

*Rubinic, D. C. (2013). Effects of Temperament, Attachment, and Parental Sensitivity on the Development of ADHD. (Doctoral Dissertation). Duquesne University, Pittsburgh, PA.

*Salihovic, S., Özdemir, M., & Kerr, M. (2014). Trajectories of Adolescent Psychopathic Traits. Journal of Psychopathology and Behavioral Assessment, 36, 47–59.

*Schmits, E., & Quertemont, E. (2018). Components of social anxiety prevent cannabis use in adolescents. Journal of Substance Use, 23, 441–450.

Schuerger, J. M., Zarrella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. Journal of Personality and Social Psychology, 56, 777.

Schwaba, T. & Bleidorn, W. (2018). Individual differences in personality change across the adult lifespan. Journal of Personality, 86, 450-464.

Schwaba, T. & Bleidorn, W. (2019). Personality development across the transition to retirement. Journal of Personality and Social Psychology, 116, 651-665.

Schwaba, T. and Bleidorn, W. (2021). Internet use and cognitive engagement in older adulthood. Manuscript submitted for publication.

*Schwaba, T., & Bleidorn, W. (2018). Individual differences in personality change across the adult life span. Journal of Personality, 86, 450–464.

*Schwaba, T., Robins, R. W., Grijalva, E., & Bleidorn, W. (2019). Does Openness to Experience matter in love and work? Domain, facet, and developmental evidence from a 24-year longitudinal study. Journal of Personality, 87, 1074–1092.

Schwaba, T., Robins, R. W., Grijalva, E., & Bleidorn, W. (2019). Does openness to experience matter in love and work? Evidence from a 24-year longitudinal study (in press). Journal of Personality.

*Šerek, J., Lacinová, L., & Macek, P. (2012). Does family experience influence political beliefs? Relation between interparental conflict perceptions and political efficacy in late adolescence. Journal of Adolescence, 35, 577–586.

*Settles, R. E., Zapolski, T. C. B., & Smith, G. T. (2014). Longitudinal test of a developmental model of the transition to early drinking. Journal of Abnormal Psychology, 123, 141–151.

*Shulman, E. P., Harden, K. P., Chein, J. M., & Steinberg, L. (2016). The Development of Impulse Control and Sensation-Seeking in Adolescence: Independent or Interdependent Processes? Journal of Research on Adolescence, 26, 37–44.

*Sijtsema, J. J., Garofalo, C., Jansen, K., & Klimstra, T. A. (2019). Disengaging from Evil: Longitudinal Associations Between the Dark Triad, Moral Disengagement, and

Antisocial Behavior in Adolescence. Journal of Abnormal Child Psychology, 47, 1351–1365.

*Skinner, O. D., McHale, S. M., Wood, D., & Telfer, N. A. (2019). Gender-Typed Personality Qualities and African American Youth's School Functioning. Journal of Youth and Adolescence, 48, 680–691.

*Slagt, M., Dubas, J. S., Deković, M., Haselager, G. J. T., & van Aken, M. A. G. (2015). Longitudinal Associations between Delinquent Behaviour of Friends and Delinquent Behaviour of Adolescents: Moderation by Adolescent Personality Traits. European Journal of Personality, 29, 468–477.

*Slane, J. D., Klump, K. L., Donnellan, M. B., McGue, M., & Iacono, W. G. (2013). The Dysregulated Cluster in Personality Profiling Research: Longitudinal Stability and Associations With Bulimic Behaviors and Correlates. Journal of Personality Disorders, 27, 337–358.

*Slining, M. M., Adair, L., Goldman, B., Borja, J., & Bentley, M. (2009). Infant temperament contributes to early infant growth: A prospective cohort of African American infants. International Journal of Behavioral Nutrition and Physical Activity, 6, 51.

*Song, J.-H. (2015). Theory-of-Mind Development as an Antecedent and a Consequence of Social-Behavioral Development in Children (Doctoral Thesis). University of Michigan.

*Song, J.-H., Waller, R., Hyde, L. W., & Olson, S. L. (2016). Early Callous-Unemotional Behavior, Theory-of-Mind, and a Fearful/Inhibited Temperament Predict Externalizing Problems in Middle and Late Childhood. Journal of Abnormal Child Psychology, 44, 1205–1215.

Soto, C. J., & John, O. P. (2012). Development of Big Five Domains and Facets in Adulthood: Mean-Level Age Trends and Broadly Versus Narrowly Acting Mechanisms. Journal of Personality, 80, 881-914.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. Journal of Personality and Social Psychology, 113, 117

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. Journal of Personality and Social Psychology, 94, 718.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. Journal of personality and social psychology, 100, 330-348.

Specht, J., Bleidorn, W., Denissen, J.J.A., Hennecke, M., Hutteman, R., Kandler, C., Luhmann, M., Orth, U., Reitz, A., Zimmermann, J. (2014). What drives adult personality development? A comparison of theories and empirical evidence. European Journal of Personality, 28, 216-230.

Specht, J., Egloff, B., & Schmukle, S.C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. Journal of Personality and Social Psychology, 101, 862–882.

*Spence, R., Owens, M., & Goodyer, I. (2013). The Longitudinal Psychometric Properties of the EAS Temperament Survey in Adolescence. Journal of Personality Assessment, 95, 633–639.

*Spengler, M., Gottschling, J., & Spinath, F. M. (2012). Personality in childhood – A longitudinal behavior genetic approach. Personality and Individual Differences, 53, 411–416.

Stanek, K. C., & Ones, D. S. (2018). Taxonomies and compendia of cognitive ability and personality constructs and measures relevant to industrial, work and organizational psychology. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work & organizational psychology: Personnel psychology and employee performance (pp. 366–407)*. Sage Reference.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods, 5, 60-78.

*Stavropoulos, V., Kuss, D., Griffiths, M., & Motti-Stefanidi, F. (2016). A longitudinal study of adolescent internet addiction: The role of conscientiousness and classroom hostility. Journal of Adolescent Research, 31, 442–473.

*Stavropoulos, V., Moore, K. A., Lazaratou, H., Dikeos, D., & Gomez, R. (2017). A multilevel longitudinal study of obsessive compulsive symptoms in adolescence: Male gender and emotional stability as protective factors. Annals of General Psychiatry, 16, 42.

*Stephan, Y., Sutin, A. R., Luchetti, M., Caille, P., & Terracciano, A. (2019). Cigarette smoking and personality change across adulthood: Findings from five longitudinal samples. Journal of Research in Personality, 81, 187–194.

*Suen, K. S., Lai, Y., Ho, S. M. Y., Cheung, L. K., & Choi, W. S. (2018). A longitudinal evaluation of psychosocial changes throughout orthognathic surgery. PLOS ONE, 13, e0203883.

*Sulik, M. J. (2013). Differential prediction of Internalizing and Externalizing Symptomatology from Temperament and parenting. (Doctoral Dissertation). Arizona State University, Tempe, AZ.

*Sun, J.-T. (2011). Major life goals of college students: An investigation of personality traits, vocational interests, and values. (Doctoral Dissertation). University of Illinois at Urbana-Champaign, Champaign, IL.

Sutin, A. R., Zonderman, A. B., Ferrucci, L., & Terracciano, A. (2013). Personality traits and chronic disease: Implications for adult personality development. Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68, 912-920.

*Taylor, M. J., Charman, T., Robinson, E. B., Plomin, R., Happé, F., Asherson, P., & Ronald, A. (2013). Developmental associations between traits of autism spectrum disorder and attention deficit hyperactivity disorder: A genetically informative, longitudinal twin study. Psychological Medicine, 43, 1735–1746.

*Taylor, Z. E., Doane, L. D., & Eisenberg, N. (2014). Transitioning From High School to College: Relations of Social Support, Ego-Resiliency, and Maladjustment During Emerging Adulthood. Emerging Adulthood, 2, 105–115.

*Terrett, G., O'Connor, M., Hawkins, M. T., Sanson, A., & Smart, D. (2012). Longitudinal Antecedents of School Bonding in Adolescence. The Australian Educational and Developmental Psychologist, 29, 107–128.

*Thartori, E., Zuffianò, A., Pastorelli, C., Di Giunta, L., Lunetti, C., Lansford, J. E., … Caprara, G. V. (2018). The interactive effects of maternal personality and adolescent temperament on externalizing behavior problem trajectories from age 12 to 14. Personality and Individual Differences, 134, 301–307.

*Thomas, C. A. (2006). Childhood temperament as a predictor of substance use in early adolescence. (Doctoral Dissertation). University of Pittsburgh,Pittsburgh, PA.

*Thompson, S. F., Zalewski, M., & Lengua, L. J. (2014). Appraisal and coping styles account for the effects of temperament on pre-adolescent adjustment. Australian Journal of Psychology, 66, 122–129.

*Tian, X., Wei, D., Du, X., Wang, K., Yang, J., Liu, W., … Qiu, J. (2016). Assessment of trait anxiety and prediction of changes in state anxiety using functional brain imaging: A test–retest study. NeuroImage, 133, 408–416.

Tucker-Drob, E. M., & Briley, D. A. (2019). Theoretical concepts in the genetics of personality development. The Handbook of Personality Development, 40-58.

*Tuvblad, C., Bezdjian, S., Raine, A., & Baker, L. A. (2013). Psychopathic Personality and Negative Parent-to-Child Affect: A Longitudinal Cross-lag Twin Study. Journal of Criminal Justice, 41, 331–341.

*Vaillancourt, T., & Brittain, H. (2019). Longitudinal associations among primary and secondary psychopathic traits, anxiety, and borderline personality disorder features across adolescence. Personality Disorders: Theory, Research, and Treatment, 10, 354–364.

*Valle, M. F., Huebner, E. S., & Suldo, S. M. (2006). An analysis of hope as a psychological strength. Journal of School Psychology, 44, 393–406.

*Van den Akker, A. L., Deković, M., Asscher, J., & Prinzie, P. (2014). Mean-level personality development across childhood and adolescence: A temporary defiance of the maturity principle and bidirectional associations with parenting. Journal of Personality and Social Psychology, 107, 736–750.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3, 4-70.

*Van der Graaff, J., Carlo, G., Crocetti, E., Koot, H. M., & Branje, S. (2018). Prosocial Behavior in Adolescence: Gender Differences in Development and Links with Empathy. Journal of Youth and Adolescence, 47, 1086–1099.

*van der Voort, A., Linting, M., Juffer, F., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2013). Delinquent and aggressive behaviors in early-adopted adolescents: Longitudinal predictions from child temperament and maternal sensitivity. Children and Youth Services Review, 35, 439–446.

*Van Heel, M., Bijttebier, P., Colpin, H., Goossens, L., Van Den Noortgate, W., Verschueren, K., & Van Leeuwen, K. (2019). Investigating the interplay between adolescent personality, parental control, and externalizing problem behavior across adolescence. Journal of Research in Personality, 81, 176–186.

van Scheppingen, M. A., Jackson, J. J., Specht, J., Hutteman, R., Denissen, J. J. A., & Bleidorn, W. (2016). Personality development during the transition to parenthood: A test of social investment theory. Social Psychological and Personality Science, 7, 452-462

*Vanwoerden, S., Leavitt, J., Gallagher, M. W., Temple, J. R., & Sharp, C. (2019). Dating violence victimization and borderline personality pathology: Temporal associations from late adolescence to early adulthood. Personality Disorders: Theory, Research, and Treatment, 10, 132–142.

*Vecchione, M., Alessandri, G., Barbaranelli, C., & Caprara, G. (2013). A Longitudinal Investigation of Egoistic and Moralistic Self-Enhancement. Journal of Personality Assessment, 95, 506–512.

*Vecchione, M., Alessandri, G., Roccas, S., & Caprara, G. V. (2019). A look into the relationship between personality traits and basic values: A longitudinal investigation. Journal of Personality, 87, 413–427.

*Vrieze, S. I., Vaidyanathan, U., Hicks, B. M., Iacono, W. G., & McGue, M. (2014). The Role of Constraint in the Development of Nicotine, Marijuana, and Alcohol Dependence in Young Adulthood. Behavior Genetics, 44, 14–24.

*Wagers, K. B., & Kiel, E. J. (2019). The influence of parenting and temperament on empathy development in toddlers. Journal of Family Psychology, 33, 391–400.

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements. Schmollers Jahrbuch: Zeitschrift fur Wirtschafts- und Sozialwissenschaften/ Journal of Applied Social Science Studies, 127: 139–69.

Wagner, J., Becker, M., Lüdtke, O., & Trautwein, U. (2015). The first partnership experience and personality development: A propensity score matching study in young adulthood. Social Psychological and Personality Science, 6, 455-463.

Wagner, J., Ram, N., Smith, J., & Gerstorf, D. (2016). Personality trait development at the end of life: Antecedents and correlates of mean-level trajectories. Journal of Personality and Social Psychology, 111, 411.

*Wagner, N. J., Waller, R., Flom, M., Ronfard, S., Fenstermacher, S., & Saudino, K. (2020). Less imitation of arbitrary actions is a specific developmental precursor to callous–unemotional traits in early childhood. Journal of Child Psychology and Psychiatry, 61, 818–825.

*Waller, R., Gardner, F., Hyde, L. W., Shaw, D. S., Dishion, T. J., & Wilson, M. N. (2012). Do harsh and positive parenting predict parent reports of deceitful-callous behavior in early childhood?: Early parenting predicts deceitful-callousness. Journal of Child Psychology and Psychiatry, 53, 946–953.

*Wang, F. L., Eisenberg, N., & Spinrad, T. L. (2019). Bifactor model of effortful control and impulsivity and their prospective prediction of ego resiliency. Journal of Personality, 87, 919–933.

*Wang, Y., & Xia, L. (2019). The longitudinal relationships of interpersonal openness trait, hostility, and hostile attribution bias. Aggressive Behavior, 45, 682–690.

*Wei, J. (2018). Academic Contingent Self-Worth of Adolescents in Mainland China: Distinguishing Between Success and Failure as a Basis of Self-Worth. (Doctoral Dissertation). The Chinese University of Hong Kong (Hong Kong).

*Wettstein, M., Wahl, H.-W., Siebert, J., & Schröder, J. (2019). Still more to learn about late-life cognitive development: How personality and health predict 20-year cognitive trajectories. Psychology and Aging, 34, 714–728.

*Wetzel, E., & Robins, R. W. (2016). Are parenting practices associated with the development of narcissism? Findings from a longitudinal study of Mexican-origin youth. Journal of Research in Personality, 63, 84–94.

*Weybright, E. H., Caldwell, L. L., Ram, N., Smith, E. A., & Wegner, L. (2015). Boredom Prone or Nothing to Do? Distinguishing Between State and Trait Leisure Boredom and Its Association with Substance Use in South African Adolescents. Leisure Sciences, 37, 311–331.

*Whipp, A. M., Korhonen, T., Raevuori, A., Heikkilä, K., Pulkkinen, L., Rose, R. J., … Vuoksimaa, E. (2019). Early adolescent aggression predicts antisocial personality disorder in young adults: A population-based study. European Child & Adolescent Psychiatry, 28, 341–350.

*Wille, B., & De Fruyt, F. (2014). Vocations as a source of identity: Reciprocal relations between Big Five personality traits and RIASEC characteristics over 15 years. Journal of Applied Psychology, 99, 262–281.

*Windle, M., & Windle, R. C. (2006). Adolescent temperament and lifetime psychiatric and substance abuse disorders assessed in young adulthood. Personality and Individual Differences, 41, 15–25.

*Windsor, T. D., Pearson, E. L., & Butterworth, P. (2012). Age group differences and longitudinal changes in approach–avoidance sensitivity: Findings from an 8-year longitudinal study. Journal of Research in Personality, 46, 646–654.

*Wood, D., & Roberts, B. W. (2006). Cross-Sectional and Longitudinal Tests of the Personality and Role Identity Structural Model (PRISM). Journal of Personality, 74, 779–810.

*Wortman, J., Lucas, R. E., & Donnellan, M. B. (2012). Stability and change in the Big Five personality domains: Evidence from a longitudinal study of Australians. Psychology and Aging, 27, 867–874.

*Wright, A. G. C., Pincus, A. L., & Lenzenweger, M. F. (2012). Interpersonal Development, Stability, and Change in Early Adulthood: Interpersonal Development. Journal of Personality, 80, 1339–1372.

*Xia, L.-X., Gao, X., Wang, Q., & Hollon, S. D. (2014). The relations between interpersonal self-support traits and emotion regulation strategies: A longitudinal study. Journal of Adolescence, 37, 779–786.

*You, J., Lin, M.-P., & Leung, F. (2015). A Longitudinal Moderated Mediation Model of Nonsuicidal Self-injury among Adolescents. Journal of Abnormal Child Psychology, 43, 381–390.

*Zavos, H. M. S., Eley, T. C., McGuire, P., Plomin, R., Cardno, A. G., Freeman, D., & Ronald, A. (2016). Shared Etiology of Psychotic Experiences and Depressive Symptoms in Adolescence: A Longitudinal Twin Study. Schizophrenia Bulletin, 42, 1197–1206.

*Zavos, H. M. S., Rijsdijk, F. V., & Eley, T. C. (2012). A Longitudinal, Genetically Informative, Study of Associations Between Anxiety Sensitivity, Anxiety and Depression. Behavior Genetics, 42, 592–602.

*Zerwas, S., Holle, A. V., Watson, H., Gottfredson, N., & Bulik, C. M. (2014). Childhood anxiety trajectories and adolescent disordered eating: Findings from the NICHD study of early child care and youth development: Childhood Anxiety Trajectories. International Journal of Eating Disorders, 47, 784–792.

*Zhang, J., & Ziegler, M. (2018). Why do personality traits predict scholastic performance? A three-wave longitudinal study. Journal of Research in Personality, 74, 182–193.

*Zhou, Q., Main, A., & Wang, Y. (2010). The relations of temperamental effortful control and anger/frustration to Chinese children's academic achievement and social adjustment: A longitudinal study. Journal of Educational Psychology, 102, 180–196.

*Zohar, A. H., Zwir, I., Wang, J., Cloninger, C. R., & Anokhin, A. P. (2019). The development of temperament and character during adolescence: The processes and phases of change. Development and Psychopathology, 31, 601–617.

*Zubizarreta, A., Calvete, E., & Hankin, B. L. (2019). Punitive Parenting Style and Psychological Problems in Childhood: The Moderating Role of Warmth and Temperament. Journal of Child and Family Studies, 28, 233–244.

*Zupančič, M., Sočan, G., & Kavčič, T. (2009). Consistency in adult reports on child personality over the pre-school years. European Journal of Developmental Psychology, 6, 455–480.

*Zupanèiè, M., & Kavèiè, T. (2005). Child personality measures as contemporaneous and longitudinal predictors of social behaviour in pre-school. Horizons of Psychology, 14, 17–33.