# Differential Stability and Individual Growth Trajectories of Big Five and Affective Traits During Young Adulthood

**Jatin G. Vaidya**[1], **Elizabeth K. Gray**[2], **Jeffrey R. Haig**[3], **Daniel K. Mroczek**[4], and **David Watson**[1]

[1]University of Iowa

[2]North Park University

[3]Clinton Community College

[4]Purdue University

## Abstract

Big Five and affective traits were measured at three assessments when participants were on average 18, 21, and 24 years old. Rank-order stability analyses revealed that stability correlations tended to be higher across the second compared to the first retest interval; however, affective traits consistently were less stable than the Big Five. Median stability coefficients for the Big Five increased from .62 (Time 1 vs. Time 2) to .70 (Time 2 to Time 3); parallel increases also were observed for measures of negative affectivity (median rs=.49 and .55, respectively) and positive affectivity (median rs=.48 and .57, respectively). Growth curve analyses revealed significant change on each of the Big Five and affective traits, although many of the scales also showed significant variability in individual trajectories. Thus, rank-order stability is increasing for a range of personality traits, although there also is significant variability in change trajectories during young adulthood.

The question of personality stability is of central importance to laypersons, trait psychologists, and behavioral scientists. For several years, much of the research in this area simply sought to establish whether or not traits are stable. In recent years, however, researchers have recognized that the question of stability is intricately tied to how stability is measured (e.g., Caspi & Roberts, 1999). Furthermore, investigators have sought to ask more sophisticated questions about patterns of personality stability over time (Fraley & Roberts, 2005) or across measures (Vaidya, Gray, Haig, & Watson, 2002). In the present study, we report the results of the third assessment of our ongoing Iowa Longitudinal Personality Project (ILPP; see Vaidya et al., 2002). Almost 400 participants completed a Big Five and trait-affect measure at two points in time and nearly 300 participants completed these measures at three points corresponding to when participants were 18, 21, and 24 years old, on average. Although other studies have been published using Big Five or Big Three measures of personality in this age group, this is the first study that measures both Big Five and affective traits across two approximately equal intervals over the course of young adulthood. As in our previous study, we examine differential stability—that is, differences in rank-order stability—of Big Five and affective traits across both time periods. Furthermore, capitalizing on the three-wave design, we use growth curve modeling to characterize overall sample trajectories as well individual differences in these trajectories over time.

Correspondence concerning this article should be addressed to Jatin G. Vaidya, Department of Psychiatry, University of Iowa, Iowa City, IA 52242. E-mail:jatin-vaidya@uiowa.edu.
M. Haig is no longer affiliated with Clinton Community College.

## Differential Stability

Following the logic of classical test theory, test-retest correlations have long been used as an index of the reliability of an instrument. Because dispositional constructs should not change over short periods of time, any difference in scores across time points must be attributable to measurement error. Thus, for years, researchers focused primarily on establishing whether or not a measure showed test-retest reliability. Furthermore, trait psychologists focused on establishing whether or not traits were stable over time without attempting to quantify differences in the rank-order stability of different traits (i.e., differential stability). Consequently, the stability literature, paralleling the early self-other agreement literature, largely has taken a dichotomous approach to stability—traits are either stable or not (Vaidya et al., 2002; Watson, 2004). This way of thinking obscures important quantitative differences in stability between scales.

Conley (1984) conducted one of the earliest systematic investigations of differential stability. Although the principal finding was that intelligence was more stable than personality traits, which, in turn, were more stable than attitudinal traits, this article also found evidence of differential stability within personality traits. Extraversion was the most stable of all the personality traits included in the analysis. Since then, two other meta-analyses have found Extraversion to be the most stable and Neuroticism to be one of the least stable of the Big Five traits (Roberts & DelVecchio, 2000; Schuerger, Zarrella, & Hotz, 1989).

### Trait Affect

Our previous approach, and the one we continue with here, is to compare the differential stability of the Big Five versus affective traits. It is now well established that four of the Big Five dimensions have strong links to trait affectivity (Watson, 2000; Watson & Clark, 1992). Neuroticism is strongly and broadly linked to negative affectivity while Extraversion is strongly associated with positive affectivity. Watson, Wiese, Vaidya, and Tellegen (1999) analyzed relevant data across combined samples with an overall N of 4,457. They obtained a correlation of .58 between Neuroticism and the trait form of the PANAS-X (Watson & Clark, 1999) General Negative Affect (NA) scale. Extraversion had a parallel correlation of .51 with General Positive Affect (PA). Indeed, the strong empirical associations have led to the conclusion that Neuroticism and Extraversion represent temperament-based traits that reflect individual differences in the propensity to experience negative and positive affective states, respectively (Tellegen, 1985). Additionally, both Agreeableness and Conscientiousness have substantial correlations with specific scales of the PANAS-X. Agreeableness, though not broadly correlated with PA or NA, is substantially (negatively) correlated with trait Hostility (low scorers on this trait report greater amounts of anger and irritability), whereas Conscientiousness is strongly correlated with Attentiveness (high scorers on this scale report being alert and highly determined). Openness is the sole Big Five dimension that is only weakly correlated with affectivity (Watson & Clark, 1992).

Despite these substantial correlations, trait affectivity appears to be significantly less stable over time than the Big Five. For instance, Watson and Walker (1996) obtained stability coefficients for trait PA and NA ranging from .36 to .46 over retest intervals of approximately 6–7 years in a young adult sample. The stability analyses from the first ILPP retest produced similar, although slightly larger, stability coefficients, most likely because of the shorter time interval (Vaidya et al., 2002). Specifically, the five PANAS-X NA scales had a median stability coefficient of .49, whereas the four PA scales had a median stability coefficient of .52. Further analyses revealed that Neuroticism (which was assessed using the Big Five Inventory [BFI]; John & Srivastava, 1999), with a stability coefficient of .61, was significantly more stable than all of the NA scales; in parallel fashion, BFI Extraversion, with an impressive stability coefficient of .72, was easily more stable than all of the PA scales (Vaidya et al., 2002).

### Differential Stability and Scale Content

Based on these results, Vaidya and colleagues (2002) concluded that the lower levels of stability for affectivity may reflect true differences in trait stability, as well as differences in the instructional set given to participants when responding to items on the two inventories (i.e., format differences). The PANAS-X format encourages respondents to access past affective experiences by asking them to rate how they generally feel. In contrast, the BFI format involves evaluating the extent to which each item is consistent with one's general self-image.

From one perspective, because affective states may change dramatically from moment to moment, (Izard, 1991), affective traits, by their very nature, should necessarily show lower levels of temporal stability. Therefore, it is not surprising that the BFI Extraversion scale, which contains very little affective content, shows higher levels of stability than the PANAS-X PA scales. However, content differences cannot explain differences between the BFI Neuroticism scale—which contains essentially all affective items—and the PANAS-X NA scales. In order to examine how item content influences longitudinal stability, we utilize item-level content analyses on the BFI performed by Pytlik Zillig, Hemenover, and Dienstbier (2002). In this study, raters determined the percentage of affective, cognitive, and behavioral content for items in commonly used Big Five inventories, including the BFI. If affective content contributes to lower retest stability, then items with the most affective content should be associated with lower stability levels. We test this possibility by correlating the percentage of affective, behavioral, and cognitive content—using data provided to us by L. M. Pytlik Zillig (personal communication)—with the BFI $T_{12}$ and $T_{23}$ stability coefficients. We expected affective content to be broadly correlated with lower levels of stability.

### Differential Stability Across Multiple Assessments

In the present study, we take advantage of our three-wave design to compare the $T_{12}$ and $T_{23}$ stability coefficients. Fraley and Roberts (2005) emphasize the importance of examining the pattern of rank-order stability coefficients over time in order to better understand the processes influencing stability and change. They point out that a single stability coefficient tells us little about how personality changes or the factors that influence continuity and change. This is, in part, because researchers rarely make specific point estimates about what a stability coefficient will be. For instance, is a .65 coefficient evidence for change or stability? Multiwave designs, however, make it possible to determine how stability is changing, if at all, over time.

Using this three-panel design, we sought to determine if the lower PANAS-X stabilities are developmentally specific. Because the initial testing occurred shortly after the beginning of the fall semester—and for a majority of the participants, shortly after beginning college— this may have been a particularly eventful and turbulent time. Because the PANAS-X scales are more sensitive to major life events (Vaidya et al., 2002), the PANAS-X stability levels may have been lower across the first retest simply because of the timing of the first assessment. Put another way, the PANAS-X scales simply may have a different time course for achieving greater stability. If this is true, BFI and PANAS-X stability levels should be slowly becoming more similar over time. This pattern of results would also suggest that differences between the BFI and PANAS-X in scale content, and not format (e.g., if participants are asked to indicate whether statements are consistent with their self-image, as in the BFI, or if they are asked to indicate how often they experience various emotional states, as in the PANAS-X), primarily are responsible for the differential stability we observed at the initial retest.

### Young Adulthood

Until recently, there was little systematic research on personality development during young adulthood. This is especially surprising considering the fact that a great deal of personality research relies on undergraduate students. It is becoming increasingly clear, however, that

young or emerging adulthood represents a distinct developmental period full of significant life changes (Arnett, 2000). This time period, which roughly encompasses ages 18 to 25 (Arnett, 2000), has recently been the focus of several longitudinal studies (Neyer & Asendorpf, 2001; Roberts, Caspi, & Moffitt, 2001; Robins, Fraley, Roberts, & Trzesniewski, 2001; Vollrath, 2000).

Psychologically, this period is especially interesting because of the many changes that occur during this time. People change residence more during this period than any other. This includes moving away from parents to live in dormitories, fraternities, and sororities, with roommates, or with romantic partners (Rindfuss, 1991). Furthermore, when they move, young adults, compared to any other age group, will move further away from their previous residence (Rindfuss, 1991). This period is marked by numerous social and romantic changes as well; while first romantic experiences often occur during adolescence, deeper, more significant relationships develop during young adulthood (Arnett, 2000). Furthermore, prevalence rates for several psychiatric disorders are highest during young adulthood for men and women (Newman et al., 1996).

### Individual and Mean-Level Change

Mean-level change is essentially a measure of systematic change— that is, how the sample changes as a whole. Young adulthood is associated with striking mean-level changes on several different personality variables. In a meta-analysis, Roberts and colleagues reported that some of the most significant mean-level changes occurred during young adulthood (Roberts, Walton, & Viechtbauer, 2006). Conscientiousness (Neyer & Asendorpf, 2001; Roberts et al., 2006; Robins et al., 2001; Vaidya et al., 2002; Vollrath, 2000), Agreeableness (Roberts et al., 2006; Robins et al., 2001; Vaidya et al., 2002; Vollrath, 2000), and Constraint (McGue, Bacon, & Lykken, 1993; Roberts et al., 2001) tend to increase during young adulthood. In addition, most studies find decreases in Neuroticism-related traits (McGue et al., 1993; Neyer & Asendorpf, 2001; Roberts et al., 2001; Roberts et al., 2006; Robins et al., 2001). Mean scores on Openness tend to increase throughout adolescence and early young adulthood but then show little change for several decades (Roberts et al., 2006). Finally, while changes in Extraversion have been less consistent, Roberts and colleagues (2006) found that these inconsistencies reflect the fact that while assertiveness and confidence increased during young adulthood, gregariousness showed little change throughout the life course (Roberts et al., 2006).

Our three-wave design made it possible to determine which period of young adulthood is associated with the greater amount of mean-level change. For many individuals, young adulthood can be conceptualized as having at least two distinct phases—one that coincides with the college years and another occurring after graduation. Attending college has significant psychosocial consequences and likely contributes to the relatively recent emergence of young adulthood as a distinct developmental epoch. For instance, many people choose to forgo marriage or long-term partnerships until after college. Additionally, because full-time work is still in the distant future, this time period is ripe for experimentation and identity exploration (Arnett, 2000).

Because Agreeableness and Conscientiousness appear to show moderate-to-strong increases throughout young adulthood, we expected these traits to show significant change across both retests. Increases in Openness appear to occur during a relatively small time window during young adulthood. Therefore, we expected much weaker changes on this trait across the second compared to the first retest interval. Finally, by conducting assessments at three time points, we hoped to resolve some of the inconsistencies in the literature on how certain traits change during this time period. For instance, contrary to some other studies, we found no change in Neuroticism in the first ILPP retest (Vaidya et al., 2002). Other studies with longer retest intervals did find evidence of a mean-level decrease in Neuroticism (e.g., Neyer & Asendorpf,

2001; Robins et al., 2001; Roberts et al., 2001). Therefore, we predicted Neuroticism scores would show a significant decline from Time 2 to Time 3 in our sample. Because findings with Extraversion are less consistent and may be idiosyncratic to the specific content of a given scale, we did not make specific predictions on this trait.

### Individual-Level Stability

In recent years, developmental trait psychologists have come to realize that there are different ways of measuring personality change and that each approach may yield a different conclusion. Mean-level analyses characterize the degree of systematic change occurring on a given trait over time. Thus, weak mean-level findings may mask significant individual-level effects if individuals are changing in opposite directions. For instance, in our first retest, the PANAS-X PA scales showed little mean-level change while the NA scales showed significant decreases (Vaidya et al., 2002). Do the weak mean-level findings for PA reflect little individual-level change or, rather, change in different directions?

A number of recent studies have conducted individual-level analyses to characterize the type and magnitude of change for each individual using reliable change index (RCI) scores ( Jacobson & Truax, 1991; Roberts et al., 2001). Using the RCI method, most studies have demonstrated that individual-level changes coincided with mean-level findings. For instance, Conscientiousness shows dramatic mean-level increases and, at the individual-level, many people show significant increases on this trait while very few show significant decreases (Robins et al., 2001; Roberts et al., 2001; Vaidya et al., 2002). In contrast, however, Vaidya and colleagues (2002) reported little mean-level change on Joviality. Individual-level results revealed that this was because some individuals increased on Joviality, whereas others decreased.

In the present study, we take advantage of our three-wave design and conduct individual growth curve modeling (Raudenbush & Bryk, 2002; Rogosa, Brandt, & Zimowski, 1982; Willett & Sayer, 1994) to characterize individual differences in trajectories on each of the Big Five and trait-affect scales. Individual growth modeling is a type of multilevel model (also known as mixed, random effects, or hierarchical linear modeling) that estimates individual-level trajectories as well as an overall, sample trajectory (linear growth curve results for the sample are essentially equivalent to the mean-level analyses using repeated measures ANOVA). The results of the growth curve modeling indicate whether there was significant variability across individuals in the trajectories. Growth curve analysis offers a number of advantages over ANOVA techniques. Most notably, because growth curve analyses can accommodate missing data points, we were able to incorporate all individuals who completed at least the first two assessments (almost 400 participants) into the analyses.

It is difficult to make specific predictions because no studies have investigated individual-level change using growth curve analyses of established Big Five or trait affect measures in young adult samples. Studies with older adults, however, have obtained evidence for significant variability in slopes for the Big Five (Mroczek & Spiro, 2003; Small, Hertzog, Hultsch, & Dixon, 2003; Terracciano, McCrae, Brant, & Costa, 2005) and PA and NA (Griffin, 2004; Griffin, Mroczek, & Spiro, in press). Because our subjects are likely to have experienced a number of important life events at distinct time points, we expected there would be significant variability in the slopes in our sample as well.

## METHODS

### Participants

Participants in this study were individuals taking part in the Iowa Longitudinal Personality Project (ILPP). At Time 1, all participants were students enrolled in an introductory psychology class at the University of Iowa. They completed a battery of questionnaires in partial fulfillment of a class research exposure requirement. Time 1 data were collected from a total of 759 undergraduates in September of 1996. A total of 394 participants (51.6% of the initial sample) completed all of the questionnaires at Time 2, slightly more than 2.5 years later in the spring of 1999. Specific demographic information was unavailable on the original pool of participants. However, this was available at the second testing. The sample at Time 2 consisted of 299 women and 95 men.[1] The mean age of the participants was 21.1 years. A large number of participants were still attending college, and most of these students (75.6%) were still enrolled at the University of Iowa.

The third testing was undertaken in the spring of 2002, approximately 3 years after the second assessment. Only individuals who participated in the Time 2 assessment were contacted. Data were collected on 312 of the Time 2 participants, representing 79.6% of the Time 2 sample. However, because of missing data on the BFI at both Time 2 and Time 3, we were able to obtain complete data on 299 participants (of which 224 were female). At Time 3, participants were 24 years old, on average. There were systematic changes in relationship status, as more than 20% of the sample now was either married or engaged (see Table 1), as opposed to only 5% at Time 2. As would be expected, however, the biggest difference was in student and employment status. At Time 2, most of the participants (81.6%) were still undergraduate students. In marked contrast, only 6.4% of the participants still were undergraduates at Time 3. Thus, more than 80% of the participants were either working full time (61.2%) or were graduate students (20.1%) at Time 3.

### Procedure

Participants were mailed the questionnaires in early 2002. Mailing addresses were obtained through the university or by using information provided by some of the participants at Time 2. Most of the questionnaires were returned by May of 2002. However, a few questionnaires were accepted as late as June 2002. Participants were reimbursed $20 for returning the completed questionnaires.

### Measures

**Big Five—**Personality ratings on the Big Five were obtained using the BFI (John & Srivastava, 1999). The version of the BFI used in this study contains two 8-item scales that measure Extraversion and Neuroticism, two 9-item scales that measure Agreeableness and Conscientiousness, and one 10-item scale that assesses Openness. Participants are asked to rate each item using a 5-point scale (1= *very uncharacteristic of myself*, 5= *very characteristic of myself*). The BFI's reliability and validity are well documented. Watson, Hubbard, and Wiese (2000) report coefficient alphas ranging from .76 to .85 for the five scales. In addition, the BFI scales are highly correlated with other Big Five measures. For instance, Watson and Hubbard (1996) report convergent correlations between the scales of the BFI and the NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) ranging between .68 (Openness) and .85 (Conscientiousness).

---

[1]InVaidya et al. (2002) we reported the results of 392 participants because 2 participants had not returned their questionnaires in time for the first retest analyses. Also, at the first retest, one female subject was incorrectly entered as a male; thus, Vaidya et al., reported a sample of 96 males.

**Trait affectivity**—Individual differences in trait affectivity were assessed using the PANAS-X (Watson & Clark, 1999). This 60-item instrument asks individuals to indicate on a 5-point scale (1= *very slightly or not at all*, 2= *a little*, 3=*moderately*, 4= *quite a bit*, 5=*extremely*) "to what extent you generally feel this way, that is, how you feel on average". The PANAS-X includes two 10-item scales that measure the general dimensions of PA and NA. Examples of PA terms include *active*, *alert*, and *interested*; NA terms include *afraid*, *irritable*, and *upset*. In addition, this instrument also contains 11 factor-analytically derived scales that measure more specific moods and emotions. The PANAS-X includes four scales that measure specific negative emotions that are strong markers of the higher order NA dimension (Watson & Clark, 1999): Fear (6 items; e.g., *scared, frightened*), Sadness (5 items; e.g., *lonely, blue*), Guilt (6 items; e.g., *blameworthy, angry at self*) and Hostility (6 items; e.g., *angry, scornful*). The PANAS-X also includes three scales that measure more specific positive emotional states that are strongly linked to the general PA dimension: Joviality (8 items; e.g., *happy, enthusiastic*), Self-Assurance (6 items; e.g., *proud, confident*), and Attentiveness (4 items; e.g., *alert, concentrating*). Finally, four scales assess affective states that are less strongly and consistently related to the two higher order dimensions: Shyness (4 items; e.g., *bashful, timid*), Fatigue (4 items; e.g., *sluggish, drowsy*), Serenity (3 items; e.g., *calm, relaxed*), and Surprise (3 items; e.g., *surprised, amazed*).

The PANAS-X is a commonly used measure of affectivity and its reliability and validity are well documented. For instance, Watson, Clark, and Tellegen (1988) reported coefficient alphas for the higher order scales ranging from .84 to .87 for Negative Affect and from .86 to .90 for Positive Affect. Additionally, Watson and Clark (1997) reported median coefficient alphas ranging from .76 (Serenity) to .93 ( Joviality) for the 11 lower-order scales computed across 11 samples.

## Data Analysis

We report rank-order correlations and mean-level change (using repeated measures ANOVA) on the complete sample of 299 participants. To estimate sample and individual-level trajectories, we utilized growth curve modeling implemented with SAS Proc Mixed. Because growth curve modeling can accommodate missing data, we incorporated individuals who completed at least the first two assessments. Individuals with only two data points do not contribute to estimation of variability in slope and change but contribute to the fixed effects (Singer & Willet, 2003). Growth curve models were fit for each of the BFI and PANAS-X scales yielding estimates of fixed effects, which describe the slope and intercept for the entire sample, and random effects, which describe the variability (i.e., individual differences) in intercepts and slopes, and whether they significantly differ from zero. Age was centered on the grand mean (i.e., age 21) to avoid artificially inflating associations between intercept and slope (Rogosa & Willett, 1985; Willett, 1988). A formal definition of the model using Extraversion can be expressed as:

$$\text{Extraversion}_{ij} = \pi_{0i} + \pi_{1i}(\text{age}_{ij}) + \epsilon_{ij} \tag{1}$$

The amount of Extraversion for individual $i$ at measurement occasion $j$ is a function of the person's age at that measurement occasion ($\text{age}_{ij}$). The intercept, $\pi_{0i}$, is the predicted amount of Extraversion at age 21. The linear coefficient, $\pi_{1i}$, is the rate of change (slope); it is the predicted annual amount of change in extraversion for person $i$. $\epsilon_{ij}$ represents the errors for each person $i$ at occasion $j$. In a sample, each participant's trajectory is described by this equation. Together, these intercepts and slopes define the overall, sample-level intercept and slope, that is, the fixed effects. The variability in intercepts and slopes across $i$ persons are the

random effects. The fixed effect for the intercept is the estimated mean at age 21, and the fixed effect for the slope is the estimated change per year.

## RESULTS

### Attrition Analyses

Almost 80% of the Time 2 participants completed the questionnaires at Time 3. In spite of this relatively high retention rate, it is still important to determine if there was evidence of differential attrition. Focusing initially on demographic variables, we determined if individuals who participated at all three assessments differed from those who completed only the first two assessments. Retest participants (RPs) were no older than nonparticipants (NPs), $t$ (384) 5 .05, $p$=.96. In addition, men and women were equally likely to continue participating in the study, $\chi^2$ (1, $N$=394)=.64, $p$=.42.

Further attrition analyses examined how RPs and NPs differed on the BFI and PANAS-X General PA and General NA scales at Time 1. These analyses were conducted using an ANOVA, with participation status (1=participant at Time 1, 2, and 3; 2= participant at Time 1 and 2 but not Time 3; 3=participant at Time 1 but not Time 2 and Time 3) as the between-subjects factor. Significant effects were obtained for Agreeableness and Conscientiousness but not the other BFI or PANAS-X scales. Post hoc comparisons revealed that individuals who participated only at Time 1 ($M$=33.35) were significantly less agreeable initially than individuals who participated at Time 1 and Time 2 ($M$=35.23; $t$(454)=3.04, $p$=.002) and those who participated at all three assessments ($M$=34.62; $t$(658)=3.00, $p$=.003). However, individuals who participated at Time 1 and 2, but not Time 3, did not differ significantly from those who participated at all three assessments on Time 1 Agreeableness ($t$(392)=.97, $p$=.33). In terms of Conscientiousness, subjects who participated only at Time 1 ($M$ =30.54) scored significantly lower than those who completed all three assessments ($M$=31.68; $t$(658)=2.89, $p$=.004) but did not differ from those who participated only at Time 1 and Time 2 ($M$=30.71; $t$(454)=.28, $p$=.78). There also was no difference on this scale between individuals who only completed the first two rounds of questionnaires and those who participated at all three assessments ($t$(392)=.97, $p$=.33).

Thus, RPs were generally more agreeable and conscientious than NPs. Therefore, the participants in this study do not constitute a completely representative sample of the original participants. The results with this retest sample may be somewhat uncharacteristic of those that would be obtained with a fully representative sample. In fact, subsequent mean-level analyses may actually suggest somewhat weaker changes in this sample on Agreeableness and Conscientiousness than actually occurred in light of the fact that individuals who completed all three assessments initially scored higher on these two traits.

### Relations Between the Big Five and Trait Affectivity

Earlier, we discussed the fact that most of the Big Five are strongly correlated with trait affectivity. It is important to emphasize that these associations were replicated in the current sample at each assessment. Table 2 reports correlations between the BFI and PANAS-X scales based on the 299 participants who completed all three assessments. The displayed correlations represent the median of the three correlations at Time 1, Time 2, and Time 3. Consistent with previous research, Neuroticism was broadly related to negative emotionality, with median correlations ranging from .41 (Hostility) to .59 (General NA). As expected, Extraversion was positively correlated with various types of positive emotionality, including Joviality ($r$=.57) and Self-Assurance ($r$=.44); not surprisingly, Extraversion also was inversely related to Shyness ($r$=−.70). In contrast, Conscientiousness and Agreeableness were more specifically

associated with Attentiveness (r5.54) and (low) Hostility (*r*=−.57), respectively. Finally, Openness was weakly related to affectivity.

## Stability Analyses

**Rank-order stability—**Table 3 displays stability coefficients and coefficient alphas for the BFI and PANAS-X scales for both retest intervals for the full sample. On the whole, these data point to increasing levels of stability for the BFI and PANAS-X scales. For instance, the median stability coefficients for the BFI scales increased from .62 to .70. Similarly, stability coefficients for the PANAS-X NA and PA scales increased from a median of .49 and .48 to a median of .55 and .57, respectively. In order to compare the stability of each trait across the two intervals more systematically, we conducted significance tests using the Williams modification of the Hotelling test for two correlations with one common variable (Kenny, 1987). As displayed in Table 3, Table 1Table 5 of 18 $T_{23}$ coefficients (83.3%) were higher than the corresponding $T_{12}$ correlations. Moreover, eight of these differences (44.4% of the total) were statistically significant.

In addition, many of the trait scales showed impressive levels of stability across the second retest interval. Most notably, our participants produced a stability correlation of .77 on BFI Extraversion across the 3-year interval between Times 2 and 3. These results are consistent with previous research (see Roberts & DelVecchio, 2000) and highlight the increasing stability of personality across the life span. Table 3 also clearly demonstrates that the PANAS-X scales show consistently lower levels of stability compared to the Big Five scales, even for scales that are conceptually and empirically related.

To examine this effect more systematically, we used the Pearson-Filon test, which tests the difference between two correlations consisting of four nonoverlapping variables from the same sample (Kenny, 1987). These analyses revealed that BFI scales were more stable than the associated PANAS-X scales, and, if anything, the difference was even more evident at the second retest. Extraversion was significantly more stable than the General PA scale, at both the first ($Z$=5.38) and second ($Z$=4.94; *ps*<.01) retest. Neuroticism was not significantly more stable than the General NA scale in the initial retest ($Z$=1.60, p>.05) but did have a higher stability coefficient at the second retest ($Z$=3.03, *p*<.01). Similarly, although Agreeableness was not significantly more stable than Hostility in the initial retest ($Z$= 1.45, *p*>.05), it did show a higher level of stability at the second retest ($Z$=2.67, *p*<.01). Finally, Conscientiousness showed higher levels of stability than Attentiveness across both the first ($Z$=3.65; *p*<.01) and second retest intervals ($Z$=2.36; *p*<.05).

As a method of summarizing these findings, we computed correlations between the first three columns in Table 3. This allowed us to determine if the general pattern of stability coefficients is similar across the different retest intervals. Indeed, the results were highly consistent; we obtained correlations of .88 ($T_{12}$ versus $T_{23}$ coefficients), .85 ($T_{12}$ versus $T_{13}$ coefficients), and .92 ($T_{23}$ versus $T_{13}$ stability coefficients). Thus, some scales are consistently more stable than others.

**Rank-order stability and scale content—**Not only are the Big Five scales more stable than the PANAS-X, there is also substantial variability in the levels of stability for the BFI scales. Extraversion, for instance, had an impressive stability coefficient of .72 and .77 for the two retest intervals, whereas Agreeableness had corresponding stability coefficients of only .59 and .69. Further reinforcing this fact, we found systematic differences in stability at the item level. For this analysis, we computed individual $T_{12}$ and $T_{23}$ stability coefficients for the 44 BFI items. After conducting an *r*-to-z transformation, we correlated the two sets of stability coefficients and obtained a correlation of .72. Thus, some items consistently were more stable than others across our two retest intervals.

To determine if stability is a function of the item content of the scales, we utilized data provided by L. Pytlik Zillig (personal communication) based on the analyses reported in Pytlik Zillig and colleagues (2002). For each BFI item, three experts rated the percentage of affective, behavioral, and cognitive content present for each scale. The raters showed substantial agreement on each item; scores were therefore summed into one overall rating of the percentages of affective, behavioral, and cognitive content. We then correlated these ratings for each item with their $T_{12}$ and $T_{23}$ stability coefficients (after conducting an $r$-to-z transformation of these stability correlations). Therefore, the effective $N$ for each correlation coefficient is the number of items on the BFI (i.e., $N=44$). It is also important to note that these content ratings necessarily are all highly correlated with one another, although the specific pattern of associations varies from scale to scale. For example, items with the highest percentages of cognitive content necessarily also tend to have the lowest percentages of affective or behavioral content. Although these analyses are underpowered because of the small $N$, the data from the two retest intervals are clear and highly consistent. The $T_{12}$ stability coefficients had correlations of $r=.01$, $r=-.10$, and $r=.10$ with the percentage of behavioral, cognitive, and affective content, respectively. The associations with the $T_{23}$ stability coefficients were similarly low, with correlations of $r=-.10$, $r=-.01$, $r=.13$, respectively. Thus, these results clearly suggest that the basic nature of the item content (i.e., behavioral vs. cognitive vs. affective) has no systematic relation to temporal stability levels.

### Repeated Measures Analysis of Mean-Level Change

Table 4 provides descriptive and inferential statistics on analyses of mean-level change. An overall F-test for mean-level change across the three assessments, the corresponding effect size for the F-test, significant differences between mean scores at each assessment, and the corresponding effect sizes are provided. Over the course of the entire time period, Openness and Conscientiousness showed the greatest amount of mean-level change. In fact, Conscientiousness changed by almost 3/4 of 1 standard deviation from Time 1 to Time 3. In contrast, Agreeableness only changed by 1/4 of 1 standard deviation from Time 1 to Time 3.

These mean-level analyses also indicate that while some traits continue to show significant change later in young adulthood, on the whole, most traits showed less change across the second retest period. Thus, while Openness, Agreeableness, and Conscientiousness all significantly increased from Time 2 to Time 3, the degree of change was less pronounced compared to the first retest period. Interestingly, Neuroticism did not show a significant decrease across the first retest—a finding somewhat inconsistent with both prior research and the findings for the PANAS-X NA scales (see below). However, Neuroticism did show a significant mean-level decrease across the second retest interval, suggesting that differences in sample characteristics may be related to the timing and trajectory of change. Extraversion scores seemed to have peaked, on average, at Time 2 and remained largely steady at Time 3. The General NA scale as well as the more specific NA scales all decreased across both retest periods, although the $T_{23}$ effects tended to be weaker. In contrast, while the PA scales showed significant increases, these effects were relatively weaker than those for the NA scales.

### Growth Curve Analyses

We conducted growth curve analyses by regressing each personality trait on age, which was centered on the grand mean (i.e., average age at Time 2). Polynomial functions were not fit because three time points are not considered sufficient to detect nonlinear change using growth curve analyses. An unstructured covariance matrix was specified for the analyses. Random effects were included in the model, allowing individuals to vary in both intercept and rate of change. The results of these analyses are presented in Table 5, Table 6, Table 7, and 8 for the BFI and PANAS-X NA, PA, and other affect scales, respectively. For each scale, fixed and random effects estimates are provided. The fixed effects shown in the top half of the tables

represent the intercept and slope for the entire sample. Thus, the fixed effects slopes should largely correspond with the mean-level findings discussed earlier. The random effects estimates shown in the bottom of the include three results: (a) variability in intercept, (b) variability in slope, and (c) covariance between intercept and slope. The latter two results are of primary interest. Significant slope variability means that there are significant individual differences in growth trajectories. Significant covariance between slope and intercept means that initial standing on a trait predicts how much change occurred on that trait.

**Fixed effects—**The fixed effects slopes are all consistent with the previously reported mean-level findings. Extraversion, Openness, Agreeableness, and Conscientious all had positive slopes indicating that sample as a whole was increasing on these traits. The fixed effect slopes for Neuroticism and the PANAS-X NA scales were negative, indicating that the sample as a whole was decreasing on these scales. Finally, the fixed effects estimates for the PANAS-X PA scales indicated that the sample was increasing on these scales over time.

**Random effects—**Turning to the random effects estimates, the variance estimates for the slopes were significant for all of the BFI scales indicating that there was significant variability in the rate of change for each scale. Thus, while the sample as a whole increased on Extraversion, Openness, Agreeableness, and Conscientiousness, and decreased on Neuroticism, there was significant variability across individuals in the rate of change. In order to characterize the nature of this change, we plotted the mean trajectories for individuals with slopes 1 standard deviation above and below the mean slope. As evident in Figure 1 through Figure 5, individuals differed in the magnitude and, in many cases, the direction of change. Participants showed both increases and decreases on Neuroticism and Extraversion. For Openness and Agreeableness, participants tended to increase or show modest declines over time. Most participants tended to increase on Conscientiousness but differed in the rate of that change.

There was significant variability in the slopes for all but one of the PANAS-X NA scales, while Joviality was the only PA scale that showed significant variability (see Table 6 and Table 7). Three of the PANAS-X NA scales also showed significant covariance between slope and intercept. Due to space considerations, we only show in graphical form the results for Joviality and Hostility (see Figure 6 and Figure 7). As with the BFI scales, these results suggest that there was significant variability in direction of change. For Joviality, the slopes were of roughly equal magnitude but in opposite directions. Therefore, the weak mean-level results for Joviality should not be interpreted as indicating that there is no systematic change. Because the interaction between slope and intercept was significant for Hostility, we graphed two lines representing intercepts at one standard deviation above and below the group intercept mean with each line representing a slope 1 standard deviation below and above the mean slope, respectively. The results for the Hostility scale, which were similar to those for Sadness, Guilt, and Fear, show that there was an overall tendency to decrease on negative emotionality over time; however, individuals with higher scores at the first assessment showed the most dramatic declines.

## DISCUSSION

We reported the results of an ongoing longitudinal study of personality development during young adulthood. Using a three-wave design, we examined the rank-order, mean-level, and individual-level stability of Big Five and affective traits. Rather than simply asking if traits are stable over time, we sought to answer more complex questions about the temporal stability of traits. The present study yielded several notable findings. First, personality showed increasing rank-order stability over time, as the stability correlations in the second retest tended to be higher than those in the first. Second, the PANAS-X scales show weaker rank-order stability

than the BFI scales across both retest intervals. Third, participants as a whole showed continued mean-level changes as measured by both traditional repeated measures ANOVA as well as by growth curve analyses. Overall, however, the strongest mean-level changes are occurring during the earlier phase of young adulthood. Finally, growth curve analyses revealed that there is significant variability in the developmental trajectories. We discuss each of these findings in greater detail below.

## Differential Stability

The retest intervals in the present study correspond to distinct periods within young adulthood for most of the participants in our sample. Between the second and third assessments, most participants had graduated from college and begun working full-time, and many had gotten involved in long-term romantic relationships. During this time, there were systematic increases in rank-order stability from the first to the second retest on many of the BFI and PANAS-X scales. There was also clear evidence of differential stability both within and across measures. Within the BFI, Extraversion showed the highest level of stability across both retest intervals; similar results were reported in a recent analysis of personality stability over a 40-year time span (Hampson & Goldberg, 2006) and in three different meta-analyses (Conley, 1984; Roberts & DelVecchio, 2000; Schuerger et al., 1989). Stability coefficients for the PANAS NA and PA scales increased from a median of .49 and .48 to a median of .55 and .57, respectively. Similarly, the median stability coefficient for the BFI scales increased from .62 to .70. Although stability coefficients are generally higher in the second retest, they are still lower than stability levels reported in older adult samples over similar time periods (e.g., Costa & McCrae, 1988) and may not peak until well into adulthood (Roberts & DelVecchio, 2000).

Although the $T_{23}$ stability coefficients generally were higher than the $T_{12}$ coefficients for the PANAS-X scales, these scales were significantly less stable than the BFI scales across both retest intervals. This is an intriguing finding in light of the fact that many of the BFI scales are conceptually and empirically related to the PANAS-X. Indeed, Neuroticism and Extraversion represent temperament-based traits that reflect individual differences in the experience of negative and positive affective states, respectively (Tellegen, 1985). Our three-wave design allowed us to determine if the lower PANAS-X stability levels we previously reported for the first retest (Vaidya et al., 2002) reflect the fact that affective traits achieve comparable levels of temporal stability at different points in time. On the contrary, the results clearly indicate that while the PANAS-X stabilities were higher across the second retest, these coefficients remained substantially lower than those for the BFI scales. It is noteworthy, furthermore, that PANAS-X scales that have significant empirical and conceptual associations with the BFI also demonstrated significantly lower levels of stability. Thus, the basic pattern of differential stability is highly consistent over time.

Moreover, a virtually identical pattern of stability coefficients was obtained over a 2-month retest (Vaidya et al., 2002, Study 2; see also Watson, 2004). While some systematic change may be possible even over 2 months (Watson, 2004), these results strongly suggest that the lower retest reliability of the PANAS-X contributes to the basic pattern of differential stability that was evident across both retest intervals in our study.

An important question is why the Big Five traits consistently show greater retest reliability and rank-order stability than the PANAS-X scales. Because affective states may fluctuate dramatically, one possibility is that the BFI scales are more stable because of important content differences. Indeed, the BFI Extraversion items ask respondents to indicate whether they are talkative, reserved (reverse-keyed), and outgoing. Only two of the Extraversion items explicitly tap affective content. We examined this issue by correlating each BFI item's stability coefficient with ratings of its percentage of behavioral, affective, and cognitive content. Although it is important to conduct further analyses of this issue, our data tentatively suggest

that affectively laden items are not necessarily associated with lower levels of stability. Other aspects of our data indicate that affective content alone cannot explain the pattern of differential stability we obtained in this study. For instance, the BFI Agreeableness scale has only a moderate amount of affective content (Pytlik Zillig et al., 2002) but showed stability levels comparable to Neuroticism.

Vaidya and colleagues (2002) suggested that format related differences between the BFI and PANAS-X may help to explain why the BFI Neuroticism scale—which largely contains affective content— shows greater stability than the PANAS-X NA scales. The PANAS-X format encourages participants to construct a mental summary of past affective experiences, whereas the BFI asks individuals to indicate whether a given item is consistent with their self-image. Thus, these different assessment approaches may activate different modes of information processing (e.g., Robinson & Clore, 2002). Specifically, the PANAS-X format— which asks respondents to access past feeling states—might encourage more autobiographical or episodic processing, while the BFI format—which involves rating how characteristic a given statement is of oneself—might encourage semantic processing (see Watson, 2004). Because autobiographical information is thought to decay more rapidly from memory (Robinson & Clore, 2002) and because autobiographical ratings of past moods are influenced by current mood states (Schwarz & Clore, 1983), the PANAS-X format may promote a type of processing that may lead to greater error in measurement at any one point in time.

Watson (2004) reported stability coefficients for a new "hybrid" trait affectivity questionnaire. The Temperament and Emotion Questionnaire (TEQ) was created by taking all of the PANAS-X terms and placing them into a sentence context to more closely resemble traditional personality measures such as the BFI. For instance, the PANAS-X terms "cheerful" and "sad" were translated into the TEQ in the following way: "I am a cheerful person" and "I often feel a bit sad". In addition, the response format was changed to a 5-point agree/disagree scale; thus, similar to the BFI, the TEQ asks respondents to indicate whether the items are consistent with their self-view.

Compared with their PANAS-X counterparts, all of the TEQ NA scales showed higher levels of stability, including statistically significant differences for Fear, Sadness, and Hostility. However, there was virtually no difference in stability between the TEQ and PANAS-X PA scales (Watson, 2004). These results suggest that there may be factors influencing trait stability that have yet to be identified. Alternatively, format-related differences may influence the stability of negative, but not positive, affect scales. It will be important for future research to test these different possibilities. The results of these studies will have clear implications for constructing tests with good test-retest reliability as well as for interpreting and comparing stability coefficients derived from scales with differing formats and content.

### Mean- and Individual-Level Change

Our mean-level ANOVA results were highly consistent with the growth curve analyses and largely replicated the meta-analytic findings recently reported by Roberts and colleagues (2006). During young adulthood, participants tend to decrease on Neuroticism and increase on Openness, Agreeableness, and Conscientiousness on average. We also found moderate increases in Extraversion. Although others have not obtained evidence for increases on this trait (e.g., Robins et al., 2001), differences in scale content may explain some of these discrepancies (Roberts et al., 2006). Inclusion of other measures of personality beyond the BFI would have allowed us to detect differences in change related to scale content. Finally, NA scales decreased over time while PA scales showed moderate increases.

While these results characterize overall patterns of change, there is significant variability across individuals in change trajectories. Some individuals are increasing while others are decreasing

on many personality traits (see also, e.g., Griffin et al., in press; Mroczek & Spiro, 2003; Small et al., 2003; Terracciano et al., 2005). Interestingly, while all the BFI scales in our sample showed significant variability in rate of change, Conscientiousness was associated with a more restricted range of trajectories; participants either did not change or showed increased scores on these scales.

The PANAS-X NA scales appear to mirror the BFI scales in that they also demonstrated greater mean-level change across the first retest period. The PA scales, on the other hand, showed modest but consistent increases across both retests. There may be several possible reasons for this slightly different trend. One possibility is that negative affect scores become elevated during adolescence (Larson & Asmussen, 1991) while PA scores remain relatively steady; thus, NA scores may have more room for downward change after the sharp peak in negative affectivity during adolescence. Another possibility is that young adults may show more systematic changes in NA during this time, while patterns of change on PA are less consistent across individuals. In fact, most of the NA scales showed significant variability in their slopes. Joviality was the only PA scale that showed significant variability in trajectories across individuals. Thus, the weaker mean-level findings for Joviality, but not the other PA scales, may be a function of the fact that study participants were both increasing and decreasing on this trait over time.

It is not clear why Joviality, but not the other PA scales, showed individual variability in trajectories. Although it may be tempting to conclude that Joviality is less "trait-like", our rank-order analyses indicate that the stability coefficients were generally as high or higher for Joviality compared to the other PA scales. Although the PA scales are highly correlated with one another (the average correlation in our sample across all three assessments was .51 between Joviality and Attentiveness and .55 between Joviality and Self-Assurance), Joviality shows somewhat stronger negative correlations with Sadness than do the other PA scales (see Watson & Clark, 1999). Thus, the significant variability in change trajectories for Joviality may represent construct overlap with negative affect measures. These results further highlight the fact that while a number of the personality changes evident during young adulthood appear to be in a positive direction (Roberts et al., 2001), this clearly is not true of everyone.

Although we now have a good understanding of how personality changes during young adulthood, relatively little is known about what causes these changes (Pervin, 2002). These changes have been interpreted as reflecting increased maturity (e.g., Roberts, Robins, Caspi, & Trzesniewski, 2003), but more research is needed on the processes that account for these developmental trends. Regardless of how one interprets these developmental trends, the results of the present study suggest that some of the biggest personality changes occur between the ages 18 to 21 (i.e., during the college years). Thus, while there is some evidence that certain types of work experiences can increase scores on Conscientiousness-related traits (e.g., Elder, 1969; Roberts et al., 2003), a complete explanation of personality development must strive to explain why so much personality change is occurring before young adults even reach the work force (see also Roberts et al., 2006). Other factors such as person-environment fit (Caspi, 1998; Roberts & Robins, 2004) and experiences in close relationships also may play an important role. In this regard, it is noteworthy that changes in romantic partner status were associated with modest changes in personality traits in two studies using representative samples of young adults (Neyer & Asendorpf, 2001; Robins, Caspi, & Moffitt, 2002).

## Limitations and Future Directions

Using a three-wave design, the goal of the present study was to provide a better understanding of the differential stability of traits during young adulthood and to use growth curve modeling to characterize overall sample trajectories as well individual differences in these trajectories

over time. While there are a number of important strengths of the study, there are also some limitations. One limitation is the relatively homogenous nature of our sample. A large majority of the sample was female (although there was no evidence for greater attrition among males) and all of our subjects were college students at some point. Without including college and non-college-going participants, it is impossible to determine if the greater stability levels in this study from the first to second retest are due to our participants taking on more adult-like roles or due simply to aging itself. In this regard, however, it is noteworthy that our stability analyses are largely consistent with other studies that have used more representative samples (Neyer & Asendorpf, 2001; Roberts et al., 2001). The mean-level and rank-order stability analyses on the BFI are also largely consistent with the meta-analyses reported by Roberts and colleagues (Roberts & DelVecchio, 2000; Roberts et al., 2006).

A second limitation of this study is that ratings were based entirely on self-report. Results based solely on self-report may be evidence for the stability of self-concepts, not necessarily trait stability (see Costa & McCrae, 1988). Unfortunately, few studies have assessed other-reports using established Big Five or Big Three personality measures (for an extended discussion of this issue, see Watson & Humrichouse, 2006). Also, as with many longitudinal studies, we had some attrition in our sample from the initial testing, and this attrition was not random. Having more individuals with three data points would have provided better estimates of the mean and variability in growth curves. Finally, because personality data were not available on our participants prior to the age of 18, we were unable to determine whether some traits showed greater changes during young adulthood in response to earlier shifts on these traits during adolescence. For instance, NA may show more dramatic changes than PA in young adulthood because of a sharp peak in negative emotional states during adolescence (Larson & Asmussen, 1991).

A more powerful design would involve initially collecting trait data on a large group of participants in high school, using both self- and other-reports. Collecting data on participants in high school would provide a baseline against which to compare later personality changes. Then, participants could be tested at regular intervals throughout young adulthood while tracking the trajectory of change for individuals who experience distinct life paths. Indeed, in order to have a comprehensive understanding of personality change during young adulthood, researchers must strive to collect longitudinal data on young adults who do not go to college. Such a design would provide a further test of the maturational hypothesis which suggests that mean-level personality changes are biologically, as opposed to environmentally, determined (McCrae et al., 1999). In addition, by collecting data from informants, researchers would be able to establish whether patterns of stability and change are consistent across self- and other ratings or if they are method specific. Results that are inconsistent across the two modalities may be function of self- or other-perception biases.

In recent years, researchers have made significant strides in characterizing the nature of personality change during young adulthood. Research in this area will benefit from investigators asking increasingly more complex questions regarding the differential stability of traits over time and across measures. This study represents a significant step in answering some of these important questions.

## REFERENCES

Arnett JJ. Emerging adulthood: A theory of development from the late teens through the twenties. American Psychologist 2000;55:469–480. [PubMed: 10842426]

Caspi, A. Personality development across the life course. In: Eisenberg, N.; Damon, W., editors. Social, emotional and personality development, Handbook of child psychology. New York: John Wiley & Sons, Inc; 1998. p. 311-388.(Volume Ed.) (Series Ed.)

Caspi, A.; Roberts, BW. Personality change and continuity across the life course. In: Pervin, LA.; John, OP., editors. Handbook of personality theory and research. New York: Guilford Press; 1999. p. 300-326.

Conley JJ. The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. Personality and Individual Differences 1984;5:11–25.

Costa PT Jr, McCrae RR. Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. Journal of Personality and Social Psychology 1988;54:853–863. [PubMed: 3379583]

Costa, PT., Jr.; McCrae, RR. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment; 1992.

Elder GH. Occupational mobility, life patterns, and personality. Journal of Health and Social Behavior 1969;10:308–323.

Fraley RC, Roberts BW. A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. Psychological Review 2005;112:60–74. [PubMed: 15631588]

Griffin PW. Positive and negative affect in middle to late adulthood: Longitudinally examining and predicting individual differences and intraindividual change. (Doctoral dissertation, Fordham University). Dissertation Abstracts International 2004;65:1578.

Griffin PW, Mroczek DK, Spiro A III. Variability in affective change among aging men: Longitudinal findings from the VA Normative Aging Study. Journal of Research in Personality. in press

Hampson SE, Goldberg LR. A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. Journal of Personality and Social Psychology 2006;91:763–779. [PubMed: 17014298]

Izard, CE. The psychology of emotions. New York: Plenum; 1991.

Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59:12–19. [PubMed: 2002127]

John, OP.; Srivastava, S. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, LA.; John, OP., editors. Handbook of personality: Theory and research. Vol. 2nd ed. New York: Guilford; 1999. p. 102-138.

Kenny, DA. Statistics for the social and behavioral sciences. Boston: Little, Brown and Company; 1987.

Larson, R.; Asmussen, L. Anger, worry, and hurt in early adolescence: An enlarging world of negative emotions. In: Colten, ME.; Gore, S., editors. Adolescent stress: Causes and consequences. New York: Aldine de Gruyter; 1991. p. 21-41.

McCrae RR, Costa PT Jr, Lima MP, Simões A, Ostendorf F, Angleitner A, et al. Age differences in personality across the adult lifespan: Parallels in five cultures. Developmental Psychology 1999;35:466–477. [PubMed: 10082017]

McGue M, Bacon S, Lykken DT. Personality stability and change in early adulthood: A behavioral genetic analysis. Developmental Psychology 1993;29:96–109.

Mroczek DK, Spiro A III. Modeling intraindividual change in personality traits: Findings from the Normative Aging Study. Journals of Gerontology: Series B: Psychological Sciences and Social Sciences 2003;58B:P153–P165.

Newman DL, Moffitt TE, Caspi A, Magdol L, Silva PA, Stanton WR. Psychiatric disorder in a birth cohort of young adults: Prevalence, comorbidity, clinical significance, and new case incidence from ages 11 to 21. Journal of Consulting and Clinical Psychology 1996;64:552–562. [PubMed: 8698949]

Neyer FJ, Asendorpf JB. Personality-relationship transactions in young adulthood. Journal of Personality and Social Psychology 2001;81:1190–1204. [PubMed: 11761317]

Pervin, LA. Current controversies and issues in personality. Vol. 3rd ed. New York: Wiley; 2002.

Pytlik Zillig LM, Hemenover SH, Dienstbier RA. What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. Personality and Social Psychology Bulletin 2002;28:847–858.

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. Vol. 2nd ed.. Thousand Oaks, CA: Sage; 2002.

Rindfuss RR. The young adult years: Diversity, structural change, and fertility. Demography 1991;28:493–512. [PubMed: 1769399]

Roberts BW, Caspi A, Moffitt TE. The kids are alright: Growth and stability in personality development from adolescence to adulthood. Journal of Personality and Social Psychology 2001;81:670–683. [PubMed: 11642353]

Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. Psychological Bulletin 2000;126:3–25. [PubMed: 10668348]

Roberts BW, Robins RW. A longitudinal study of person-environment fit and personality development. Journal of Personality 2004;72:89–110. [PubMed: 14686885]

Roberts, BW.; Robins, RW.; Caspi, A.; Trzesniewski, K. Mortimer, J.; Shanahan, M. Handbook of the life course. New York: Kluwer Academic; 2003. Personality trait development in adulthood; p. 579-598.

Roberts BW, Walton KE, Viechtbauer W. Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psychological Bulletin 2006;132:1–25. [PubMed: 16435954]

Robins RW, Caspi A, Moffitt TE. It's not just who you're with, it's who you are: Personality and relationship experiences across multiple relationships. Journal of Personality 2002;70:925–964. [PubMed: 12498360]

Robins RW, Fraley RC, Roberts BW, Trzesniewski KH. A longitudinal study of personality change in young adulthood. Journal of Personality 2001;69:617–640. [PubMed: 11497032]

Robinson MD, Clore GL. Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychological Bulletin 2002;128:934–960. [PubMed: 12405138]

Rogosa DR, Brandt D, Zimowski M. A growth curve approach to the measurement of change. Psychological Bulletin 1982;92:726–748.

Rogosa DR, Willett JB. Understanding correlates of change by modeling individual differences in growth. Psychometrika 1985;50:203–228.

Schuerger JM, Zarrella KL, Hotz AS. Factors that influence the temporal stability of personality by questionnaire. Journal of Personality and Social Psychology 1989;56:777–783.

Schwarz N, Clore GL. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. Journal of Personality and Social Psychology 1983;45:513–523.

Singer, JD.; Willett, JD. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.

Small BJ, Hertzog C, Hultsch DF, Dixon RA. Stability and change in adult personality over 6 years: Findings from the Victoria Longitudinal Study. Journals of Gerontology: Psychological Sciences and Social Sciences 2003;58B:166–176.

Tellegen, A. Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Tuma, AH.; Maser, JD., editors. Anxiety and the anxiety disorders. Hillsdale, NJ: Erlbaum; 1985. p. 681-706.

Terracciano A, McCrae RR, Brant LJ, Costa PT Jr. Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. Psychology and Aging 2005;20:493–506. [PubMed: 16248708]

Vaidya JG, Gray EK, Haig J, Watson D. On the temporal stability of personality: Evidence for differential stability, and the role of life experiences. Journal of Personality and Social Psychology 2002;83:1469–1484. [PubMed: 12500825]

Vollrath M. Personality and hassles among university students: A three-year longitudinal study. European Journal of Personality 2000;14:199–215.

Watson, D. Mood and temperament. New York: Guilford Press; 2004.

Watson D. Stability versus change, dependability versus error: Issues in the assessment of personality over time. Journal of Research in Personality 2004;38:319–350.

Watson D, Clark LA. On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. Journal of Personality 1992;60:441–476. [PubMed: 1635050]

Watson D, Clark LA. Measurement and mismeasurement of mood: Recurrent and emergent issues. Journal of Personality Assessment 1997;68:267–296. [PubMed: 16370781]

Watson D, Clark LA. The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form. Retrieved from University of Iowa, Department of Psychology. 1999Web site: http://www.psychology.uiowa.edu/ Faculty/Watson/Watson.html

Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology 1988;54:1063–1070. [PubMed: 3397865]

Watson D, Hubbard B. Adaptational style and dispositional structure: Coping in the context of the five-factor model. Journal of Personality 1996;64:737–774.

Watson D, Hubbard B, Wiese D. Self-other agreement in personality and affectivity: Effects of acquaintanceship, trait visibility, and assumed similarity. Journal of Personality and Social Psychology 2000;78:546–558. [PubMed: 10743880]

Watson D, Humrichouse J. Personality development in emerging adulthood: Integrating evidence from self- and spouse-ratings. Journal of Personality and Social Psychology 2006;91:959–974. [PubMed: 17059313]

Watson D, Walker LM. The long-term temporal stability and predictive validity of trait measures of affect. Journal of Personality and Social Psychology 1996;70:567–577. [PubMed: 8851741]

Watson D, Wiese D, Vaidya J, Tellegen A. The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. Journal of Personality and Social Psychology 1999;76:820–838.

Willett, JB. Questions and answers in the measurement of change. In: Rothkopf, EZ., editor. Review of research in education. Vol. Vol. 15. Washington, DC: American Educational Research Association; 1988. p. 345-422.

Willett JB, Sayer AG. Using covariance structural analysis to detect correlates and predictors of individual change over time. Psychological Bulletin 1994;116:363–381.
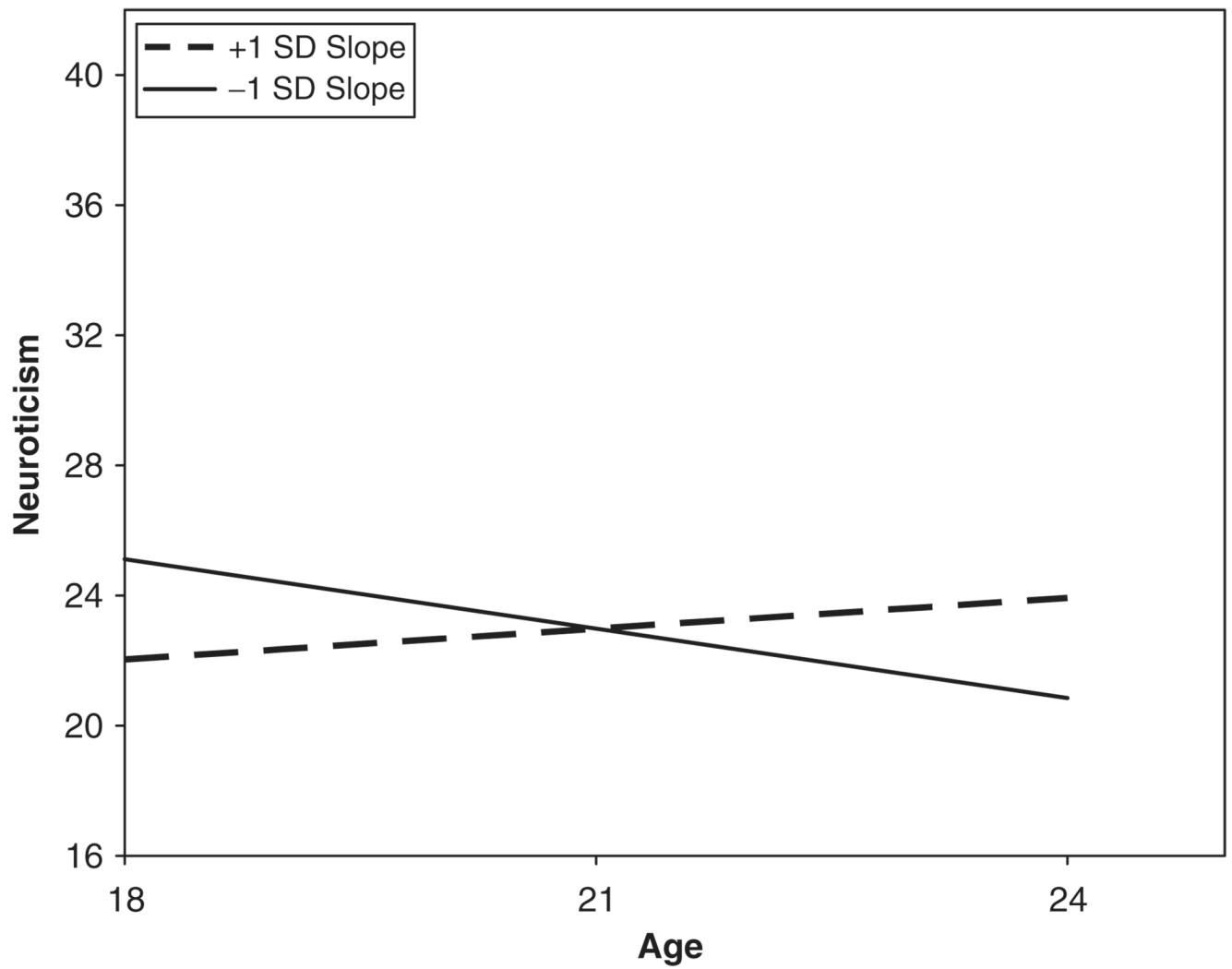
**Figure 1.**
Growth curve analyses for Neuroticism.
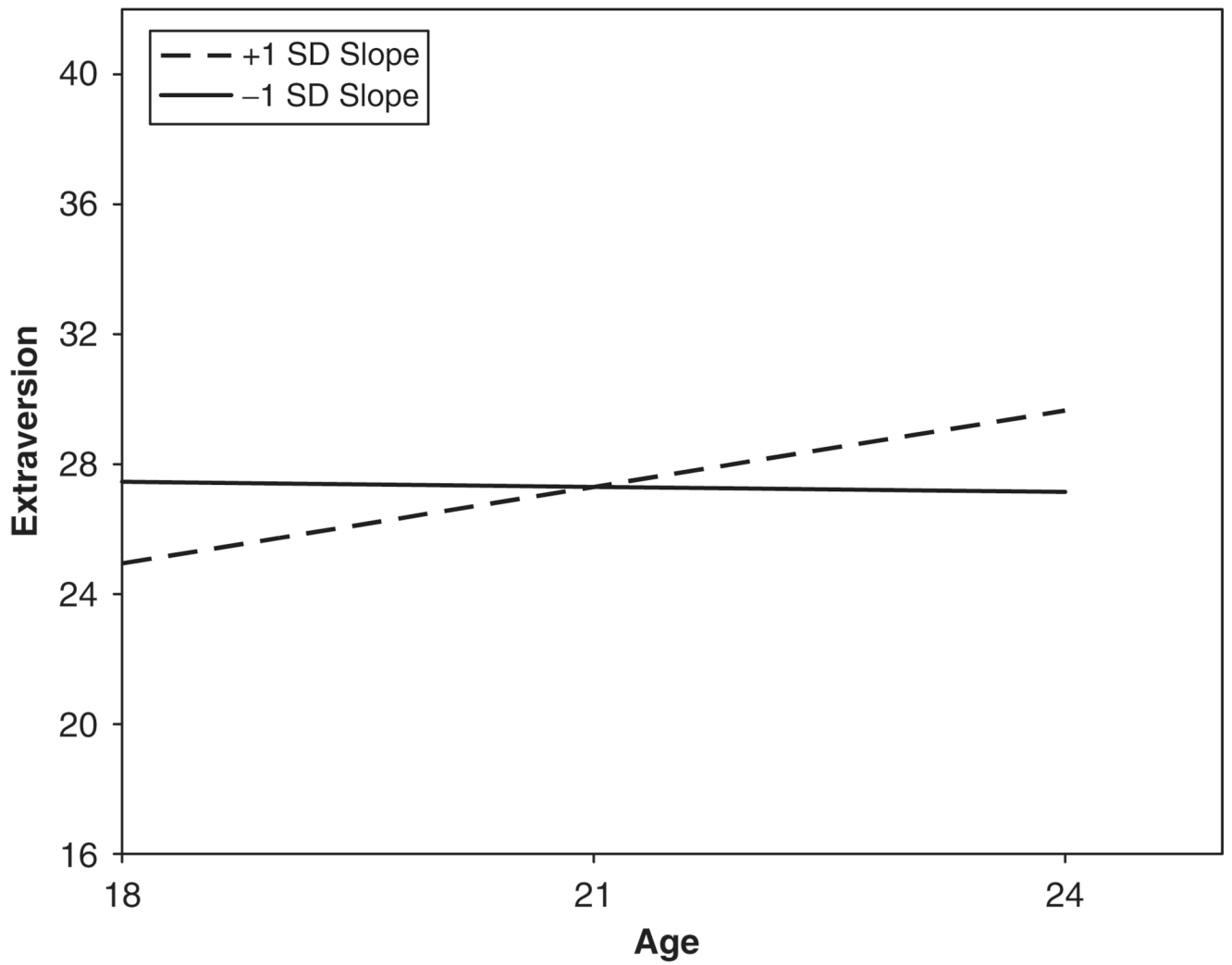
**Figure 2.**
Growth curve analyses for Extraversion.
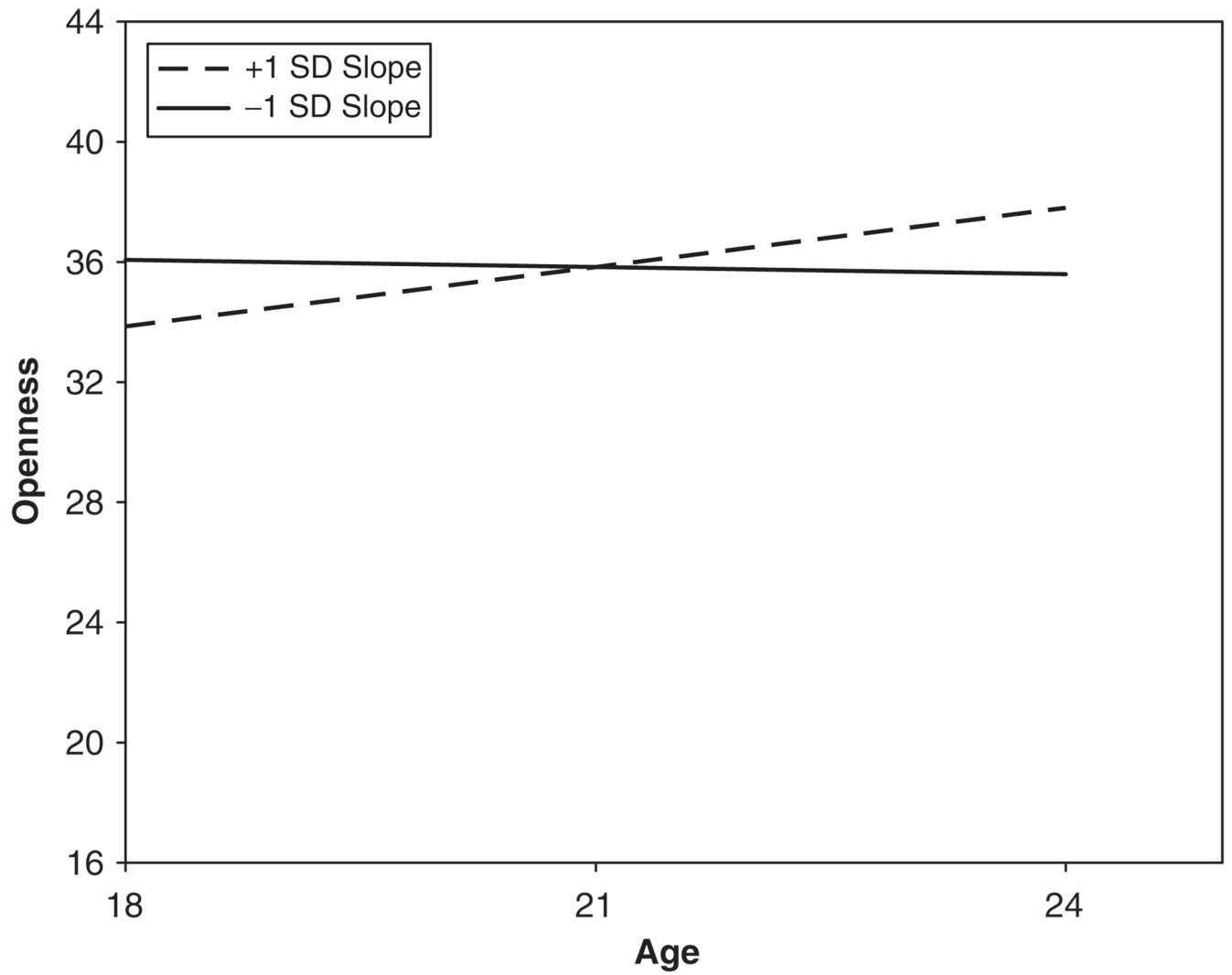
**Figure 3.**
Growth curve analyses for Openness.
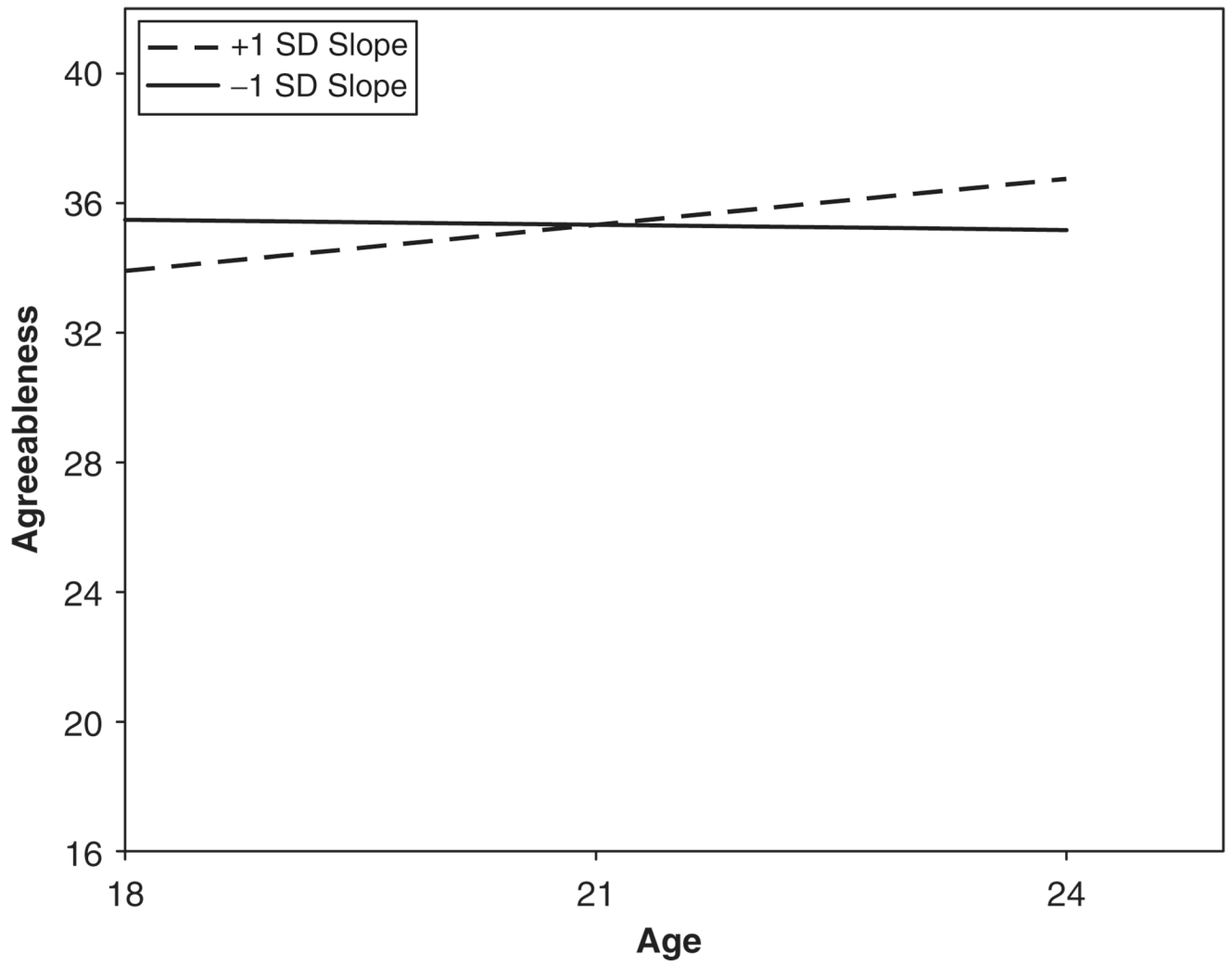
**Figure 4.**
Growth curve analyses for Agreeableness.
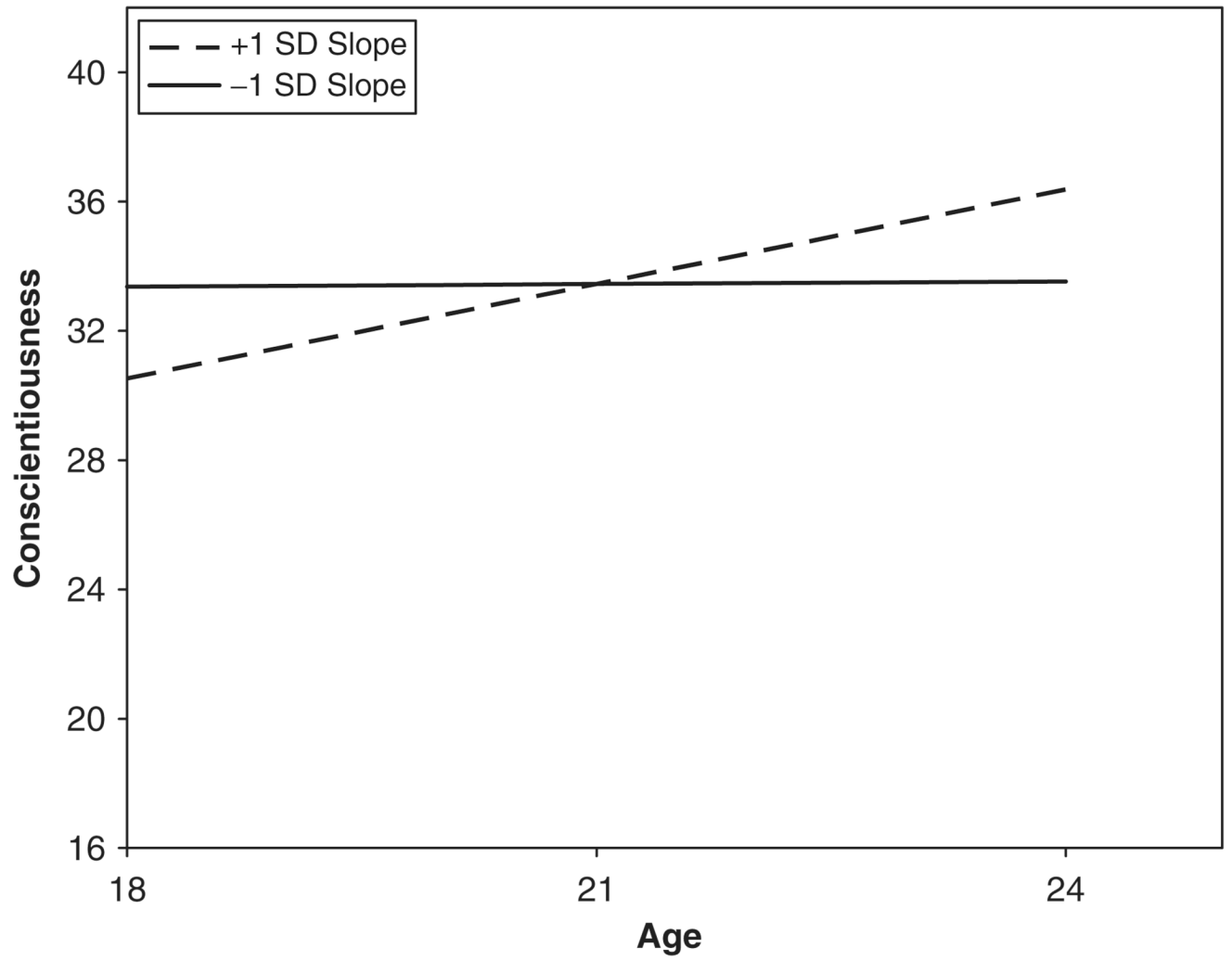
**Figure 5.**
Growth curve analyses for Conscientiousness.
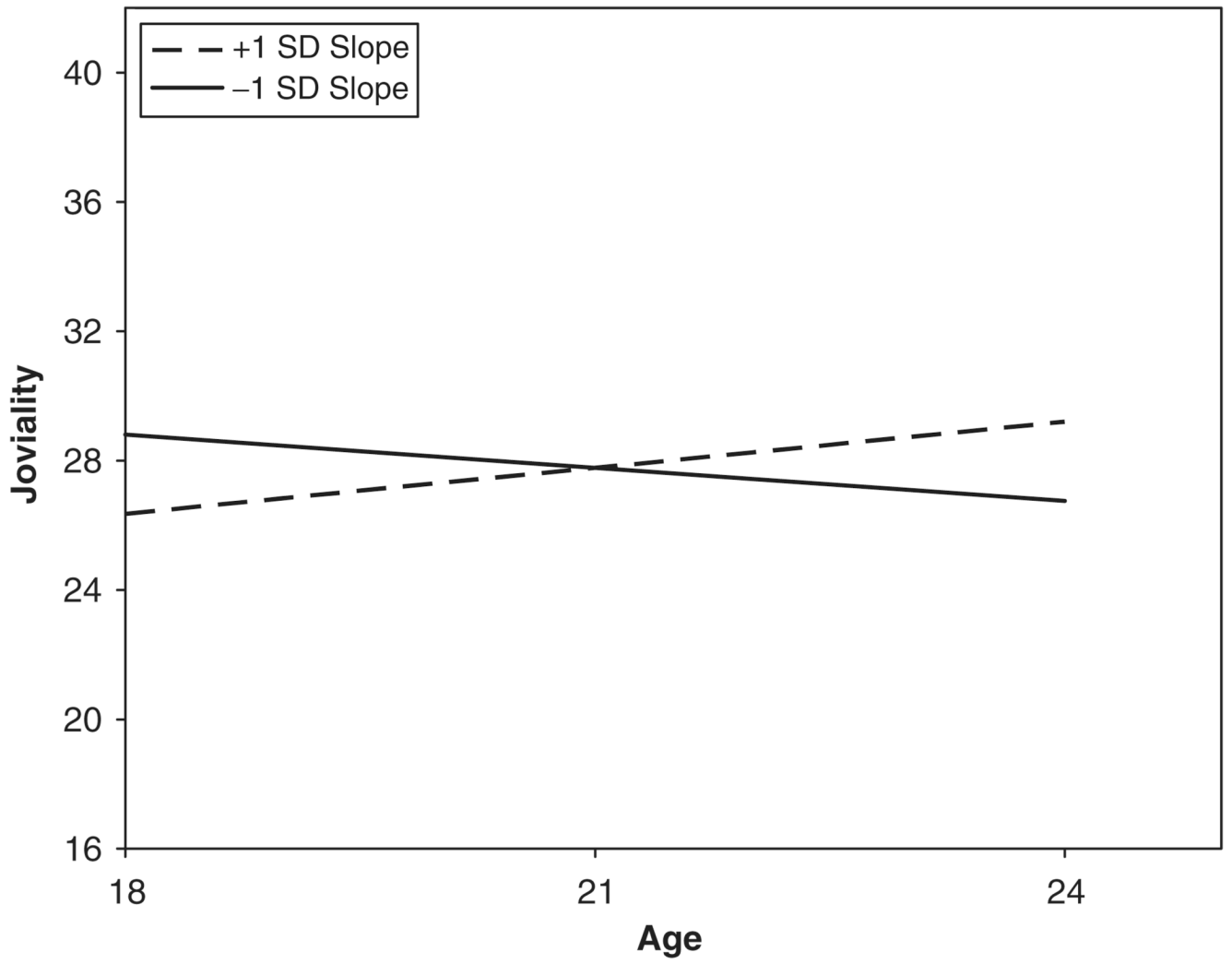
**Figure 6.**
Growth curve analyses for Joviality.
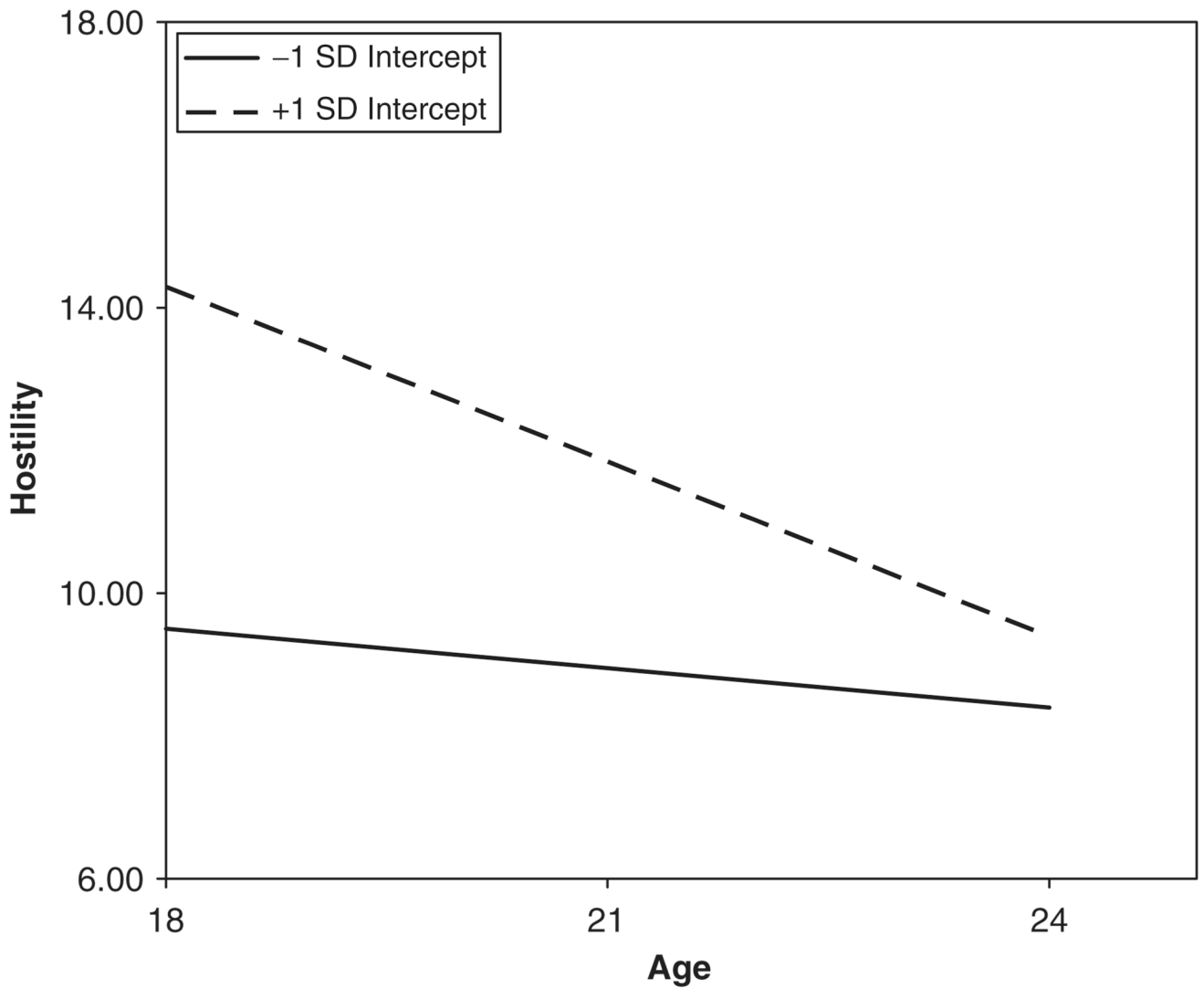
**Figure 7.**
Growth curve analyses for Hostility. Lines represent slopes 1 standard deviation above and below the sample mean. Because there was significant covariance in slope and intercept, one line represents 1 standard deviation above the sample slope and intercept and the other line represents 1 standard deviation below the sample slope and intercept.

**Table 1**

Demographic Information at Time 2 and Time 3 for Complete Sample

|  | Time 2 | Time 3 |
| --- | --- | --- |
| Age (Mean) | 21.1 | 24.0 |
| *Relationship status* | | |
| Single | 29.8% | 26.4% |
| Dating | 64.9% | 51.8% |
| Engaged or Married | 5.0% | 21.4% |
| *Work/student status* | | |
| Undergraduate student | 81.6% | 6.4% |
| Graduate Student | 0.01% | 20.1% |
| Working full-time | 6.7% | 61.2% |

*Note: N*=299.

**Table 2**

Correlations Between the BFI and PANAS-X Scales

| | BFI Scale | | | | |
|---|---|---|---|---|---|
| **Scale** | **N** | **E** | **O** | **A** | **C** |
| *NA Scales* | | | | | |
| General NA | **.59** | -.25 | -.05 | -.38 | -.28 |
| Fear | **.55** | -.24 | -.04 | -.18 | -.17 |
| Sadness | **.46** | -.31 | .01 | -.29 | -.25 |
| Guilt | **.42** | -.19 | .01 | -.32 | -.31 |
| Hostility | **.41** | -.25 | -.10 | -.57 | -.30 |
| *PA Scales* | | | | | |
| General PA | -.34 | **.46** | .20 | .31 | .40 |
| Joviality | -.39 | **.57** | .14 | .40 | .17 |
| Self-assurance | -**.41** | **.44** | .19 | .03 | .10 |
| Attentiveness | -.19 | .29 | .10 | .24 | **.54** |
| *Other Affect Scales* | | | | | |
| Surprise | -**.08** | .15 | .04 | .15 | .07 |
| Serenity | -**.65** | .15 | .09 | .26 | .05 |
| Fatigue | .32 | -.24 | -.07 | -.25 | -.29 |
| Shyness | .32 | -**.70** | -.15 | -.06 | -.17 |

*Note: N=299.*

Correlations represent median inter-scale correlation from Time 1, Time 2, and Time 3.

Correlations greater than |.40| are in bold.

**Table 3**

Rank-Order Stability of BFI and PANAS-X Scales

| Scale | Stability Coefficients | | | Coefficient Alpha | | |
|---|---|---|---|---|---|---|
| | $T_{12}$ | $T_{23}$ | $T_{13}$ | T1 | T2 | T3 |
| *BFI Scales* | | | | | | |
| Neuroticism | .58[a] | .70[b] | .49 | .83 | .88 | .81 |
| Extraversion | .72[a] | .77[a] | .63 | .85 | .88 | .86 |
| Openness | .64[a] | .73[b] | .57 | .78 | .83 | .80 |
| Agreeableness | .59[a] | .69[b] | .52 | .77 | .78 | .79 |
| Conscientiousness | .62[a] | .62[a] | .49 | .72 | .75 | .72 |
| (Median) | (.62) | (.70) | (.52) | (.78) | (.83) | (.80) |
| *NA Scales* | | | | | | |
| General NA | .49[a] | .55[a] | .48 | .82 | .81 | .84 |
| Hostility | .51[a] | .56[a] | .46 | .78 | .81 | .79 |
| Guilt | .46[a] | .56[b] | .42 | .81 | .82 | .85 |
| Sadness | .52[a] | .50[a] | .36 | .86 | .87 | .82 |
| Fear | .43[a] | .50[a] | .43 | .83 | .78 | .83 |
| (Median) | (.49) | (.55) | (.43) | (.82) | (.81) | (.83) |
| *PA Scales* | | | | | | |
| General PA | .44[a] | .55[b] | .44 | .77 | .82 | .84 |
| Joviality | .51[a] | .61[b] | .50 | .91 | .92 | .91 |
| Self-Assurance | .52[a] | .59[a] | .50 | .72 | .77 | .79 |
| Attentiveness | .41[a] | .49[a] | .37 | .69 | .74 | .72 |
| (Median) | (.48) | (.57) | (.47) | (.75) | (.80) | (.82) |
| *Other Affect Scales* | | | | | | |
| Shyness | .60[a] | .68[b] | .54 | .86 | .86 | .86 |
| Fatigue | .44[a] | .45[a] | .42 | .85 | .86 | .86 |
| Serenity | .45[a] | .58[b] | .45 | .80 | .81 | .78 |
| Surprise | .42[a] | .37[a] | .34 | .72 | .78 | .73 |
| (Median) | (.45) | (.52) | (.44) | (.83) | (.84) | (.82) |

*Note: N*=299.

$T_{12}$ and $T_{23}$ stability coefficients with different subscripts indicate significant differences between the two stability coefficients at *p*<.05.

All correlations are significant, *p*<.01

**Table 4**

Mean-Level Stability of the BFI and PANAS-X Scales

| Scale | Means | | | F | $\eta^2$ | Cohen's d | | |
|---|---|---|---|---|---|---|---|---|
| | Time 1 | Time 2 | Time 3 | | | $T_{12}$ | $T_{23}$ | $T_{13}$ |
| *BFI Scales* | | | | | | | | |
| Neuroticism | 23.51$_a$ | 22.98$_a$ | 22.18$_b$ | 8.99 ** | 0.03 | -0.09 | -0.13 | -0.22 |
| Extraversion | 26.09$_a$ | 28.24$_b$ | 28.01$_b$ | 35.13 ** | 0.11 | -0.33 | -0.04 | 0.30 |
| Openness | 33.93$_a$ | 36.23$_b$ | 36.90$_c$ | 52.63 ** | 0.15 | 0.36 | 0.11 | 0.48 |
| Agreeableness | 34.62$_a$ | 35.19$_b$ | 36.06$_c$ | 14.30 ** | 0.05 | 0.11 | 0.16 | 0.26 |
| Conscientiousness | 31.68$_a$ | 34.30$_b$ | 35.37$_c$ | 101.46 ** | 0.25 | 0.51 | 0.22 | 0.74 |
| *NA Scales* | | | | | | | | |
| General NA | 21.62$_a$ | 18.97$_b$ | 17.45$_c$ | 84.97 ** | 0.22 | -0.46 | -0.29 | -0.78 |
| Hostility | 12.40$_a$ | 10.62$_b$ | 9.89$_c$ | 78.79 ** | 0.21 | -0.48 | -0.22 | -0.76 |
| Guilt | 11.95$_a$ | 10.61$_b$ | 9.40$_c$ | 49.61 ** | 0.14 | -0.30 | -0.30 | -0.64 |
| Sadness | 11.43$_a$ | 9.95$_b$ | 8.91$_c$ | 55.80 ** | 0.16 | -0.35 | -0.27 | -0.65 |
| Fear | 12.57$_a$ | 10.86$_b$ | 10.10$_c$ | 65.03 ** | 0.18 | -0.46 | -0.22 | -0.71 |
| *PA Scales* | | | | | | | | |
| General PA | 34.19$_a$ | 35.37$_b$ | 36.16$_c$ | 18.98 ** | 0.06 | 0.21 | 0.14 | 0.36 |
| Joviality | 27.27$_a$ | 28.23$_b$ | 28.41$_b$ | 8.30 ** | 0.03 | 0.17 | 0.03 | 0.21 |
| Self-Assurance | 18.11$_a$ | 18.38$_a$ | 18.83$_b$ | 5.60 ** | 0.02 | 0.07 | 0.11 | 0.18 |
| Attentiveness | 13.93$_a$ | 14.44$_b$ | 15.01$_c$ | 27.80 ** | 0.09 | 0.22 | 0.25 | 0.46 |
| *Other Affect Scales* | | | | | | | | |
| Shyness | 9.84$_a$ | 7.92$_b$ | 7.66$_b$ | 92.00 ** | 0.24 | -0.55 | -0.08 | -0.69 |
| Fatigue | 11.68$_a$ | 10.97$_b$ | 9.61$_c$ | 51.31 ** | 0.15 | -0.21 | -0.41 | -0.62 |
| Serenity | 9.69$_a$ | 9.85$_a$ | 10.28$_b$ | 10.54 ** | 0.03 | 0.07 | 0.19 | 0.26 |
| Surprise | 7.34$_a$ | 6.44$_b$ | 6.42$_b$ | 25.70 ** | 0.08 | -0.39 | -0.01 | -0.40 |

*Note: N=299.*

Means that share a subscript do not significantly differ at *p*<.05.

** *p*<.01.

Growth Curve Estimates for BFI Scales

**Table 5**

| Effect | Neuroticism | Extraversion | Openness | Agreeableness | Conscientiousness |
|---|---|---|---|---|---|
| **Fixed Effects Estimates** | | | | | |
| Intercept | 22.98 (.27) | 27.30 (.29) | 35.83 (.28) | 35.33 (.23) | 33.45 (.23) |
| t-statistic | 84.38 *** | 93.71 *** | 129 43 *** | 115.94 *** | 145.54 *** |
| Slope | −.22 (.06) | .35 (.05) | .65 (.06) | .23 (.05) | .67 (.05) |
| t-statistic | −3 89 *** | 6.62 *** | 11.75 *** | 4.60 *** | 13.56 *** |
| **Random Effects Estimates** | | | | | |
| Variance of Intercept | 24.03 (2.13) | 29.11 (2.40) | 25.16 (2.18) | 16.35 (1.46) | 16.90 (1.51) |
| z-statistic | 11.29 *** | 12.15 *** | 11.56 *** | 11.16 *** | 11.22 *** |
| Variance of Slope | .29 (.11) | .24 (.08) | .19 (.09) | .21 (.08) | .17 (.07) |
| z-statistic | 2.67 ** | 2.85 ** | 1.99 * | 2.61 ** | 2.44 ** |
| Covariance of Intercept, Slope | −.07 (.32) | −.50 (.31) | −.35 (.30) | −.06 (.21) | −.34 (.24) |
| z-statistic | −.24 | −1.59 | −1.16 | −.27 | −1.45 |
| Residual Variance | 12.81 | 10.57 | 12.66 | 9.58 | 9.78 |
| −2LL | 6660 | 6575 | 6634 | 6312 | 6321 |

*Note* $N$=392; number of observations =1087. Standard errors are in parentheses. −2LL =−2 log likelihood, a fit index.

*
$p \leq$ .05.

**
$p \leq$ .01.

***
$p \leq$ .001.

**Table 6**

Growth Curve Estimates for Negative Affect Scales

| Effect | General NA | Fear | Sadness | Guilt | Hostility |
|---|---|---|---|---|---|
| Fixed Effects Estimates | | | | | |
| Intercept | 19.75 (.25) | 11.38 (.16) | 10.43 (.18) | 11.00 (.19) | 11.13 (.16) |
| $t$-statistic | 79.42 *** | 71.54 *** | 59.18 *** | 57.30 *** | 70.85 *** |
| Slope | −.75 | −.44 (.04) | −.45 (.04) | −.44 (.05) | −.43 (.04) |
| $z$-statistic | −12.79 *** | −11.02 *** | −10.32 *** | −9.28 *** | −11.24 *** |
| Random Effects Estimates | | | | | |
| Variance of Intercept | 18.17 (1.79) | 7.20 (.74) | 9.03 (.91) | 10.65 (1.08) | 7.29 (.72) |
| $z$-statistic | 10.14 *** | 9.73 *** | 9.95 *** | 9.89 *** | 10.14 *** |
| Variance of Slope | .11 (.09) | .08 (.04) | .10 (.05) | .10 (.06) | .10 (.04) |
| $z$-statistic | 1.24 | 1.90 * | 2.02 * | 1.70 * | 2.62 ** |
| Covariance of Intercept, Slope | −.45 (.29) | −.18 (.13) | −.36 (.16) | −.43 (.18) | −.38 (.12) |
| $z$-statistic | −1.54 | −1.39 | −2.27 * | −2.31 * | −3.04 ** |
| Residual Variance | 16.13 | 7.16 | 8.28 | 10.01 | 6.16 |
| −2LL | 6683 | 5794 | 5971 | 6161 | 5690 |

*Note: N* = 392; number of observations = 1087. Standard errors are in parentheses. NA = Negative Affect; − 2LL = − 2 log likelihood, a fit index.

*
$p \le .05$.

**
$p \le .01$.

***
$p \le .001$.

**Table 7**

Growth Curve Estimates for Positive Affect Scales

| Effect | General PA | Joviality | Self-Assurance | Attentiveness |
|---|---|---|---|---|
| Fixed Effects Estimates | | | | |
| Intercept | 35.13 (.24) | 27.78 (.24) | 18.45 (.16) | 14.40 (.10) |
| *t*-statistic | 145.35*** | 117.03*** | 112.46*** | 143.01*** |
| Slope | .33 (.06) | .17 (.05) | .11 (.04) | .18 (.03) |
| *t*-statistic | 5.86*** | 3.19** | 2.82** | 7.00*** |
| Random Effects Estimates | | | | |
| Variance of Intercept | 17.31 (1.69) | 17.41 (1.61) | 8.02 (.77) | 2.85 (.29) |
| *z*-statistic | 10.23*** | 10.82*** | 10.44*** | 9.62*** |
| Variance of Slope | .09 (.09) | .17 (.08) | .06 (.05) | .02 (.02) |
| *z*-statistic | 1.04 | 2.12* | 1.29 | 1.42 |
| Covariance of Intercept, Slope | .04 (.28) | −.08 (.26) | .11 (.12) | .001 (.05) |
| *z*-statistic | .13 | −.31 | .91 | .02 |
| Residual Variance | 14.85 | 12.14 | 6.70 | 2.96 |
| −2LL | 6603 | 6482 | 5759 | 4817 |

*Note N*=392; number of observations=1087. Standard errors are in parantheses. PA = Negative Affect; −2LL 5 −2 log likelihood, a fit index.

*
 $p \le .05$.

**
 $p \le .01$.

***
 $p \le .001$.

**Table 8**

Growth Curve Estimates for Other Affect Scales

| Effect | Shyness | Fatigue | Serenity | Surprise |
|---|---|---|---|---|
| Fixed Effects Estimates | | | | |
| Intercept | 8.61 (.15) | 10.89 (.14) | 9.88 (.10) | 6.86 (.09) |
| *t*-statistic | 57.37[***] | 79.24[***] | 100.71[***] | 74.68[***] |
| Slope | −.38 (.03) | −.37 (.04) | .10 (.02) | −.17 (.02) |
| *t*-statistic | −12 31[***] | −10.46[***] | 4.25[***] | −6 97[***] |
| Random Effects Estimates | | | | |
| Variance of Intercept | 7.06 (.65) | 4.98 (.55) | 2.82 (.28) | 2.08 (.25) |
| *z*-statistic | 10.92[***] | 9.09[***] | 10.11[***] | 8.42[***] |
| Variance of Slope | .02 (.03) | .02 (.03) | .02 (.02) | <.01 |
| *z*-statistic | .71 | .56 | 1.08 | <.01 |
| Covariance of Intercept, Slope | −.25 (.09) | −.02 (.10) | .02 (.05) | −.02 (.04) |
| *z*-statistic | −2.71[**] | −.30 | .48 | −.53 |
| Residual Variance | 4.78 | 6.55 | 2.58 | 3.35 |
| −2LL | 5433 | 5579 | 4689 | 4788 |

*Note:* N = 392; number of observations =1087. Standard errors are in parantheses. −2LL = −2 log likelihood, a fit index.

[*]
 *p* ≤ .05.

[**]
 *p* ≤ .01.

[***]
 *p* ≤ .001.