Title

Developmental Associations between Temperament and Psychological Adjustment from Late Childhood to Young Adulthood

Author
Lawson, Katherine Murray

Publication Date
2022

Developmental Associations between Temperament and Psychological Adjustment from Late Childhood to Young Adulthood

By

KATHERINE MURRAY LAWSON
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____
Richard W. Robins, Chair

____
Wiebke Bleidorn

____
Christopher J. Hopwood

Committee in Charge

2022

Acknowledgements

First, I would like to thank my Ph.D. advisor and dedicated mentor, Rick Robins, who has been a tremendous source of knowledge, support, and guidance for me during the past five years. Rick has expertly balanced fostering my independence while teaching me something new every day, and he is an exemplar of how to be a thoughtful and curious researcher. I feel incredibly lucky to have benefitted from Rick's mentorship, and I am very grateful that Rick always made an effort to foster my own love of teaching and mentoring by creating every possible opportunity for me to work with students, even though this often meant taking on much more responsibility himself. I so appreciate the integral role that Rick has played in my graduate school experience, and I feel very fortunate to know that he will continue to be a lifelong mentor.

Two other members of my dissertation committee, Wiebke Bleidorn and Chris Hopwood, have also provided consistent, supportive mentorship throughout my time in graduate school. They always ask the tough questions and encourage me to think carefully and deeply. I'd also like to thank the many other UC Davis faculty members who have served as collaborators, advisors, and informal mentors, including Emilio Ferrer, Mijke Rhemtulla, Alison Ledgerwood, Johnna Swartz, Joanna Scheib, and Simine Vazire. My journey to graduate school began long before I arrived at UC Davis, and I feel extremely lucky to have been mentored by Dan Krauss, Debby Burke, and Julie Tannenbaum during my time at Pomona College. Together, they built the academic foundation that facilitated my success during graduate school. Further, they have all served as wonderful inspiration for my own career path by demonstrating what it means to be an amazing teacher and mentor. I'm also quite appreciative of my former lab member and good friend, Olivia Atherton, from whom I learned so much about research, graduate school, and being a good colleague.

My students and mentees at UC Davis have been a tremendous source of inspiration for me, encouraging me to think about my research and the world around me in new and interesting ways.

My dissertation research was made possible by the amazing California Families Project (CFP) interviewers and staff, as well as the CFP families who have given me a snapshot into so much of their lives and trusted me to help tell their stories.

During my time in grad school, many conversations, meetings, and writing sessions happened at Cloud Forest Café, and this space holds a special place in my heart. Thank you to T.H. and the rest of the Cloud Forest staff for providing the perfect space to think and discuss each Friday.

I am forever indebted to my family for giving me the chance to pursue higher education and for always nurturing my tenacity and self-esteem. Thank you to my parents, Dawn and Jim Lawson, for fostering my love of learning from a very young age and providing me every opportunity to pursue my dreams. Your belief in me has helped me to believe in myself. My brother, Nick, has been an amazing example for me and never once complained as I followed (sometimes too closely) in his footsteps.

I'm also incredibly fortunate to have such amazing friends who bring me copious amounts of joy and consistently remind me that there is so much more to life than academia.

Most of all, thank you to my partner and best friend, Ethan Caleb Prater-Fahey. Ethan has been the most wonderful source of personal and professional support over the past 9 years. He is the most thoughtful listener and I've yet to encounter an instance when his incisive feedback was not 100% accurate. Ethan's authenticity, diligence, and self-assuredness serves as daily inspiration for me. He has truly done everything possible to make my academic career a reality and I could not have done it without him. You're the best, Ethan!

Table of Contents

Abstract

Youth undergo numerous social, biological, and cognitive changes from late childhood through young adulthood. Many of these changes impact their temperament, or individual differences in reactivity and self-regulation that are present from an early age and relatively enduring. Temperament, in turn, is related to various domains of psychological adjustment, especially mental health. This dissertation is comprised of three chapters. In Chapter 1, I examined the latent structure of the most commonly used measure of adolescent temperament, the Early Adolescent Temperament Questionnaire-Revised (EATQ-R), and then showed that the resulting factors (i.e., Effortful Control, Negative Emotionality, and Positive Emotionality) had theoretically meaningful concurrent associations with several measures of adolescent functioning, supporting the construct validity of the EATQ-R. In Chapter 2, I examined how temperament from age 12 to 16 is associated with the onset of suicidal ideation and behaviors during adolescence and young adulthood. Finally, in Chapter 3, I examined how temperament develops across adolescence (age 10 to 16) and whether the developmental trajectories of temperament are associated with anxiety/depression during young adulthood (ages 19 and 21). Together, the results suggest that (1) the EATQ-R is a valid measure of adolescent temperament, (2) adolescents undergo meaningful temperament change across adolescence, and (3) these changes are associated with various aspects of psychological adjustment, especially mental health.

Chapter 1

The Structure of Adolescent Temperament and Associations with Psychological Functioning: A Replication and Extension of Snyder et al. (2015)

The content of this chapter has been previously published in the *Journal of Personality and Social Psychology*. I want to extend my thanks to the American Psychological Association for approving this copyrighted material to be used in my dissertation. Below is the citation for the corresponding published article.

Cite: Lawson, K. M., Atherton, O. E., & Robins, R. W. (2021). The structure of adolescent temperament and associations with psychological functioning: A replication and extension of Snyder et al. (2015). *Journal of Personality and Social Psychology, 121*(5), e19-e39

Abstract

The present study attempts to replicate and extend Snyder et al. (2015, *JPSP*). The original study examined the latent factor structure of the EATQ-R, a commonly used measure of adolescent temperament, and then showed that the resulting latent factors (i.e., Effortful Control, Negative Emotionality, and Positive Emotionality) had theoretically meaningful concurrent associations with several measures of adolescent functioning (depression, anxiety, ADHD, relational aggression, and school performance and behavior). We performed these same analyses using data from a large sample of Mexican-origin youth (*N*=674), and also examined prospective associations between the three EATQ-R factors and measures of adolescent functioning assessed two years later. We found some evidence supporting the bifactor model fit reported in the original study but poor replication of the correlations among latent factors. Additionally, model comparisons demonstrated that correlated factors models produced more interpretable factors than the bifactor models. In contrast, we replicated most of the concurrent correlations (and extended the findings to prospective associations) between the EATQ-R factors and measures of adolescent functioning, supporting the construct validity of the EATQ-R as a measure of adolescent temperament. Thus, these findings raise concerns about the generalizability of the factor structure identified by Snyder et al. (2015), but bolster claims about the generalizability of the concurrent and predictive validity of the EATQ-R. Overall, differences between the present findings and those of Snyder et al. highlight the importance of ongoing construct validation in youth temperament research, especially with participants from groups traditionally underrepresented in psychological research.

Introduction

The path to a cumulative, replicable science of personality requires the use of psychometrically sound measures (Flake, Pek, & Hehman, 2017). One area of personality psychology that is underdeveloped in terms of construct validation is youth temperament research, which examines individual differences in reactivity and self-regulation that are present from an early age and relatively enduring (Rothbart, 2007; Rothbart, 2011). There is no agreed upon taxonomy of the most important dimensions of youth temperament or consensus about which measures to use (Tackett & Durbin, 2017). This chaotic state impedes progress toward understanding how temperament differences develop during the highly consequential adolescent years. Recently, Tackett and Durbin (2017) warned the field that youth temperament research is "in dire need of more sophisticated and thoughtful measurement work", and proposed that researchers should "contribute to the long-standing but tedious journey to construct validity" (p. 1).

The present study heeds Tackett and Durbin's call to action and focuses on the construct validity of the most widely used measure of adolescent temperament - the *Early Adolescent Temperament Questionnaire – Revised* (EATQ-R; Ellis & Rothbart, 2001).[1] The EATQ-R was designed as a self- and informant report of temperament for youth aged from 9- to 15-years, but it is commonly used with older participants, including those up to 19-years old (e.g., Snyder et al., 2015).[2] The questionnaire assesses the three key constructs in Rothbart's highly influential temperament model (Rothbart, Ahadi, & Evans, 2000): effortful control (EC), negative emotionality (NE), and positive emotionality (PE). According to Rothbart's model, the EC domain involves one's capacity to plan and suppress inappropriate impulses (Inhibitory Control), perform an action or pursue goals when there are competing desires (Activation Control), and focus and shift attention when needed (Attention). The NE domain involves unpleasant affect derived from anticipating distress (Fear), negative affect related to ongoing tasks being interrupted (Frustration), and behavioral inhibition to social interaction (Shyness). The PE domain involves pleasure derived from high intensity or novel activities (High Intensity Pleasure/Surgency; hereafter referred to as Surgency), pleasure derived from low-intensity environmental stimulation (Pleasure Sensitivity), awareness of low-intensity environmental stimulation (Perceptual Sensitivity), and a desire for close, warm interpersonal connections (Affiliation). The EATQ-R also includes two additional scales: Aggression (hostile reactivity to negative stimuli including person- and object-directed violence) and Depressed Mood (unpleasant affect, lowered mood, and lack of enjoyment in activities). These scales were not considered part of any temperament domain in the original development of the EATQ-R, but they are conceptually linked to the NE domain, and are sometimes scored as part of the NE superordinate factor, along with the Fear, Frustration, and Shyness scales.

The development of the EATQ-R was guided by Rothbart and colleague's broad theory of temperament (e.g., Rothbart & Ahadi, 1994; Rothbart & Bates, 2006). This theory suggests that the structure of temperament is captured by reactivity (NE, PE) and regulation (EC), and specifies how each domain (as measured by the EATQ-R, or any other Rothbart-developed temperament questionnaire[3]) should be associated with real-world outcomes. In particular, internalizing problems (e.g., Anxiety, Depression) should be associated with high levels of NE (Brandes et al., in press; Nigg, 2006), whereas externalizing problems (e.g., ADHD) should be associated with high levels of NE and PE and low levels of EC (Brandes et al., in press; Eisenberg et al., 2007; Eisenberg et al., 2009; Muris, Meesters, & Blijlevens, 2007). Additionally, worse interpersonal functioning should be associated with low levels of EC and high levels of NE (Coplan & Bullock, 2012; De Bolle & Tackett, 2013; Eisenberg et al., 2009; Ojanen, Findley, & Fuller, 2012; Tackett et al., 2014). Further, better academic performance and school behavior should be associated with high levels of EC and low levels of NE (Valiente et al., 2013), but not associated with PE. Assuming Rothbart's theory is valid, the expected theoretical associations between the temperament domains and real-world outcomes allow researchers to evaluate the construct validity of her measure of temperament, the EATQ-R. Specifically, concurrent relations between temperament domains and life outcomes provide an opportunity to assess the extent to which there is *concurrent validity* of the EATQ-R, whereas temperament domains predicting future life outcomes assesses *predictive validity* of the measure. Past work on the concurrent and predictive validity of the EATQ-R has yielded mixed findings, in part because the field lacks consensus about how best to score the EATQ-R to assess the presumed underlying constructs (i.e., EC, NE, PE) (see Footnote 2 in Snyder et al., 2015).

The mixed findings in past research may also be due to problems with the internal structure of the measure. Notably, Rothbart's temperament measures were developed through a theory-driven, top-down approach with very little empirical, bottom-up validation (e.g., Kim, Brody, & Murry, 2003; Muris & Meesters, 2009). The limited work examining item-level analyses of Rothbart's temperament measures suggests that there are structural problems with many of Rothbart's temperament measures, including the EATQ-R. Kotelnikova and colleagues (2016) examined the structure of the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, & Hershey, 1994) and found little evidence for the theorized higher-order three-factor structure and many items did not load well onto their designated lower-order factor. Kotelnikova and colleagues (2017) examined these same properties in the Temperament in Middle Childhood Questionnaire (TMCQ; Simonds, 2006) and found similar problems with both the higher-order and the lower-order structure. Other than Snyder et al. (2015), only one study has examined the factor structure of the EATQ-R. In particular, Latham et al. (2020), using data from a large sample of Australian adolescents (10-12 years old), found poor fit for a hierarchical model that specified four higher-order factors (EC, NE, Surgency, Affiliativeness) along with subfactors corresponding to the EATQ-R subscales. However, they found that fit improved substantially when items were allowed to have cross-loadings on subfactors other than their designated subfactor (e.g., an Inhibitory Control item was allowed to load onto both the Inhibitory Control and the Activation Control subfactors), and when subfactors were allowed to have cross-loadings on domains other than their designated domain (e.g., Inhibitory Control was allowed to load onto both the EC and Affiliativeness domains). Together, these findings suggest that Rothbart's temperament measures have poor structural validity. However, despite these psychometric issues, there is still substantial evidence for the convergent and discriminant validity of the EATQ-R (Muris & Meesters, 2009), which is consistent with the notion that internal structure is only one element of construct validity (Hopwood & Donnellan, 2010; Loevinger, 1957).

Snyder et al. (2015)

In 2015, Snyder and colleagues published an article in the *Journal of Personality and Social Psychology* (*JPSP*) that endeavored to clarify the latent structure of the EATQ-R and the extent to which the underlying factors showed theoretically expected associations with various measures of adolescent functioning. Specifically, the authors used data from a large sample of adolescents ($N = 2{,}026$) to identify the factor structure of the EATQ-R via confirmatory factor analysis (CFA) and then examined associations between the resulting factors and concurrent measures of depression, anxiety, ADHD, peer interactions, and school grades and behavior. That is, Snyder and colleagues examined the extent to which the latent factors of the EATQ-R demonstrate concurrent validity with theoretically related life outcomes.

Snyder et al. (2015) used bifactor CFA models to assess the structure of the EATQ-R. These models allow researchers to study how items on a scale are related to both the general and domain-specific aspects of a broader latent construct (Chen, West, & Sousa, 2006; Reise, Moore, & Haviland, 2010) and are increasingly popular in research on personality and psychopathology (e.g., when studying the *p* factor; Caspi et al., 2014). In these models, each item can load onto a common factor, a domain-specific factor, or both a common and a domain-specific factor. These item loadings provide information about how to conceptualize the hierarchical structure of the construct. Further, the common and domain-specific factor scores can be related to external variables to better understand the psychological meaning of the shared (i.e., common) and unique (i.e., domain-specific) aspects of the construct.

Results from the best-fitting bifactor models in Snyder et al. (2015) generally support Rothbart's theoretical conception of the EC and NE domains. In particular, most of the variance of EC was captured by a common general factor, with the remaining variance being accounted for by a specific Activation Control factor. For NE, the model that best fit the data included a common general NE factor and several specific factors corresponding to each of the NE subscales (i.e., Fear, Frustration, Shyness, Aggression, Depressed Mood). However, the latent structure of PE was more complicated than expected. In particular, there was no psychologically coherent common factor for PE; instead, the Surgency items, which are arguably the theoretical core of the PE construct, formed a completely separate factor that was not strongly correlated (r = .29) with a common factor comprised of the remaining PE scales (Pleasure Sensitivity, Perceptual Sensitivity, Affiliation). The disjointed structure of PE led Snyder and colleagues to conclude that the PE subscale may actually assess two distinct constructs (i.e., surgency/sensation-seeking and general sensitivity to stimuli) rather than a single coherent domain of positive emotionality.

After examining the latent factor structure of the EATQ-R via bifactor models, Snyder et al. (2015) investigated the extent to which the resulting common and specific factors show concurrent associations with various indicators of adolescent functioning. Consistent with Rothbart's theoretical model and past empirical work, high EC was associated with decreased psychopathology (i.e., depression, anxiety, ADHD), decreased antisocial behavior toward and victimization by peers, higher grades, and fewer school discipline problems. Also consistent with theory and past work, the common NE factor was associated with more instances of relational aggression and increased psychopathology, with specific NE factors differentially predicting specific psychopathology symptoms. In particular, and not surprisingly, the specific Depressed Mood factor was associated with depression symptoms, the specific Fear factor was associated with Harm Avoidance symptoms of anxiety, and the specific Aggression factor was associated with more relational aggression, whereas there were no significant concurrent associations between measures of adolescent functioning and the Frustration-specific and Shyness-specific factors. Unlike EC and NE, the PE factors had more complex associations with adolescent functioning; for example, the Surgency factor was only associated with fewer separation/panic symptoms of Anxiety, whereas the common PE factor was positively associated with harm avoidance symptoms of Anxiety. Additionally, the specific Affiliation factor was positively associated with Anxiety and Depression symptoms as well as both antisocial behavior toward peers and victimization by peers. Further, the common PE factor was positively associated with relational aggression (both perpetration and victimization). Together, this pattern led Snyder and colleagues to conclude that PE as measured by the EATQ-R may be tapping into a broad sensitivity to one's environment, rather than positive emotionality as conceptualized by Rothbart and colleagues.

The work of Snyder et al. (2015) provides an important and timely empirical investigation of the latent structure of adolescent temperament as measured by the EATQ-R, as well as the extent to which the EATQ-R has concurrent validity with respect to relevant real-world outcomes. Moreover, the authors present results from their bifactor CFAs as an updated, empirically-derived scoring method for use in future research aiming to measure adolescent temperament via the EATQ-R. However, there are several limitations to Snyder et al.'s work, which we discuss below.

Methodological Limitations. Snyder et al. (2015) concluded that a bifactor model fits their EATQ-R data best. However, there is evidence of statistical bias in favor of bifactor models, even when another (simpler) model fits similarly well (Murray & Johnson, 2013). Thus, when comparing bifactor models with other models, it is especially important to consider multiple model fit indices that take model parsimony into account (Murray & Johnson, 2013), which Snyder et al. did not do. In addition, there is evidence that bifactor models are vulnerable to overfitting, meaning that "model fit statistics are unreliable indicators of the validity of bifactor models" (Watts, Poore, & Waldman, 2019, p. 3) and the resulting parameter estimates are often unstable and/or difficult to interpret. Together, these two issues suggest that model fit indices may have trouble differentiating between bifactor and other models, and so substantive issues should also be considered when determining which factor model to retain (Morgan, Hodge, Wells, & Watkins, 2015).[4]

These methodological issues highlight potential problems with Snyder et al.'s bifactor model findings. In particular, they used change in chi-square to compare models, which is notoriously sensitive to large sample sizes (Cheung & Rensvold, 2002), and only use a single fit index (RMSEA) that takes into account model parsimony (Hu & Bentler, 1998). Importantly, Snyder et al. found that the correlated factors model fit almost as well as the bifactor model (see Table 1, Snyder et al., 2015), which makes their selection of the bifactor model somewhat problematic given what we now know about biased fit indices and problems of overfitting with these models. The present study attempts to address some of these issues by comparing several different factor structures (i.e., a single-factor model, a correlated-factors model, a modified bifactor model, and a hierarchical model) using multiple fit indices (i.e., chi-square, RMSEA, CFI, AIC, BIC, and sample-size adjusted BIC) and attending to substantive concerns when adjudicating between models.

Given the widespread use of the EATQ-R and the potential for future research to implement this proposed scoring method, it is crucial to test the authors' claim that the findings of Snyder et al. (2015) are "robust and likely to generalize" (p. 1141). The present study seeks to examine the generalizability of the latent factor structure and concurrent validity findings of Snyder et al. (2015) in a large sample of Mexican-origin youth. Additionally, the present study aims to extend Snyder et al.'s findings by examining not only concurrent validity but also the predictive validity of the latent EATQ-R factors, using outcomes that allow us to evaluate both convergent and discriminant validity.

Replicability and Generalizability

Snyder and colleagues (2015) tested the replicability of their results by deriving the bifactor models using 50% of their sample and then examining the fit of these derived models in the remaining 50% of their sample (i.e., a randomly selected hold-out sample). This approach demonstrates, importantly, that the model fit of the latent structure of the EATQ-R replicates in a sample with the same characteristics as the derivation sample. However, it provides little insight into the generalizability of the findings, or whether the results depend on an originally unmeasured variable (e.g., ethnicity; Asendorpf et al., 2013). Generalizability is especially important for claims about construct validity as the degree to which a measure demonstrates construct validity varies by the specific use of the scale and "can often be context or population dependent" (Flake et al., 2017, p. 371). Because Snyder et al. (2015) found evidence for replicability, and replicability is necessary, but not sufficient, for generalizability, the next logical step is to examine the generalizability of the original findings in a sample that differs from the original sample in a measurable way (Finkel, Eastwick, & Reis, 2017).

In the present study, we examine the generalizability of both the derived latent factor model fit and associations with adolescent functioning in a sample of Mexican-origin youth. This study will help calibrate the theoretical conclusions drawn in Snyder et al. (2015) about the construct validity of the EATQ-R and facilitate the ongoing process of understanding how to best characterize adolescent temperament (Tackett & Durbin, 2017).

Hypotheses

Our goal for the present study was to examine the generalizability of the findings from Snyder et al. (2015). We expected to replicate Snyder et al.'s (2015) findings in a sample of Mexican-origin youth, thereby demonstrating the generalizability of the findings to an ethnic minority sample. Specifically, we expected to find: (1) adequate model fit (defined below) for the EC and NE bifactor models presented in Snyder et al. (2015); (2) adequate fit for a modified PE bifactor model[5]; (3) associations between the EC, NE, and PE latent factors (as specified by

Snyder et al., 2015) and measures of adolescent functioning that are in the same direction, and of similar magnitude, as those found in Snyder et al. (2015); and (4) evidence that an alternative scoring method tested by Snyder et al. (2015) shows weaker and less specific associations with adolescent functioning measures than the latent factors.

In addition to testing these hypotheses, we also extended Snyder et al. (2015) in several exploratory ways. First, we examined the influence of correlated residuals on model fit for the retained bifactor models. Second, given the methodological limitations of the original study, we conducted new model comparisons to see whether the bifactor models were a better conceptual and empirical fit to the data than three competing models (single factor, correlated factors, hierarchical factor). Third, given research documenting gender differences in temperament (Else-Quest et al., 2006), we tested for measurement invariance across gender in the structure of adolescent temperament. Fourth, we examined whether the derived latent factors predict adolescent functioning assessed two years after the EATQ-R data were collected. In other words, we extended Snyder et al. (2015) by examining predictive, as well as concurrent, validity.

Method

Participants

The present study used data from the California Families Project, a longitudinal study of Mexican-origin youth and their parents ($N = 674$).[6] Children were drawn at random from rosters of students from the Sacramento and Woodland, CA school districts, in 2006-07. The focal child had to be in the 5th grade, of Mexican origin, and living with his or her biological mother, in order to participate in the study. Approximately 72.6% of the eligible families agreed to participate in the study, which was granted approval by the University of California, Davis Institutional Review Board (Protocol # 217484-21). The children (50% female) were assessed annually from 5^{th} grade to two years post-high school. To most closely match the age of participants in Snyder et al. (2015), the present study used data from Wave 3, when the children were in 7^{th} grade ($M_{age} = 12.81$, $SD = 0.49$). To extend the analyses from Snyder et al. (2015) to include an examination of predictive validity, the present study also used data from Wave 5, when the children were in 9^{th} grade ($M_{age} = 14.75$, $SD = 0.49$). The retention rate relative to the first assessment (Wave 1, 5^{th} grade) was 86% at Wave 3 ($N = 586$) and 90% at Wave 5 ($N = 605$).

Participants were interviewed in their homes in Spanish or English, depending on their preference. Interviewers were all bilingual and most were of Mexican heritage. Sixty-three percent of mothers and 65% of fathers had less than a high school education (median = 9^{th} grade for both mothers and fathers); median total household income was between \$30,000 and \$35,000 (overall range of income = < \$5,000 to > \$95,000). With regard to generational status, 83.6% of mothers and 89.4% of fathers were 1^{st} generation, and 16.4% of mothers and 10.6% of fathers were either 2^{nd} or 3^{rd} generation. One hundred and twenty-four of the families were single-parent households (mothers only), and 549 of the families were two-parent households.

Compared to the participants in Snyder et al. (2015), participants in the present study differ in several ways (see Table 1.1). Notably, the majority of participants in the combined sample in Snyder et al. (2015) were White and lived in various geographic locations (i.e., Washington, Colorado, and New Jersey, United States; The Netherlands; Belgium) whereas all participants in the present study were Mexican-origin and living in northern California.

Table 1.1
Comparison of Sample Characteristics in Snyder et al. (2015) versus the Present Study

	Snyder et al. (2015)		Present study
	Factor Structure	Concurrent Validity	All Analyses
N	1013	562	586
Age	13.0 (2.57)	13.6 (2.36)	12.8 (0.49)
Gender (% female)	55%	55%	50%
Race/Ethnicity	Primarily White	Primarily (69%) White	Latinx (Mexican-origin)
Nationality	American, Dutch, Belgian	American	American
Geographic location	WA, CO, NJ, The Netherlands, Belgium	CO, NJ	Northern CA
Household income	No information	Median = \$86,500	Median = \$32,500
Parent Education Level	No information	No information	Median = 9^{th} grade

Note. Age is presented in the following format: Mean (*SD*). WA = Washington, CO = Colorado, NJ = New Jersey, CA = California.

Measures

The measures described in the present study are identical to those used in Snyder et al. (2015) unless otherwise indicated.

Temperament. Adolescent temperament was measured via adolescent self-reports using the *Early Adolescent Temperament Questionnaire – Revised* (EATQ-R; Ellis & Rothbart, 2001).[7] The EATQ-R was designed to measure three domains of temperament – EC ($M = 3.02$, $SD = 0.40$, $\alpha = .74$, $\omega = .72$), NE ($M = 1.96$, $SD = 0.38$, $\alpha = .84$, $\omega = .87$), and PE ($M = 2.70$, $SD = 0.40$, $\alpha = .55$, $\omega = .66$). Descriptive statistics for all subscales are shown in Table S1.1.

EC Subscales*.* The EC scale (16 items) has three facets: Activation Control (5 items), Attention (6 items), and Inhibitory Control (5 items). Activation Control assesses the ability to perform an action or pursue goals when there are competing desires. Attention assesses the ability to focus and shift attention when needed. Inhibitory Control assesses the ability to plan and suppress inappropriate impulses.

NE Subscales. The NE scale (17 items) has three facets: Fear (6 items), Frustration (7 items), and Shyness (4 items). Fear assesses unpleasant affect derived from anticipating distress. Frustration assesses negative affect related to ongoing tasks being interrupted. Shyness assesses behavioral avoidance of novelty and social challenges. Aggression (6 items) assesses hostile reactivity to negative stimuli including person- and object-directed violence. Depressed Mood (6 items) assesses unpleasant affect, lowered mood, and lack of enjoyment in activities. The Aggression and Depressed Mood scales also fall into the NE scale, as shown by Snyder et al. (2015).

PE Subscales. The PE scale (11 items) has two facets: Surgency (6 items) and Affiliation (5 items). Surgency assesses pleasure derived from activities involving high intensity or novelty. Affiliation assesses the desire for warmth and closeness with others. Notably, the PE measure in Snyder et al. (2015) also included four Perceptual Sensitivity items (assessing awareness of low-intensity stimulation in the environment) and five Pleasure Sensitivity items (assessing pleasure related to activities or stimuli involving low intensity), which were not assessed in the present study. Given the absence of these Perceptual and Pleasure Sensitivity items in the present study, we limit our findings about the EATQ-R PE domain to Surgency and Affiliation.

Adolescent functioning. Consistent with Snyder et al. (2015), adolescents reported on their depression, anxiety, and ADHD symptoms, as well as aggression towards and victimization by peers, whereas mothers completed reports of their child’s school behavior and performance. All adolescent functioning measures were assessed at both age 13 (for concurrent validity analyses) and age 15 (for predictive validity analyses).

Depression, Anxiety, and ADHD. To assess depression, anxiety, and ADHD symptoms, the present study used the NIMH Diagnostic Interview Schedule for Children-IV (*DISC-IV*). The *DISC-IV* is a comprehensive, psychiatric interview that assesses mental health problems for children and adolescents using DSM-IV criteria; it is the most widely-used mental health interview that has been tested in both clinical and community populations and validated in both English and Spanish (Costello, Edelbrock, & Costello, 1985; Schwab-Stone et al., 1996; translated into Spanish by Bravo, Woodbury-Farina, Canino, & Rubio-Stipec, 1993). For the present study, we used the Depression (22 items), Anxiety (14 items), and Attention-Deficit Hyperactivity Disorder (ADHD; 24 items) modules of the NIMH *DISC-IV*. Responses were recorded dichotomously (0 = *no*, 1 = *yes*) as the symptom being present or not in the past year. The Depression module included questions about feeling sad and irritable such as, "[Was there] a time in the past year when you were very upset or depressed?" and physical symptoms of depression such as, "[Did you] sleep more during the day than usual in the last year?". The Anxiety module included questions about general worry and concern such as, "[Are you the] type of person who is tense and finds it hard to relax?" and physical symptoms of anxiety such as, "[Did you] often have stomachaches in the last year?". The ADHD module included questions about attention-related behaviors such as, "[Did you have] trouble keeping your mind on task for more than a short period of time?" and hyperactivity problems such as, "[Did you] often climb on things/run around when you weren't supposed to?". For Depression, Anxiety, and ADHD, we computed a symptom count variable by summing the responses for each symptom (present vs. absent) to create separate composite scores of Depression, Anxiety, and ADHD. In addition, we computed symptom counts for the inattention (11 items) and hyperactivity (12 items) facets of ADHD.

Although the *DISC-IV* is similar to the measures Snyder et al. (2015) used to assess psychopathology, it is important to note that the DISC-IV generates symptom counts whereas the Snyder et al. measure used Likert-type continuous rating scales of Depression (27-item self-report Children's Depression Inventory; Kovacs, 1985), Anxiety (39-item Manifest Anxiety Scale for Children; March, Parker, Sullivan, Stallings, & Conners, 1997), and ADHD (18-item MTA SNAP-IV; Swanson et al., 2001). Consequently, we expected to have more zero-inflated Depression, Anxiety, and ADHD scores than Snyder et al. (2015), as discussed in personal communication with the original first-author (personal communication, May 10, 2019), which would likely result in smaller effect sizes.

***Antisocial behavior toward peers and victimization by peers*.** Adolescents completed a 12-item Relational Aggression scale (for details, see Aizpitarte et al., 2018; Atherton et al., 2017) that includes all seven items from the shortened Revised Peer Experiences Questionnaire (Prinstein, Boergers, & Vensberg, 2001) used by Snyder et al. (2015) to measure relational aggression. Participants completed two versions of the scale, one where they were asked about being a *victim* of relational aggression and the other where they were asked about being a *perpetrator* of relational aggression. Sample items include, "In the past three months, a kid your age told mean stories or lies about you." ["In the past three months, you told mean stories or lies about a kid your age."] and "In the past 3 months, a kid your age left you out of what he or she was doing." ["In the past 3 months, you left a kid your age out of what you were doing on purpose."]. Responses were made on a 4-point scale ranging from 1 (*almost never or never*) to 4 (*almost always or always*) and scores were calculated by averaging item responses for each version of the scale (i.e., mean victim score and mean perpetrator score for each participant).

School grades. School performance was measured using a one-item parent report assessing school grades, "On average, what are [ADOLESCENT'S] grades?" and the response scale ranging from 1(*mostly F's*) to 5 (*mostly A's*). This is the same item from Snyder et al. (2015), in which "Parents reported on their child's typical letter grades, from 1 = *mostly A's* to 5 = *mostly F's*" (p. 1136).

School behavior. School behavior was also assessed using a one-item parent report assessing adolescent misbehavior at school, "In the past 12 months, how frequently has [ADOLESCENT] been in trouble at school for things like arguing and fighting, being very disruptive in class, or other things like these?" and the response scale ranged from 1 (*never*) to 4 (*often*). This item is similar (but not exactly the same) as the item from Snyder et al. (2015), in which "Parents … reported the number of times their child had been sent to the office for misbehavior during the year from 1 = *none* to 6 = *more than five times*" (p. 1136).

Gender. At age 10, youth reported their gender (1 = girl; 2 = boy).

Data Analyses

All analyses were conducted in R version 3.6.0 (R Core Team, 2019) and RStudio Version 1.2.1335 using the *lavaan* (Rosseel, 2012) and *psych* packages (Revelle, 2018).[8] We used maximum likelihood (ML) estimation for all models, unless otherwise noted.

Replication of Snyder et al. EATQ-R factor structure analyses. Our first set of analyses focused on evaluating the fit of the latent structure of adolescent temperament reported in Snyder et al. (2015). Unless otherwise noted, the statistical methods detailed in this section are the same as those described in the Methods section of Snyder et al. (2015). Specifically, we tested the fit of the EC, NE, PE, and full bifactor models reported in Snyder et al. (2015) (Table S1.2 in the present study). To avoid overfitting to our data and to provide an accurate depiction of the generalizability of the findings from the original paper, no modifications were made from the original EC and NE bifactor models. As detailed above, we tested a modified PE bifactor model given that the present study excludes Perceptual Sensitivity and Pleasure Sensitivity scales. To fit the bifactor models, we used the “cfa” function in *lavaan*. Factor variance was set to 1 so that all factor loadings were estimated. Additionally, item loadings were standardized with respect to latent variable variance (i.e., std.lv = TRUE in *lavaan*). The same correlated residuals that were included in Snyder et al. (2015) were specified in our models (see Figures S1.1 – S1.3 for path diagrams).

Consistent with the thresholds outlined in the original article, absolute model fit was assessed as good if RMSEA < .05 and CFI > .95 and adequate if RMSEA < .08 and CFI > .90. Given the lack of consensus about rules of thumb for model fit indices (Marsh, Hau, & Wen, 2004) and complexities with interpreting model fit indices in factor analyses of personality measures (Hopwood & Donnellan, 2010), we report exact values for all fit indices. We also report (see Supplemental Materials, Tables S1.2-S1.3) several additional statistical indices for the bifactor models recommended by Rodriguez, Reise, and Haviland (2015) (see also Dueber, 2017), including omega hierarchical/omega hierarchical subscale, explained common variance (ECV), percent of uncontaminated correlations (PUC), item common variance attributed to a general dimension (I-ECV), a measure of construct reliability (i.e., H), and a measure of factor score determinacy (i.e., FD). In addition, we examined the congruence between the item-level factor loadings reported by Snyder et al. (2015) with those observed in the present study; that is, when rank-ordered by magnitude of the factor loading for each temperament domain (i.e., EC, NE, PE), how similar is the order of items between Snyder et al. and the present study. To do this, we calculated Pearson correlations between factor loadings from Snyder et al. and the present study.

After fitting the bifactor models for each domain, we estimated a full model with all three domains (EC, NE, and PE) together. This model allowed us to examine associations among all latent factors, both within (e.g., NE Fear-specific factor with NE Frustration-specific factor) and across temperament domains (e.g., NE Fear-specific factor with EC Activation Control-specific factor).

Model comparisons for different EATQ-R factor structures. Next, we attempted to account for limitations in the methods used by Snyder and colleagues to determine the best-fitting model of the factor structure of the EATQ-R in the present sample. Given that bifactor models can be erroneously identified as the best-fitting model when a more parsimonious model fits similarly well (e.g., Morgan et al., 2015), we compared a number of different CFA models. In particular, for each temperament domain (i.e., EC, NE, PE), we examined (1) a single-factor model where all of the items load onto a single domain factor, (2) a correlated-factors model where items load onto their respective subscale and then these latent subscale scores are allowed to correlate, (3) a bifactor model without correlated residuals and without excluding any EATQ-R items, and (4) a hierarchical model with one higher-order factor (e.g., EC) and the relevant subscales (e.g., activation control, attention control, and inhibitory control) as specific factors subsumed within the higher-order factor. Unlike Snyder et al. (2015), we did not remove items that load below .30, as doing so could reduce construct validity because items with low loadings may tap into relatively unique content that is nonetheless theoretically important to the construct. Two of these models (i.e., single-factor and correlated-factors models) were tested in Snyder et al. (2015) and deemed to exhibit worse fit than their retained bifactor model. However, Snyder and colleagues used change in chi-square as a primary method to compare model fit, which is problematic in large samples because even trivial differences in fit can be statistically significant (Cheung & Rensvold, 2002). Thus, to better compare the models (i.e., single-factor, correlated-factors, bifactor, hierarchical) in terms of model fit, we used the following indices – RMSEA, CFI, AIC, BIC, and sample-size adjusted BIC. We consider the best-fitting model to be the one that had the best fit in the majority of these five indicators. We also considered substantive concerns when adjudicating between models – in particular, having interpretable factor loadings (e.g., items loading in the expected direction) (Morgan et al., 2015). We then conducted all subsequent analyses using the best-fitting model from our model comparisons and the retained bifactor model from Snyder et al.[9]

Measurement invariance. As an extension of Snyder et al., we examined measurement invariance across gender. To do this, we compared a series of multiple group models. In particular, we examined four models: (a) freely estimating the factor loadings across for boys and girls (i.e., configural invariance); (b) constraining the respective factor loadings to be equal for boys and girls (i.e., weak invariance); (c) constraining the factor loadings and intercepts for boys and girls (i.e., strong invariance); and (d) constraining the factor loadings, intercepts, and residual variances for boys and girls (i.e., strict invariance). If the more constrained models did not fit worse than the lesser constrained models, then we concluded that the structure of the adolescent temperament is similar for boys and girls.

Correlations with adolescent functioning. To address questions about concurrent validity, we examined the concurrent correlations between the latent temperament factors and each measure of adolescent functioning. As an extension of Snyder et al. (2015), we also examined the predictive validity of the latent temperament factors by examining the correlations between the latent temperament factors at age 13 and each measure of adolescent functioning at age 15. For both concurrent and predictive validity analyses, we used the “rcorr” function in the *Hmisc* package (Harrell, 2019). For zero-inflated count data (i.e., psychiatric symptoms), we used Spearman’s *rho* correlations and for all other adolescent functioning variables, we computed Pearson’s *r* correlations.

Alternative scoring method. In addition to examining concurrent and longitudinal correlations between the derived latent factor structure and adolescent functioning outcomes, we also examined the concurrent and longitudinal associations between adolescent functioning outcomes and the alternative scoring method for the EATQ-R presented in Snyder et al. (2015). This alternative scoring method (also called the “traditional method” by Snyder et al.) differs from the other scoring methods examined in this study in two ways. First, the content included in each domain differs, such that EC is a composite of Attention, Activation Control, and Inhibition; NE is a composite of Fear, Aggression, Frustration, and Shyness (without Depressed Mood); and PE is a composite of Surgency and Affiliation (given that the present study excludes Pleasure Sensitivity and Perceptual Sensitivity). Second, this alternative scoring method uses observed (or manifest) composites for each domain rather than latent variables. Given that latent variables are free of non-systematic measurement error, correlations between observed variables should be smaller (Bollen, 2002; Borsboom, 2008). The results of all analyses examining this alternative scoring method are provided in the Supplemental Materials (see Table S1.8).

Power analyses. Power analyses for the correlations between derived latent factors and measures of adolescent functioning were performed using the *pwr* package (Champely, 2018). Using our sample size of 586 and setting alpha = .05 with a two-sided alternative hypothesis, we had 90% power to detect an effect size of $r = .13$ and 95% power to detect an effect size of $r =$

.15. This was sufficient power to detect all significant correlations in Snyder et al. (2015) (see italicized values in Table 1.7).

Results

Replication of Snyder et al. EATQ-R Factor Structure Analyses

We fit the bifactor models from Snyder et al. (2015) to each of the three temperament domains. Table 1.2 shows bifactor model fit statistics from the present study. Factor loadings and I-ECV for each item are shown in Table S1.2. Additional statistical indices (i.e., omega, omega hierarchical, ECV, PUC, H, and FD) are reported in Table S1.3.

Effortful Control. For EC, Snyder et al. found that the best-fitting latent factor model included a common EC factor and an Activation Control-specific factor, with no Inhibitory Control-specific or Attention-specific factors. Additionally, we removed item 41 ("You are good at keeping track of several different things that are happening around you") because it was excluded from Snyder et al.'s EC model for its weak loading (i.e., below .30) on the Attention subscale. When we fit this EC model to our data, we found that the model had poor fit by both CFI and RMSEA. However, the χ^2-value is almost identical to the one from the hold-out replication sample in Snyder et al. When we compared the factor loadings between the original study and the present study, we found a correlation of r = .31 for the Common EC factor and r = -.22 for the Activation-specific factor.

Table 1.2
Replication of Snyder et al. Bifactor Model Fit Statistics

Model	χ^2 (*df*)	χ^2/df	CFI	RMSEA
Effortful Control	411.65(81)* [414.10(81)*]	5.08 [5.11]	.72 [.84]	.084 [.066]
Negative Emotionality	853.57(343)* [1064.67(343)*]	2.49 [3.10]	.87 [.89]	.051 [.047]
Positive Emotionality	108.08(25)* [380.32(116)*]	4.32 [3.28]	.86 [.91]	.076 [.049]
Full Model	3243.15(1252)* [4693.71(1731)*]	2.59 [2.71]	.73 [.79]	.052 [.043]

Note. CFI = confirmatory fit index; RMSEA = root mean square error of approximation. Values in brackets are fit statistics from the hold-out sample in Snyder et al. (2015) for comparison. * $p < .001$.

Negative Emotionality. For NE, Snyder et al. (2015) found that the best-fitting latent factor model included a common NE factor and specific factors for all subscales (i.e., Fear, Frustration, Shyness, Aggression, and Depressed Mood). In the original article, item 37 ("You get sad when a lot of things are going wrong") had a weak negative loading on Depressed Mood and was eliminated from the Depressed Mood-Specific factor (but not the common NE factor), so we also loaded this item only on the common NE factor. When we fit this NE model to our data, we found that the model had poor fit via CFI and adequate fit via RMSEA (Table 1.2). Notably, the CFI and RMSEA fit indices are quite similar to the original study for the NE domain. When we compared the factor loadings between the original study and the present study, we found a correlation of $r = .58$ for the Common NE factor, $r = -.12$ for the Depressed Mood-specific factor, $r = .42$ for the Fear-specific factor, $r = .48$ for the Shyness-specific factor, $r = .73$ for the Frustration-specific factor, and $r = .87$ for the Aggression-specific factor.

Positive Emotionality. For PE, Snyder et al. (2015) found that the best-fitting latent factor model included a common PE factor, specific factors for Affiliation, Perceptual

Sensitivity, and Pleasure Sensitivity, and Surgency as a separate factor (not in the common PE factor). Given that the present study only included data on Affiliation and Surgency, we only included these two factors when fitting the PE model. In particular, we fit a model where Affiliation and Surgency formed two specific factors that were allowed to correlate, which is consistent with findings from Snyder et al. (2015) and with recommendations from the first-author (Dr. Hannah Snyder, personal communication, May 10, 2019). Additionally, we removed two Surgency items (item 3 "You think it would be exciting to move to a new city" and reverse-scored item 19 "You wouldn't like living in a really big city, even if it was safe") because they were excluded from Snyder et al.'s PE model for their weak loadings (i.e., below .30) on the Surgency subscale. When we fit this modified PE model to our data, we found that the model had poor fit via CFI and adequate fit via RMSEA (Table 2). When we compared the factor loadings between the original study and the present study, we found a correlation of $r = .92$ for the Surgency factor and of $r = .72$ for the Affiliation factor.[10]

Full combined model. Snyder et al. (2015) estimated a full model that included the EC, NE, and PE domains modeled together. Our full model was modified because it included the modified PE factors. When we fit this modified full model to our data, we found that the model had poor fit via CFI and adequate fit via RMSEA (Table 1.2), similar to Snyder et al.'s findings for their full model.

This full model allowed us to examine associations between all of the latent EC, NE, and PE factors. The original article found that EC was negatively correlated with most NE factors and that Surgency was negatively correlated with the Activation Control-Specific factor, the Fear-Specific, and the Shyness-Specific factor. Based on significance tests, these findings from Snyder et al. (2015) did not generalize to the present sample, as we found far fewer significant correlations (Table 1.3). Indeed, only 2 out of 14 (14%) significant correlations replicated between the original and present studies; in both studies there were significant negative associations of the common EC factor with the specific Aggression factor and the specific Depressed Mood factor. However, if we consider only the direction of the effect, we did find that 18 out of 22 (82%) of the correlations were in the same direction across the two studies.

Further, we found a number of significant associations between the Affiliation-specific factor and NE and EC factors. With the caveat that the Affiliation-specific factor was specific to our study and cannot be directly compared to results from the original study, the pattern of correlations does parallel those found by Snyder and colleagues (2015) for the common PE factor in terms of direction and significance.

Table 1.3
EATQ-R Full Model Estimated Factor Intercorrelations

	Common EC	Activation-specific	Surgency	Affiliation
Common NE	-.09 [-.36*]	-.72* [.30*]	.18 [-.25*]	.55*
Aggression-specific	-.36* [-.43*]	-.16 [-.17]	.04 [.17*]	-.27*
Depressed mood-specific	-.73* [-.40*]	.35 [.09]	-.03 [-.22*]	-.77*
Fear-specific	.17 [-.17]	.26 [.29]	-.18 [-.48*]	-.09
Frustration-specific	-.20 [-.41*]	-.22 [-.13]	-.20 [.17*]	-.26
Shyness-specific	-.13 [-.10*]	-.02 [-.08]	-.09 [-.21*]	-.33*
Common EC			.04 [.11]	.66*
Activation-specific			-.21 [-.26*]	-.62*
Surgency	.04 [.11]	-.21 [-.26*]		

Note. Blanks in the table indicate auto-correlations or factors constrained not to correlate (e.g., specific factors within each domain do not correlate with each other or with their Common factor because their shared variance is already captured by their Common factor). Values in brackets are coefficients from the hold-out dataset in Snyder et al. (2015) for comparison. *$p < .0005$.

Exploratory Bifactor Model Analyses

Given the generally suboptimal fit of the EC, NE, PE, and full bifactor models, we ran exploratory analyses to assess whether the model deficiencies were due, at least in part, to correlated residuals that were not being modeled. More specifically, we examined (1) if the correlated residuals from Snyder et al. (2015) actually increased model fit in the present study, and (2) whether adding new correlated residuals might lead to even better model fit in the present study. We believe these exploratory analyses are needed to thoroughly evaluate the generalizability of the findings from Snyder et al. (2015) because they added item residual covariances "until good model fit was achieved" (p. 1136). In other words, item covariances were included based on modification indices that reflect purely empirical (rather than theoretical) associations. Consequently, this practice is likely to capitalize on sample-specific associations and may decrease generalizability of the model when fit to other datasets.

First, to examine whether the correlated residuals from Snyder et al. (2015) actually increased model fit in the present study, we re-ran bifactor models with these residual covariances removed and compared fit indices to those of the original model. As another comparison, we also conducted analyses where we added correlated residuals between *random* pairs of items. (We ran these analyses three times, each with unique random pairs of items, and averaged the results across these three trials.) We found that the correlated residuals from Snyder et al. (2015) led to better model fit than not having any correlated residuals or having random correlated residuals, as the CFIs were higher and the RMSEAs were lower when fitting the original model versus the model without any correlated residuals and random correlated residuals (Table 1.4). Together, these results suggest that the data-dependent model modifications in Snyder et al. (2015) provide some generalizable benefit across samples.

Second, we examined whether there were different residual covariances that would result in better model fit for our data than the ones included in Snyder et al. (2015). To do this, we used the models with no correlated residuals and examined modification indices for residual covariances that would result in the greatest increase in model fit (i.e., reduce discrepancies between the observed and model-implied matrices). For consistency with Snyder et al. (2015), we included the same number of residual covariances that were included in their retained models (i.e., 4 residual covariances to EC, 6 residual covariances to NE, and 1 residual covariance with PE). Results indicate that the model with modified correlated residuals resulted in better fit for the EC and NE subscales and equal fit for the PE subscale.[11] These findings are not surprising, given that these model modifications were made via purely empirical (rather than theoretical) decisions. Further, the results highlight how data-dependent model modifications may generalize across multiple studies but they also might provide less benefit outside of a particular sample.

Table 1.4
Modified Bifactor Model Fit Statistics

Model	χ^2 (*df*)	χ^2/df	CFI	RMSEA
Effortful Control				
Original (Snyder et al.) model	411.65(81)*	5.08	.72	.084
No correlated residuals	446.71(85)*	5.26	.70	.086
Modified correlated residuals	368.68(81)*	4.55	.76	.078
Random correlated residuals	423.41(81)*	5.23	.71	.085
Negative Emotionality				
Original (Snyder et al.) model	853.57(343)*	2.49	.87	.051
No correlated residuals	906.04(349)*	2.60	.86	.053
Modified correlated residuals	785.71(343)*	2.29	.89	.047
Random correlated residuals	897.458(343)*	2.62	.86	.053
Positive Emotionality				
Original (Snyder et al.) model	108.08(25)*	4.32	.86	.076
No correlated residuals	135.42(26)*	5.21	.82	.085
Modified correlated residuals	108.08(25)*	4.32	.86	.076
Random correlated residuals	126.85(25)*	5.07	.83	.084

Note. CFI = confirmatory fit index; RMSEA = root mean square error of approximation. "Original (Snyder et al.) model" refers to the fit of models retained in Snyder et al. (2015) when evaluated using data from the present study. "No correlated residuals" refers to the models with no correlated residuals. "Modified correlated residuals" refers to the models with adjusted correlated residuals based on the largest modification indices. "Random correlated residuals" refers to models with random pairs of correlated residuals. * $p < .001$.

New Model Comparisons

Next, we conducted additional model comparisons examining single-factor, correlated factors, bifactor, and hierarchical factor models. Results from the new model comparisons for each of the temperament domains (EC, NE, and PE) are shown in Table 1.5. For EC, we found that the bifactor model fit the best given the statistical fit indices.[12] However, in some cases, the factor loadings in this model did not correspond to the presumed conceptual meaning of the factor. In particular, four items (two from the Attention-specific factor and two from the Inhibitory Control-specific factor) did not load in the expected direction. For example, the positively keyed item, "You can stick with your plans and goals", had a *negative* loading (-.15) on the Inhibitory Control-specific factor, making it difficult to interpret this factor as reflecting high levels of inhibitory control. For NE, the correlated factors model fit as well as (or better than) the bifactor model according to all of the fit indices, and better than both the single factor and hierarchical models. For PE, the bifactor model fit best based on the fit indices. However, this model was not positive definite because of a negative variance. The next best model was the correlated factors model, which fit slightly better than the hierarchical model.[13]

Overall, these model comparisons did not clearly identify a best-fitting model. For NE, the correlated factors model fit the best and is the most parsimonious. However, for EC and PE, the bifactor model fit best according to the statistical fit indices, but had some confusing factor loadings for EC and was not positive definite for PE. Given that the correlated factors model was the best or second-best model (considering both empirical and conceptual issues) for all three domains, and given that we are already examining the bifactor models from Snyder et al., we decided to retain the correlated factors models from our model comparisons for EC, NE, and PE (see Figures S1.4-S1.6 for path diagrams).

Table 1.5
New Model Comparison Fit Statistics

Temperament Domain	Model	$\chi^2(df)$	χ^2/df	CFI	RMSEA	AIC	BIC	Sample-size adjusted BIC
Effortful Control	Single factor	580.20(104)	5.58	.65	.089	22757	22967	22814
	Correlated factors	562.81(102)	5.52	.66	.088	22744	22962	22803
	Bifactor model	400.67(88)	4.55	.77	.078	22610	22889	22686
	Hierarchical model	--	--	--	--	--	--	--
Negative Emotionality	Single factor	2190.17(377)	5.81	.54	.091	39658	40037	39761
	Correlated factors	923.20(367)	2.52	.86	.051	38411	38834	38526
	Bifactor model	911.53(348)	2.62	.86	.053	38437	38943	38575
	Hierarchical model	1071.38(372)	2.88	.82	.057	38549	38950	38658
Positive Emotionality	Single factor	276.01(44)	6.27	.64	.096	16710	16854	16749
	Correlated factors	189.89(43)	4.42	.77	.077	16626	16774	16666
	Bifactor model	73.69(33)	2.23	.94	.046	16529	16721	16582
	Hierarchical model	189.89(42)	4.52	.77	.078	16628	16780	16669

Note. CFI = confirmatory fit index; RMSEA = root mean square error of approximation. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. Hierarchical EC model did not converge.

Measurement Invariance across Gender

Snyder et al. bifactor models. Results of the measurement invariance analyses for the retained bifactor models from Snyder et al. are shown in Tables S1.6. We found evidence for strict invariance for EC, strict invariance for NE, and weak invariance for PE.

Correlated factors models. Results of the measurement invariance analyses for the correlated factors models are shown in Tables S1.7. We found no form of invariance for EC (the configural invariance model did not converge), weak invariance for NE, and weak invariance for PE.

Concurrent and Prospective Associations with Measures of Adolescent Functioning

Snyder et al. Bifactor Models

Next, we used the bifactor models from Snyder et al. (2015) to examine the concurrent associations between the derived temperament scores and measures of adolescent functioning. Descriptive statistics (i.e., means, standard deviations, ranges) for all measures of adolescent functioning are shown in Table 1.6. 95% confidence intervals for correlations are presented in Table S1.14.

Table 1.6
Descriptive Statistics for Measures of Adolescent Functioning

	Concurrent Assessment (Age 13)			Predictive Assessment (Age 15)		
	Mean	*SD*	Range	*Mean*	*SD*	Range
Depression	3.99	3.83	0 - 18	4.08	4.22	0-18
Anxiety	2.37	1.77	0-11	1.96	1.84	0-10.5
ADHD – Total	2.02	2.47	0-14	2.03	2.47	0-15.5
ADHD – Inattention	1.01	1.32	0-8	1.07	1.42	0-8
ADHD – Hyperactivity	1.01	1.49	0-10	.97	1.41	0-8
Antisocial Interpersonal Functioning	1.08	.18	1-4	1.08	.18	1-4
Victim Interpersonal Functioning	1.14	.26	1-4	1.09	.20	1-4
School Grades	4.18	.91	1-5	3.95	1.08	1-5
School Discipline	1.23	.58	1-4	1.25	.60	1-4

Correlations of Effortful Control (EC) with adolescent functioning. Snyder et al. (2015) found that higher common EC scores were associated with fewer concurrent symptoms of depression, anxiety, and ADHD; less antisocial behavior toward peers and victimization by peers; better grades; and fewer disciplinary problems. On the whole, these findings generalized to the present sample (Table 1.7), with two exceptions; using the Bonferroni adjusted alpha level of .0003 from Snyder et al. (2015), we did not find a significant association between common EC scores and lower levels of victimization by peers (r = -.11, p = .006) or the common EC scores and fewer symptoms of anxiety (r = -.13, p = .002). Therefore, the vast majority of the concurrent associations found by Snyder and colleagues replicated in the present sample.

As an extension of Snyder et al. (2015), we examined associations between EC scores at age 13 and measures of adolescent functioning two years later (Table 1.7). We found similar, but weaker, prospective associations related to those associations we found concurrently. In particular, we found that that higher common EC scores at age 13 were associated with fewer ADHD symptoms ($r = -.20$, $p < .001$), less antisocial behavior toward peers ($r = -.18$, $p < .001$), better grades ($r = .23$, $p < .001$), and fewer disciplinary problems at age 15 ($r = -.22$, $p < .001$). We did not find any significant prospective associations between the Activation Control-specific factor and any of the adolescent functioning measures.

Table 1.7

Bifactor Models: Concurrent and Prospective Correlations between EATQ-R Factors and Measures of Adolescent Functioning (Compared with Snyder et al. Findings)

	Effortful Control	
	Common EC	Activation specific
Depression	-.33*[-.16*] /-.58*	-.16*[-.11] /.06
Anxiety	-.13 [-.05] /-.38*	-.05 [-.04] /-.17
ADHD – Total	-.39*[-.20*] /-.25*	-.20*[-.15] /.23
ADHD - Inattention	-.42*[-.24*] /-.21	-.18*[-.12] /-.07
ADHD – Hyperactivity	-.26*[-.10] /.15	-.17*[-.13] /.05
Antisocial Interpersonal Functioning	-.22*[-.18*] /-.45*	-.14 [-.09] /-.04
Victim Interpersonal Functioning	-.11 [-.14] /-.35*	-.06 [.04] /.07
School Grades	.31*[.23*] /.36*	.03 [.01] /-.06
School Discipline	-.18*[-.22*] /-.18*	-.01 [-.03] /-.03

	Negative Emotionality					
	Common NE	Aggression spec.	Depression spec.	Fear spec.	Frustration spec.	Shyness spec.
Depression	.38*[.26*]/.57*	.12 [.04] / .12	.27* [.20*] / .50*	-.06 [-.05] / -.14	.16* [.04] / -.07	-.04 [.02] / -.01
Anxiety	.41*[.23*]/.75*	-.06 [-.08] / -.09	.09 [.05] / .00	.18* [.07] / .05	.10 [.03] / -.18	.12 [.01] / .16
ADHD – Total	.30*[.28*]/-.08	.25*[.19*]/ .16	.19* [.09] / .21	-.14 [-.05] / -.01	.23* [.11] / .19	-.10 [-.06] / -.12
ADHD – Inattention	.29*[.23*]/-.02	.22*[.18*]/ .05	.18* [.09] / .19	-.09 [-.05] / .10	.20* [.09] / .14	-.04 [-.02] / .08
ADHD – Hyperactivity	.21*[.25*]/-.10	.22*[18] /-.12	.12 [.01] / .14	-.13 [-.04] / .08	.19* [.10] / .07	-.14 [-.11] / .10
Antisocial Int. Funct.	.21*[.15]/.36*	.32*[.32*]/.46*	.13 [.12] / .02	-.09 [-.10] / -.13	.14 [.08] / -.16	-.11 [-.11] / -.09
Victim Int. Funct.	.25*[.16*]/.34*	.08 [.06]/.27*	.20* [.14] / .19	.07 [.03] / -.05	.07 [.02] / -.12	-.03 [-.00] / -.09
School Grades	-.07 [.06] / -.11	-.15*[-.17*]/-.35*	-.11 [-.09] / -.10	.01 [.07] / -.25*	.00 [.03] / -.04	.05 [.05] / -.01
School Discipline	.00 [.07] / -.11	.19*[.17*]/.26*	.01 [.09] / .20	-.12 [-.08] / .04	.11 [.05] / .20	-.08 [-.06] / -.10

	Positive Emotionality	
	Surgency	Affiliation
Depression	.07 [.08] / -.15	.04 [.08]
Anxiety	.06 [.09] / -.14	.04 [.07]

35

ADHD – Total	**.10** [.18*] / *.12*	**.07** [.11]
ADHD - Inattention	**.08** [.15]/ *-.02*	**.03** [.05]
ADHD - Hyperactivity	**.09** [.12] / *.01*	**.09** [.12]
Antisocial Interpersonal Functioning	**.03** [.03] / *-.01*	**-.00** [.04]
Victim Interpersonal Functioning	**.04** [.01] / *.05*	**.03** [.04]
School Grades	**-.05** [.03] / *-.05*	**.11** [.19*]
School Discipline	**.03** [.01] / *.11*	**-.06** [-.07]

Note. Bolded values are concurrent associations from the present study. Values in brackets are prospective associations from the present study. Values in italics after the slash are concurrent coefficients from Snyder et al. (2015) for comparison. EATQ-R = Early Adolescent Temperament Questionnaire-Revised; EC = Effortful Control; NE = Negative Emotionality; PE = Positive Emotionality; ADHD = Attention deficit/hyperactivity disorder; Spec. = Specific; Int. Funct. = Interpersonal Functioning. *p < .0003

Correlations of Negative Emotionality (NE) with adolescent functioning. Snyder et al. (2015) found that higher common NE scores were associated with more symptoms of depression and anxiety, and more antisocial behavior toward peers and victimization by peers. These findings all generalize to the present sample (see Table 1.7). Additionally, we also found a positive significant association with common NE and ADHD ($r = .28$, $p < .001$). Snyder et al. (2015) found that higher Aggression-specific scores were associated with more antisocial behavior toward peers and victimization by peers, lower grades, and more school discipline problems. These associations generalize to the present sample, with the exception that we did not find evidence for higher Aggression-specific scores and more victimization by peers ($r = .07$, $p = .076$). In addition to the replicated findings, we also found that youth with higher Aggression-specific scores also had more ADHD symptoms ($r = .30$, $p < .001$). Further, Snyder et al. (2015) found that higher Depressed Mood-specific scores were associated with more depression and physical anxiety symptoms. We replicated the positive association between Depressed Mood-specific scores and depression symptoms ($r = .27$, $p < .001$) and also found that higher Depressed Mood-specific scores were associated with more ADHD symptoms ($r = .19$, $p < .001$) and more victimization by peers ($r = .20$, $p < .001$). Snyder et al. (2015) found that higher Fear-specific scores were associated with more anxiety symptoms and lower grades. We replicated the significant positive association between Fear-specific scores and anxiety symptoms ($r = .18$, $p < .001$). Unlike in the original article, we found significant associations between Frustration-specific scores with depression and ADHD symptoms. Finally, as in Snyder et al. (2015), we also found no association between the Shyness-specific scores and any measure of adolescent functioning.

The prospective associations were similar to (but weaker than) the concurrent associations for the common NE and Aggression-specific scores, with two exceptions (see Table 1.7). Using the Bonferroni adjusted alpha level of .0003 from Snyder et al. (2015), there was no association between common NE scores at age 13 and antisocial behavior towards peers at age 15 ($r = .15$, $p = .001$), and no association between Aggression-specific scores at age 13 and depression symptoms at age 15 ($r = .04$, $p = .294$). Furthermore, there were no prospective associations between Depressed-mood specific, Fear-specific, Frustration-specific, Shyness-specific scores and any of the measures of adolescent functioning except for a positive correlation between Depressed-mood specific scores at age 13 and depression symptoms at age 15 ($r = .20$, $p < .001$).

Correlations of Positive Emotionality (PE) with adolescent functioning. Snyder et al. (2015) found that Affiliation-specific scores were associated with higher Depression and Anxiety symptoms, more antisocial behavior toward peers, and more victimization by peers. These findings did not generalize to the present sample, as we found no significant concurrent associations between Affiliation and measures of adolescent functioning (Table 1.7). Similar to Snyder et al. (2015), we also found no significant associations between Surgency scores and measures of adolescent functioning.

Finally, Surgency and Affiliation did not have any significant prospective associations with any measure of adolescent functioning, except for a positive correlation between Surgency at age 13 and ADHD symptoms two years later ($r = .18$, p < .001) and Affiliation at age 13 and school grades two years later ($r = .19$, $p < .001$).

Comparison with an alternative method of scoring the EATQ-R temperament factors. In addition to assessing the three temperament domains using Snyder et al.'s latent factors, we also used an alternative scoring of the temperament domains and examined their concurrent and prospective associations with adolescent functioning. In particular, when Snyder et al. (2015) examined differences between their derived latent factors and the traditional scoring method, they found that the correlation patterns were much less specific and that contamination by common variance masked some specific effects. Results of these analyses are presented in Table S1.8. We found strong evidence of generalizability, with 41 out of 50 (82%) effects reported by Snyder et al. showing significant associations in the present study. In addition, we found 25 significant associations between manifest EATQ-R scores and concurrent measures of adolescent functioning that were not significant in Snyder et al. but were all in the same direction (Table S8).

The prospective associations showed a similar, but weaker, pattern (Table S1.8). Of the 41 concurrent associations from Snyder et al. that replicated in the present study, 30 associations (73%) also showed significant prospective associations in the present study.

Correlated Factors Models

Finally, we examined concurrent and prospective associations between factors scores from the correlated factors models with measures of adolescent functioning (Table 1.8; 95% confidence intervals are presented in Table S1.15). To help us interpret these results, and how they converge and diverge with results using the bifactor models, we present in Tables S1.9-S1.11 the correlations between the factor scores derived from the bifactor models with the factor scores derived from the correlated factors models.

Correlations of Effortful Control (EC) with adolescent functioning. The three EC factors (Activation Control, Inhibitory Control, and Attention) from the correlated factors model had associations with adolescent functioning that were very similar to those observed for the common EC factor from the Snyder et al. bifactor models. In particular, all three subscales were associated with symptoms of Depression and ADHD, interpersonal aggression, and school grades and discipline. Notably, using the correlated factors model, the Activation Control subscale factor showed the expected associations with school grades ($r = .30$, $p < .001$) and discipline ($r = -.18$, $p < .001$), whereas the Activation-specific factor from the Snyder et al. retained bifactor model did not (in either their study or the present study).

The correlations between the factors derived from the bifactor and correlated factors models provide some insight into why this discrepancy exists (see Table S1.9). In particular, there are very high correlations (*rs* = .96 - .98) between factor scores from the Common EC factor from the bifactor model and the three EC subscales from the correlated factors model. However, there is a relatively low correlation ($r = .15$) between the Activation-specific factor from the bifactor model and the Activation Control factor from the correlated factors model. Together, these findings suggest that the EC items are tapping into a shared effortful control factor resulting in similar associations across the subscales.

Correlations of Negative Emotionality (NE) with adolescent functioning. The Aggression, Depression, and Frustration factors from the correlated factors model showed very similar associations as the common NE factor from the bifactor model, especially for symptoms of Depression, Anxiety, and ADHD and interpersonal aggression and victimization. The Fear factor was associated with symptoms of Depression and Anxiety and interpersonal victimization, but only the association with Anxiety overlaps with the Fear-specific factor from Snyder et al.'s bifactor model. The Shyness factor was associated with Anxiety symptoms, whereas the Shyness-specific factor was not. This suggests that, in the bifactor model, the common NE factor is capturing much of the variance that is associated with adolescent functioning measures, leaving the specific factors with relatively less valid variance. Conversely, in the correlated factors model, each subscale factor is able to correlate more strongly with the theoretically related measures of adolescent functioning because the shared variance among the subfactors has not been removed.

With regard to the correlations between the factors from the bifactor and correlated factors models (see Table S1.10), we found moderate to strong correlations (*r*s = .53 - .95) between the Common NE factor scores and scores from each of the subscales from the correlated factors model, as well as between each specific factor and its corresponding factor from the correlated factors model (*r*s = .44 to .90).

Correlations of Positive Emotionality (PE) with adolescent functioning. As with Snyder et al.'s bifactor models, the Surgency and Affiliation factors from the correlated factors models were not significantly associated with any of the adolescent functioning measures with two exceptions: the Surgency factor was prospectively (but not concurrently) associated with total ADHD symptoms ($r = .19$, $p < .001$) and the Affiliation factor was prospectively (but not concurrently) associated with school grades ($r = .20$, $p < .001$). This pattern is consistent with the near-perfect correlations (*rs* = .98 - 1.00) between the factor scores from the bifactor and correlated factors models (see Table S1.11). Thus, regardless of the method of scoring the PE domain items, PE as assessed by the EATQ-R was not significantly associated with the eight measures of adolescent functioning examined in the present study.

Table 1.8

Correlated Factors Models: Concurrent and Prospective Correlations between EATQ-R Factors and Measures of Adolescent Functioning

	Effortful Control		
	Activation Control	Attention	Inhibitory Control
Depression	-.30* [-.16*]	-.29* [-.14]	-.28* [-.12]
Anxiety	-.09 [-.06]	-.10 [-.04]	-.11 [-.03]
ADHD - Total	-.38* [-.19*]	-.36* [-.16]	-.34* [-.14]
ADHD - Inattention	-.40* [-.23*]	-.39* [-.20*]	-.38* [-.19*]
ADHD - Hyperactivity	-.26* [-.11]	-.24* [-.07]	-.23* [-.05]
Antisocial Interpersonal Functioning	-.21* [-.16*]	-.20* [-.16*]	-.20* [-.16*]
Victim Interpersonal Functioning	-.10 [-.12]	-.10 [-.14]	-.10 [-.14]
School Grades	.30* [.23*]	.30* [.23*]	.29* [.23*]
School Discipline	-.18* [-.22*]	-.18*[-.21*]	-.17* [-.20*]

	Negative Emotionality				
	Aggression	Depressed Mood	Fear	Frustration	Shyness
Depression	.32* [.20*]	.41* [.29*]	.22* [.14]	.36* [.23*]	.12 [.13]
Anxiety	.17* [.07]	.41* [.22*]	.41* [.20*]	.36* [.18*]	.30* [.11]
ADHD - Total	.40* [.31*]	.30* [.27*]	.09 [.11]	.35* [.29*]	.03 [.06]
ADHD - Inattention	.37* [.28*]	.30* [.22*]	.13 [.08]	.33* [.25*]	.07 [.07]
ADHD - Hyperactivity	.32* [.26*]	.21* [.22*]	.05 [.10]	.27* [.24*]	-.03 [.00]
Antisocial Int. Funct.	.37* [.33*]	.22* [.16*]	.04 [-.01]	.26* [.18*]	-.01 [-.04]
Victim Int. Funct.	.17* [.12]	.27* [.18*]	.20* [.12]	.23* [.14]	.08 [.07]
School Grades	-.16* [-.12]	-.10 [.02]	-.02 [.10]	-.07 [.03]	.01 [.07]
School Discipline	.17* [.17*]	-.02 [.07]	-.10 [-.02]	.07 [.09]	-.09 [-.03]

	Positive Emotionality	
	Surgency	Affiliation
Depression	.08 [.08]	.02 [.04]
Anxiety	.07 [.08]	.04 [.07]
ADHD - Total	.11 [.19*]	.04 [.07]
ADHD - Inattention	.09 [.15]	-.00 [.02]
ADHD - Hyperactivity	.10 [.13]	.06 [.10]
Antisocial Interpersonal Functioning	.03 [.02]	-.02 [.02]

42

Victim Interpersonal Functioning	.03 [.01]	.02 [.02]
School Grades	-.05 [.03]	.13 [.20*]
School Discipline	.02 [.01]	-.09 [-.08]

Note. Values are concurrent associations from the present study. Values in brackets are prospective associations from the present study. EATQ-R = Early Adolescent Temperament Questionnaire-Revised; EC = Effortful Control; NE = Negative Emotionality; PE = Positive Emotionality; ADHD = Attention deficit/hyperactivity disorder; Spec. = Specific; Int. Funct. = Interpersonal Functioning. *p < .0005

Discussion

The present study aimed to replicate and extend Snyder et al.'s (2015) findings using data from a large sample of ethnic minority youth. Overall, we found weak evidence for the generalizability of the bifactor models reported in the original study for the EATQ-R temperament domains, with relatively poor model fit observed for the EC domain and adequate, but not good, fit observed for the NE and PE domains. Further, when we conducted new model comparisons, we found that correlated factors models produced more interpretable results in our data, although the EC and PE models did not fit well by traditional standards for fit indices. Together, these results suggest that the EATQ-R does not have a clear and replicable internal structure. In contrast, we replicated most, but not all, of the concurrent associations between temperament and adolescent functioning, and showed that these associations hold up longitudinally when predicting adolescent functioning two years later. These concurrent and prospective associations support the construct validity of the EATQ-R as a measure of adolescent temperament, despite its structural problems. Below we review and discuss these findings, and then turn to broader implications and directions for future research.

Do the Bifactor Models of the EATQ-R Temperament Domains Replicate?

Effortful control. Given the conceptual cohesiveness of the EC domain (i.e., all three subscales scales are theorized to be interrelated facets of a superordinate EC domain), we expected to find adequate fit for the bifactor EC model reported by Snyder et al. (2015). However, this model fit poorly in the present sample and we found weak correlations between the magnitude of the item loadings in Snyder et al. and the present study for both the Common EC factor (r = .30) and Activation-specific factor (r = -.22), suggesting that the EC bifactor model findings reported by Snyder et al. do not generalize well to our sample.

There are a number of potential explanations for this lack of generalizability. First, the EC factor structure may not generalize to our sample due to differences in age, ethnicity, SES, geographic location, and/or nationality between the samples (see Table 1.1). Second, and related to the first issue, the poor model fit may be due, at least in part, to acquiescence bias, or the tendency for participants to endorse items without regard to the actual content. In particular, acquiescence bias might have been a problem for the EC subscale where we found strong positive correlations between items keyed in the same direction, but much weaker associations between positively and negatively keyed items, suggesting the presence of a keying factor. Indeed, past research has found that low SES participants are more likely to exhibit acquiescence bias (Meisenberg & Williams, 2008) and participants in the present study come from lower SES backgrounds than the Snyder et al. participants. Therefore, this is consistent with the possibility that participants in the present sample may have been more likely to acquiesce than those in the original study. Third, the EC bifactor models may not robustly generalize because of underlying issues with bifactor models (Morgan et al., 2015; Murray & Johnson, 2013; Watts, Poore, & Waldman, 2019). The likelihood of this explanation is bolstered by the fact that modified residual covariances resulted in substantially better model fit of the EC model than the one that included the original residual covariances.

However, for EC, we also were unable to find a good-fitting factor structure when we evaluated the fit of single-factor, correlated factors, and hierarchical factor models, in addition to bifactor models. These findings dovetail with previous discussions about how Rothbart's temperament measures were developed through a theory-driven, top-down approach with little psychometric work to ensure a coherent internal structure (e.g., Kim, Brody, & Murry, 2003; Kotelnikova et al., 2016; Kotelnikova et al., 2017; Latham et al., 2020; Muris & Meesters, 2009).

Further, these findings are consistent with a warning from Hopwood and Donnellan (2010) that, due to the inherent complexity of personality data, failure to find adequate fit for CFA models of personality data is more of the rule than the exception. In fact, they found that, when examining seven prominent, well-validated personality measures using CFA, all had CFIs (ranging from .65 to .79) and RMSEAs (ranging from .09 to .13) that were not acceptable by traditional standards. In the present study, the model fit indices for the EC correlated factors model fit into these ranges (i.e., CFI = .66, RMSEA = .088). Further, these issues may be due, at least in part, to the fact that we fit CFAs to item-level data, rather than using composite scores or parcels. As Hopwood and Donnellan (2010) noted, personality items often “tap additional if substantially minor sources of variation”, which result in general model misfit if many correlated residuals are not included (p. 334). To address this issue, we ran an exploratory analysis using item parcels for the EC correlated factors model and found that model fit improved substantially (CFI = .93, RMSEA = .063). Another potential impediment to acceptable model fit derives from our use of ML estimation, which assumes interval-level data—a standard that personality test items arguably fail to reach. When we conducted the CFAs using an estimator (WLSMV) designed for use with ordinal data, we found improved model fit (see Table S1.4), especially for the EC model (CFI = .85, RMSEA = .077, vs. .72 and .084 using ML estimation). Together, these analyses indicate that CFA models of the structure of the EATQ-R can attain improved model fit through the use of item parcels or WLSMV estimation, otherwise researchers may have to rely on numerous correlated residuals to achieve adequate fit.

Negative emotionality. Unlike the EC model, we found that the NE bifactor model fit as well in the present sample as it did in Snyder et al.’s holdout replication sample, supporting the generalizability of the NE factor structure. In particular, we found nearly good fit via RMSEA (but not adequate fit via CFI) for the NE bifactor model. In addition, we found that the model fit improved only slightly when modified correlated residuals were included, which further supports Snyder et al.'s claim that NE is a conceptually coherent construct. Further, we found moderate to high correlations between the item loadings in Snyder et al. and the present study for the Common NE factor (r = .58) and the specific factors (Fear r = .72; Shyness r = .48, Frustration r = .73, and Aggression r = .87), except for Depressed Mood (r = -.12). In new model comparisons, we found that the correlated factors model fit slightly better than the bifactor model, and we found high correlations between the respective factor scores from both of these models. Finally, we found substantially improved – and good – model fit for both NE models using WLSMV estimation. Together, these findings suggest that we are closer to a generalizable factor structure of NE (as measured by the EATQ-R) than we are for the EC domain.

Positive emotionality. In contrast to the EC and NE domains, the implications of the results for PE are less clear. Snyder et al. did not find any coherent common factor for PE, and our findings do little to clarify the structure of this domain (as measured by the EATQ-R), in part because the present study included only the Surgency and Affiliation scales and omitted the Perceptual Sensitivity and Pleasure Sensitivity scales. Nonetheless, our findings do corroborate Snyder et al.'s finding that Surgency and Affiliation (along with Pleasure Sensitivity and Perceptual Sensitivity) form separate factors (see also Latham et al., 2020). Additionally, we found high correlations between the item loadings in Snyder et al. and the present study for both Surgency (r = .92) and Affiliation (r = .72). Further, our omission of the Pleasure Sensitivity and Perceptual Sensitivity may not have mattered much with regard to the structure of this domain, given that these scales likely would have loaded onto the Affiliation factor. In the new model comparisons, we retained the PE correlated factors model, which is practically identical to the PE bifactor model because in both cases there was no common factor shared between the Surgency and Affiliation subscales. This similarity is highlighted when examining the near-perfect correlations between the respective factor scores from both the retained Snyder et al. bifactor model and correlated factors model. Finally, it is worth noting that the Surgency construct (also called High Intensity Pleasure by Rothbart and colleagues) is the most closely aligned with Rothbart's theoretical conceptualization of Positive Emotionality, as well as other conceptualizations of PE (e.g., Watson, Clark, & Carey 1988), which tend to emphasize high arousal positive affect and reward sensitivity (Rothbart & Ahadi, 1994). In contrast, the content assessed by the Affiliation scale, and the two omitted scales measuring general sensitivity to stimuli, do not seem like core components of the PE domain.

Full model. We obtained poor model fit when we attempted to replicate Snyder et al.'s full model, which included the EC, NE, and PE domains modeled together. Despite finding better fit using WLSMV estimation, it remained poor fit by CFI and only adequate by RMSEA. However, it is worth noting that Snyder et al. also found poor fit for the overall model in their own data, so it is not surprising that this model did not generalize well to a sample that is different in so many ways from their sample. Nevertheless, this full model allowed us to examine associations between all of the latent EC, NE, and PE factors. Likely due at least in part to the poor model fit and low correlations between factor loadings from both studies, we only replicated 14% of the significant correlations replicated between the latent factors (82% of the correlations were in the same direction across the two studies). Together, these results further highlight that the EATQ-R lacks a coherent internal structure.

Measurement invariance. For the EC domain, we found evidence for strict measurement invariance across gender for the Snyder et al. bifactor model, but no measurement invariance for the correlated factors model. This suggests that, for the Snyder et al. bifactor model, the measurement parameters (i.e., factor loadings, intercepts, residual variances) are similar for both boys and girls (Van De Shoot et al., 2015; Putnick & Bornstein, 2016). For the NE domain, we found strict measurement invariance across gender for the Snyder et al. bifactor model and weak invariance for the correlated factors model. This suggests that, for Snyder et al. bifactor model, the measurement parameters (i.e., factor loadings, intercepts, residual variances) are similar for both boys and girls, whereas for the correlated factors model, we can only conclude that each item contributes to the latent constructs to a similar degree for both boys and girls. For the PE domain, we found weak measurement invariance across gender, which suggests that, for both models, each item contributes to the latent constructs to a similar degree for both boys and girls.

Do the Associations between Temperament and Adolescent Functioning Replicate?

We found strong evidence for the generalizability of the vast majority of the associations between the three temperament factors and measures of adolescent functioning.

Effortful control. For the EC domain, we replicated 5 out of the 7 (71%) significant concurrent associations with adolescent functioning outcomes found in Snyder et al.; specifically, youth higher in EC tended to show lower levels of depression, ADHD, and relational aggression, and better school grades and school behavior. For the five replicated effects, the median effect size was .36 in Snyder et al. and .31 in the present study. Further, of the 5 significant concurrent associations that we replicated, all 5 (100%) replicated prospectively, suggesting that these EC domains are related to important adolescent functioning outcomes not only at the same age but also two years later.

Moreover, we found 6 additional significant concurrent associations that were not found in the original study. Specifically, youth higher in common EC had higher levels of both the Inattention and Hyperactivity facets of ADHD and youth higher in the Activation-specific factor of EC had fewer symptoms of depression and ADHD (mirroring associations of the common EC factor and measures of adolescent functioning).

Negative emotionality. We replicated 8 out of the 10 (80%) significant concurrent associations between NE and adolescent functioning found in Snyder et al. For the 8 replicated effects, the median effect size was .41 in Snyder et al. and .26 in the present study. Seven of the eight (88%) significant concurrent correlations replicated prospectively, suggesting that these NE domains maintain their implications for adolescent functioning over time.

Moreover, we found 14 additional significant concurrent associations that were not found in the original study. Specifically, youth high in common NE and specific Aggression, Depression, and Frustration scores reported more ADHD symptoms. Additionally, youth with higher Depression specific scores were more likely to report being victims of relational aggression. Finally, youth with higher specific Fear reported more symptoms of Anxiety.

Altogether, the findings from the EC and NE domains are consistent with Snyder et al.'s claim that their construct validity analyses should generalize to other samples, and reinforce their statement that their derived EC and NE factors "revealed specific, theoretically predicted, and meaningful patterns of links with [the] outcome measures" (p. 1145).

Positive emotionality. In contrast to EC and NE, there was little to replicate in the PE domain. Snyder et al. (2015) found mostly null concurrent associations between PE and adolescent functioning and the present findings revealed a similar pattern of null effects for the two PE factors (i.e., Surgency, Affiliation), both concurrently and prospectively. Therefore, we should consider why PE, as measured by the EATQ-R, does not seem to be related to adolescent functioning in either Snyder et al. or the present study. As Snyder et al. allude to throughout their article, the lack of associations may not reflect a problem with the PE scale, but rather with the outcomes chosen to assess its construct validity. Snyder et al. avoid making clear predictions for associations between PE and adolescent functioning measures, though they do posit that higher levels of PE (conceptualized as Surgency) should be associated reward-oriented tendencies such as ADHD and conduct problems. Indeed, although none of the associations between the PE factors and measures of adolescent functioning were significant in the present study, the effect sizes were largest for ADHD. To explore whether PE is meaningfully related to other measures of adolescent functioning, future research should test whether PE is related to more theoretically relevant measures, such as substance use (Depue, Luciana, Arbisi, Collins, & Leon, 1994) or romantic relationship quality (Robins, Caspi, & Moffitt, 2002).

Broader Implications and Recommendations

Overall, the results of the present study suggest that we have yet to discover a generalizable latent structure of temperament as measured by the EATQ-R. These findings echo previous discussions about Rothbart’s theory-driven approach to developing temperament measures (Kim, Brody, & Murry, 2003; Kotelnikova et al., 2016; Kotelnikova et al., 2017; Latham et al., 2020; Muris & Meesters, 2009) and claims from Hopwood and Donnellan (2010), who demonstrated that fitting CFAs to well-validated personality measures often results in poor model fit according to conventional standards (in part because personality researchers typically fit CFAs to items rather than composite scores). In fact, the fit indices obtained in the present study for EC, NE, and PE were similar to, or exceeded, those reported by Hopwood and Donnellan for several other widely used personality measures. Thus, the EATQ-R is not unique in exhibiting relatively poor model fit. Further, our findings also highlight how ML estimation might be problematic given the measurement properties of item-level personality data, so using WLSMV estimation or item parcels can result in better model fit.

Despite the absence of a clear structure – at the domain level and at the level of the overall questionnaire – the replicability of most concurrent associations between temperament domains and adolescent functioning suggest that temperament, as assessed via the EATQ-R, is consistently related, both concurrently and prospectively, to theoretically relevant adolescent outcomes. Thus, the present study suggests that the EATQ-R has a rather poor factor structure (both within and across temperament domains) but reasonably good validity, at least for EC and NE.

These results also highlight the quagmire that occurs when measures of temperament such as the EATQ-R are not scored consistently across studies, which results in it being "difficult to compare the results and build a systematic, replicable knowledge base" (Snyder et al., 2015, p. 1134). However, it is equally important to ensure that researchers closely match the scoring procedure with their conceptualization of the construct being assessed. Indeed, the lack of a clear structure actually empowers researchers to more carefully consider theoretical issues when determining how they use the EATQ-R. Below, we delve into each of the three temperament domains assessed by the EATQ-R and provide recommendations for scoring these domains based on findings from Snyder et al., the present study, and theories about adolescent temperament.

Effortful control. The construct of effortful control lies within a larger nomological network of self-regulatory traits including self-control, executive function, and conscientiousness (Carver, 2005). Rothbart and her colleagues have clearly articulated their theoretical conceptualization of EC in numerous publications, and there is little question that the EATQ-R EC subscales (Inhibitory Control, Activation Control, and Attention) and item content closely map onto this theoretical conceptualization. Thus, the EC scale has strong content validity, despite its lack of a coherent factor structure. The EC scale also shows strong construct validity, as evidenced by its generalizable associations with theoretically relevant measures of adolescent functioning (i.e., lower levels of depression and ADHD, less antisocial interpersonal functioning, higher grades, less school discipline). Based on these theoretical considerations and empirical findings, we believe that the EC domain should be scored as a composite of Inhibitory Control, Activation Control, and Attention. These facets are all necessary components of the EC construct, as conceptualized by Rothbart, and a superordinate EC construct defined by all three components constitutes a theoretically, even if not empirically, coherent construct.

Notably, however, Snyder et al. found that the Activation-specific and common EC factors showed somewhat different concurrent associations with measures of adolescent functioning. The authors note that this pattern illustrates one way that EC parallels executive function, given that "there are both specific EF abilities and a common EF ability, which spans these components" (Snyder et al., 2015, p. 1143). The divergent findings for Activation-specific and common EC factors is consistent with previous research with the present sample showing that Activation Control has a different developmental trajectory across adolescence than overall EC and the Inhibitory Control and Attention Control facets (Atherton, Lawson, & Robins, in press). On the other hand, in the present study, we did not replicate the discriminant concurrent associations reported by Snyder et al. and instead found that the common EC and Activation-specific factors showed similar associations with measures of adolescent functioning, with common EC associations tending to be stronger than Activation-specific associations. Therefore, given this discrepancy, researchers using the EATQ-R should always examine the degree to which findings observed at the EC domain level replicate across the three facets, especially for Activation Control.

Negative emotionality. Negative emotionality, as assessed by the EATQ-R, is an empirically coherent construct where both the bifactor model and the vast majority of associations with adolescent functioning generalized from Snyder et al. to the present study. Therefore, evidence suggests that researchers can use Snyder et al.'s proposed scoring method and examine both a common NE core as well as specific scores for Fear, Frustration, Shyness, Aggression, and Depressed Mood. Further, because a correlated factors model also fit the NE scale well, there is evidence that conceptualizing NE as multiple interrelated facets is also appropriate.

However, one concern is that a composite of all five facets represents an overly broad mapping of the NE domain. Although many conceptualizations of NE encompass emotions like anger and hostility that can contribute to aggression, actual aggressive *behaviors* are generally not considered part of the NE construct. Thus, a case could be made for excluding the Aggression scale from the NE domain.[14] Similarly, Shyness is an interstitial trait that includes aspects of low Extraversion in addition to high Neuroticism (the Big Five domain most closely aligned with NE). Thus, if a researcher wants to assess the NE domain as conceptualized within a Big Five framework, it would be best to form a composite based on the Fear, Frustration, and Depressed Mood subscales of the EATQ-R, and exclude the Aggression and Shyness subscales. This seems consistent with Rothbart's conceptualization of NE as akin to "neuroticism, that is, a general tendency to experience and express negative emotions" (Rothbart & Ahadi, 1994, p. 57). Finally, the construct of Negative Affectivity (Watson, Clark, & Tellegen, 1988) represents an even narrower conceptualization of the NE domain that would exclude Depressed Mood and include only the Fear and Frustration scales (i.e., only high arousal negative affect). Therefore, when using the EATQ-R, researchers should not only examine both common/domain and specific/facet scores of NE, they should also be clear about their conceptualization of NE and choose only the facets that map onto this construct.

Positive emotionality. As we (and Snyder et al.) have discussed, positive emotionality, as assessed via the EATQ-R, is not empirically coherent. Luckily, this fragmentation allows for fairly simple scoring guidance. In particular, researchers should never create a common PE factor from the various EATQ-R subscales, but instead should always separately score Surgency and Affiliation (along with Perceptual Sensitivity and Pleasure Sensitivity, if assessed). This is consistent with Snyder et al.'s recommendation that, "If surgency is the construct of interest, only the Surgency subscale should be used" (Snyder et al., 2015, p. 1144), as well as with Rothbart's (Rothbart & Ahadi, 1994) and Watson et al.'s (1988) theoretical conceptualization of PE. This recommendation also makes sense from a Big Five perspective because PE is most closely aligned with Big Five Extraversion, which is also sometimes labeled Surgency. In contrast, the EATQ-R Affiliation scale (which assesses the "desire for warmth and closeness with others, independent of shyness or extraversion) is most closely related to the Agreeableness domain, and neither the Perceptual Sensitivity scale (which assesses "Detection or perceptual awareness of slight, low-intensity stimulation in the environment") nor the Pleasure Sensitivity scale ("Amount of pleasure related to activities or stimuli involving low intensity, rate, complexity, novelty, and incongruity") have a close conceptual connection to any Big Five domain, although they may be loosely connected to Openness to Experience (both scales) and Neuroticism (Perceptual Sensitivity only). Consequently, it does not make sense from a Big Five perspective to ever form a composite of the Affiliation, Perceptual Sensitivity, and Pleasure Sensitivity scales, because these scales are conceptually associated with distinct Big Five domains. Thus, we recommend that researchers should use the Surgency scale to assess the superordinate temperament domain of Positive Emotionality, and use the other scales only as measures of their respective specific constructs.

Limitations and Future Directions

The present study has several limitations. First, and most notably, the EATQ-R data in the present study do not include the Perceptual Sensitivity and Pleasure Sensitivity scales that were included in Snyder et al. (2015). Consequently, we were not able to directly test the generalizability of Snyder et al.'s derived factor structure of the PE domain and instead examine a modified factor structure. It is worth noting, however, that the PE subscales did not form a coherent general factor in Snyder et al. (2015), and had very few concurrent associations with theoretically relevant variables. Second, participants in the present study differ from those in the original study not only because of ethnicity, but also because of their geographic location, immigrant status, and SES (as well as other unmeasured variables). Therefore, the suboptimal bifactor model fit results might be due to the fact that the samples in the original and present study differ in important ways, and we do not have enough information to determine which differences led to the discrepancies in the findings. Third, Snyder et al. (2015) used self-reports with Likert-type scales to assess depression, anxiety, and ADHD, whereas the present study used diagnostic symptom counts derived from a structured psychiatric interview. As is typically found for symptom counts, there was less variance in the depression, anxiety, and ADHD scores in the present study compared to the original study. Nonetheless, we replicated 5 out of 6 (83%) of the significant concurrent associations between depression, anxiety, and ADHD symptoms and temperament scores and we also found 19 additional significant associations between these psychopathology measures and temperament as assessed via the EATQ-R. Despite these limitations, the present study helps to calibrate our confidence in the findings from Snyder et al.; in particular, the degree to which the three primary domains of temperament have both concurrent and prospective associations with various measures of psychological functioning.

Concluding Remarks

Moving forward, researchers should continue to evaluate the structure and construct validity of the EATQ-R and examine generalizability across more diverse samples. Construct validation is a never-ending iterative process, especially when an existing scale is being used in a new context or population (Flake et al., 2017). Indeed, ongoing construct validation is especially important in research using data from participants traditionally underrepresented in psychological research, and Hernández and colleagues highlight its importance as "one of three essential components of cultural validation…" (Hernández, Nguyen, Casanova, Suárez-Orozco, & Saetermoe, 2013, p. 49). Thoughtful measurement work will also help to lay a better foundation for future applied research on youth temperament, especially with ethnic-minority samples. One method for continuing this process, as demonstrated by the present study, is taking advantage of existing longitudinal data to conduct replication-focused research. Because replications typically entail the exact analyses conducted in the original study, problems with selective reporting and *p*-hacking are minimized with replications conducted using existing data. Similarly, the hypotheses for the replication are based on findings from the original study, reducing concerns about hypothesizing after the results are known (i.e., HARKing). Further, although direct or exact replications are a vital step toward improving psychological research, the present study demonstrates that replication studies should also work to improve on previous research and not solely replicate what was done. This is especially true when there are conceptual or methodological limitations with the original research, as there often are, and when new concerns (e.g., about the limits of bifactor models) have arisen since publication of the original article. In particular, we recommend that replication studies always repeat the exact analyses conducted in the original study, which is necessary for robust examination of the replicability and generalizability of the findings, but also conduct additional (often exploratory) analyses to redress the limitations of the original study and explore alternative explanations for the findings. Together, an exact replication combined with a thoughtful extension aimed at ameliorating problems with the existing research will provide a more nuanced understanding of the replicability and generalizability of the findings.

Snyder and colleagues (2015) took an important step towards heeding Tackett and Durbin's (2017) call to focus on measurement and construct validity in youth temperament research. The present study follows in their footsteps to contribute to a more robust, generalizable science of adolescent temperament. However, the "tedious journey to construct validity" continues onward.

Supplemental Materials

Table S1.1

Descriptive Statistics for the EATQ-R Subscales

Subscale	Mean	SD	α / ω
Activation control	2.90	.56	.58 / .66
Attention	3.03	.44	.48 / .65
Inhibitory control	3.14	.48	.37 / .49
Fear	2.33	.60	.65 / .74
Frustration	2.20	.62	.80 / .85
Shyness	2.14	.70	.72 / .78
Aggression	1.39	.46	.79 / .84
Depressed Mood	1.76	.46	.59 / .70
Surgency	2.58	.50	.39 / .56
Affiliation	2.84	.57	.66 / .72

Note. SD = Standard deviation. α = alpha reliability. ω = omega reliability.

Table S1.2
Factor Loadings from Best-fitting Bifactor Models from Snyder et al. (2015) Fit to Data from the Present Study

Effortful Control		Factor Loading (Common EC)	Factor Loading (Activation-Specific)	IECV	Items
Activation	7R	.376	.152	.860	You have a hard time finishing things on time. (R)
	18R	.278	.473	.257	You do something fun for a while before starting your homework, even when you're not supposed to. (R)
	30	.532	.024	.998	If you have a hard assignment to do, you get started right away.
	39	.509	-.074	.979	You finish your homework before the due date.
	49R	.238	.448	.220	You put off working on projects until right before they're due. (R)
Attention	1	.449	-	1.00	It is easy for you to really concentrate on homework problems.
	34R	.182	-	1.00	You find it hard to shift gears when you go from one class to another at school. (R)
	38R	.210	-	1.00	When trying to study, you have difficulty tuning out background noise and concentrating. (R)
	59	.441	-	1.00	You pay close attention when someone tells you how to do something.
	61R	.279	-	1.00	You tend to get in the middle of one thing, then go off and do something else. (R)
Inhibitory	10R	.112	-	1.00	It's hard for you not to open presents before you're supposed to. (R)
	14	.384	-	1.00	When someone tells you to stop doing something, it is easy for you to stop.
	26R	.239	-	1.00	The more you try to stop yourself from doing something you shouldn't, the more likely you are to do it. (R)
	43	.237	-	1.00	It's easy for you to keep a secret.
	63	.444	-	1.00	You can stick with your plans and goals.
Removed (*Att.*)	41				You are good at keeping track of several different things that are happening around you.

Original Correlated Residuals				7R & 30, 38R & 59, 10R & 26R, 26R & 63
Modified Correlated Residuals				26R & 61R, 38R & 61R, 10R & 61R, 59 & 63
Negative Emotionality	Factor Loading (Common NE)	Factor Loading (Specific)	IECV	Items
Aggression 5	.192	.457	.150	If you are mad at somebody, you tend to say things that you know will hurt their feelings.
9	.128	.272	.181	When you are angry, you throw or break things.
13	.219	.502	.160	If you get really mad at someone, you might hit them.
22	.232	.384	.267	You tend to be rude to people you don't like.
50	.260	.353	.352	When you're really mad at a friend, you tend to explode at them.
58	.121	.226	.223	You pick on people for no real reason.
Depressed Mood 2R	.032	.186	.029	You feel pretty happy most of the day. (R)
11	.174	.152	.567	Your friends seem to enjoy themselves more than you do.
20	.354	.272	.629	It often takes very little to make you feel like crying.
29	.363	.399	.453	It often takes very little to make you feel like crying.
37	.550	-	1.00	You get sad when a lot of things are going wrong.
55	.387	.255	.697	You feel sad even when you should be enjoying yourself, like at Christmas or on a trip.
Fear 32	.234	.385	.270	You get frightened riding with a person who likes to speed.
35	.299	.327	.455	You worry about your family when you're not with them.
40	.375	.463	.396	You worry about getting into trouble.
46	.332	.271	.600	You are nervous of some of the kids at school who push people into lockers and throw your books around.
51	.383	.470	.399	You worry about your parent(s) dying or leaving you.
57	.425	.089	.958	You feel scared when you enter a darkened room at home.

Frustration	25	.333	.286	.575	It bothers you when you try to make a phone call and the line is busy.
	36	.495	.268	.773	You get very upset if you want to do something and your parents won't let you.
	47	.492	.321	.701	You get very upset if you want to do something and your parents won't let you.
	56	.450	.483	.465	It really annoys you to wait in long lines.
	60	.473	.181	.872	It really annoys you to wait in long lines.
	62	.447	.374	.588	It frustrates you if people interrupt you when you're talking.
	64	.444	.336	.636	You get upset if you're not able to do a task really well.
Shyness	8	.287	.305	.470	You feel shy with kids of the opposite sex.
	15	.371	.592	.282	You feel shy about meeting new people.
	45	.317	.679	.179	You are shy.
	53R	.047	.526	.008	You are not shy. (R)
Original Correlated Residuals					20 & 37, 29 & 37, 32 & 51, 35 & 51, 45 & 53R, 60 & 64
Modified Correlated Residuals					36 & 37, 60 & 64, 32 & 46, 2R & 35, 9 & 13, 35 & 37

Positive Emotionality			Factor Loading		Items
Affiliation					
	17	-	.447	-	You want to be able to share your private thoughts with someone else.
	27	-	.561	-	You enjoy exchanging hugs with people you like.
	31	-	.332	-	You will do most anything to help someone you care about.
	44	-	.566	-	It is important to you to have close relationships with other people.
	54	-	.320	-	You are quite a warm and friendly person.
Surgency					
	28R	-	.192	-	Skiing fast down a steep slope sounds scary to you. (R)
	42	-	.858	-	You would not be afraid to try a risky sport, like deep-sea diving.
	48	-	.592	-	You wouldn't be afraid to try something like mountain climbing.

52	-	.148	-	You enjoy going places where there are big crowds and lots of excitement.
Removed (*Surg.*) 3				You think it would be exciting to move to a new city.
Removed (*Surg.*) 19				You wouldn't like living in a really big city, even if it was safe. (R)
Original Correlated Residuals				31 & 54
Modified Correlated Residuals				31 & 54

Note. Removed = items excluded from Snyder et al. bifactor models. IECV = item explained common variance. IECV "provides the extent to which an item's responses are accounted for by variance on the latent general dimension alone" (Stucky et al., 2013, p. 51). Higher IECV values indicate greater unidimensionality of the common factor.

Table S1.3
Additional Bifactor Model Statistics

	ECV	Omega/OmegaS	OmegaH/OmegaHS	H	FD	PUC
Effortful Control						.905
Common EC	.802	.664	.637	.694	.831	
Activation-specific	.198	.562	.123	.363	.608	
Negative Emotionality						.835
Common NE	.466	.829	.672	.807	.861	
Aggression-specific	.783	.556	.436	.512	.713	
Depressed Mood-specific	.448	.441	.212	.285	.535	
Fear-specific	.517	.646	.316	.484	.687	
Frustration-specific	.353	.756	.257	.476	.648	
Shyness-specific	.785	.686	.555	.653	.814	
Positive Emotionality						-
Surgency	.522	-	-	.772	-	
Affiliation	.478	-	-	.586	-	

Note. All statistics calculated using Dueber (2017) Excel-based calculator. ECV = Explained common variance, which is the proportion of common variance explained by the factor. Omega/OmegaS = model-based estimate of internal reliability for the general factor (i.e., including all items; Omega) or the specific factors (i.e., including items loading onto the specific factor; OmegaS). OmegaH/OmegaHS= Omega Hierarchical, which estimates how much variance can be attributed to a general (i.e., OmegaH) or specific (i.e., OmegaHS) factor (Reise et al., 2013). H = measure of construct replicability. FD = Factor Determinacy, which is the correlation between factor scores and the factors. PUC = Percent of uncontaminated correlations.

Table S1.4

Modified Bifactor Model Fit Statistics using a WLSMV estimator

Model	χ^2 (*df*)	χ^2/df	CFI	RMSEA
Effortful Control	345.73 (81)	4.27	.85	.077
Negative Emotionality	895.24(343)	2.61	.92	.054
Positive Emotionality	131.02(25)	5.24	.85	.087
Full	4156.13(1252)	3.32	.80	.066

Table S1.5

New Model Comparison Fit Statistics using a WLSMV estimator

Temperament Domain	Model	$\chi^2(df)$	χ^2/df	CFI	RMSEA
Effortful Control	Single factor	493.10(104)	4.74	.81	.083
	Correlated factors	486.73(102)	4.77	.81	.083
	Bifactor model	371.95(88)	4.23	.86	.077
	Hierarchical model	-	-	-	-
Negative Emotionality	Single factor	1799.49(377)	4.77	.78	.083
	Correlated factors	671.25(367)	1.83	.95	.039
	Bifactor model	914.23(348)	2.63	.91	.054
	Hierarchical model	1087.19(372)	2.92	.89	.059
Positive Emotionality	Single factor	178.87(44)	4.07	.82	.074
	Correlated factors	177.69(43)	4.13	.82	.075
	Bifactor model	50.53(33)	1.53	.98	.031
	Hierarchical model	177.69(42)	4.23	.82	.076

Note. CFI = confirmatory fit index; RMSEA = root mean square error of approximation. AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. Hierarchical EC model did not converge.

Table S1.6

Model Fit Indices for Snyder et al. Bifactor Models for Testing Measurement Invariance across Gender

	χ^2 (*df*)	CFI	RMSEA
Effortful Control			
Configural	500.91(162)	.72	.085
Weak	541.95(180)	.70	.083
Strong	549.23(193)	.71	.080
Strict	565.82(208)	.71	.077
Negative Emotionality			
Configural	-	-	-
Weak	1388.90(737)	.84	.055
Strong	1462.47(760)	.83	.057
Strict	1512.40(789)	.82	.056
Positive Emotionality			
Configural	144.78(50)	.84	.081
Weak	152.02(57)	.84	.076
Strong	179.88(64)	.81	.079
Strict	190.93(73)	.81	.075

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; χ^2 = chi-square; df = degrees of freedom.

Table S1.7

Model Fit Indices for Correlated Factors Models for Testing Measurement Invariance across Gender

	χ^2 / *df*	CFI	RMSEA
Effortful Control			
Configural	-	-	-
Weak	-	-	-
Strong	-	-	-
Strict	-	-	-
Negative Emotionality			
Configural	1420.17(734)	.83	.057
Weak	1454.17(758)	.83	.056
Strong	1555.89(782)	.81	.059
Strict	1610.36(811)	.80	.058
Positive Emotionality			
Configural	236.82(86)	.77	.078
Weak	256.47(95)	.75	.077
Strong	288.36(104)	.71	.078
Strict	301.57(115)	.71	.075

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; χ^2 = chi-square; df = degrees of freedom.

Table S1.8

Estimated Factor Correlations Between EATQ-R Manifest Variables and Concurrent and Predictive Measures of Adolescent Functioning (Compared with Snyder et al. Findings)

	Effortful Control			
	EC	Activation	Attention	Inhibition
Depression	__-.38*__ [-.20*] / _-.51*_	__-.30*__ [-.20*] / _-.45*_	__-.34*__ [-.16*] / _-.44*_	__-.26*__ [-.10] / _-.39*_
Anxiety	__-.20*__ [-.11] / _-.14_	__-.08__ [-.09] / _-.10_	__-.23*__ [-.12] / _-.16*_	__-.17*__ [-.01] / _-.10_
ADHD - Total	__-.45*__ [-.24*] / _-.19*_	__-.39*__ [-.23*] / _-.15*_	__-.40*__ [-.21*] / _-.24*_	__-.28*__ [-.13] / _-.10_
ADHD - Inattention	__-.46*__ [-.27*] / _-.23*_	__-.40*__ [-.24*] / _-.21*_	__-.41*__ [-.23*] / _-.28*_	__-.28*__ [-.16] / _-.12_
ADHD - Hyperactivity	__-.33*__ [-.14] / _-.13_	__-.27*__ [-.16] / _-.09_	__-.27*__ [-.12] / _-.18*_	__-.22*__ [-.06] / _-.04_
Antisocial Interpersonal Funct.	__-.26*__ [-.20*] / _-.39*_	__-.22*__ [-.16*] / _-.29*_	__-.19*__ [-.15] / _-.33*_	__-.21*__ [-.17*] / _-.35*_
Victim Interpersonal Funct.	__-.16*__ [-.14] / _-.27*_	__-.10__ [-.06] / _-.21*_	__-.17*__ [-.17*] / _-.25*_	__-.11__ [-.10] / _-.24*_
School Grades	__.28*__ [.20*] / _.27*_	__.24*__ [.18*] / _.25*_	__.24*__ [.13] / _.24*_	__.22*__ [.17*] / _.19*_
School Discipline	__-.16*__ [-.22*] / _-.17*_	__-.15*__ [-.18*] / _-.15*_	__-.08__ [-.17*] / _-.12_	__-.14*__ [-.17*] / _-.15_

	Negative Emotionality					
	NE	Aggression	Depression	Fear	Frustration	Shyness
Depression	__.34*__ [.23*] / _.53*_	__.28*__ [.17*] / _.43*_	__.45*__ [.31*] / _.60*_	__.13__ [.07] / _.23*_	__.32*__ [.21*] / _.35*_	__.11__ [.13] / _.24*_
Anxiety	__.42*__ [.20*] / _.59*_	__.13__ [.05] / _.17*_	__.33*__ [.18*] / _.48*_	__.34*__ [.15] / _.58*_	__.33*__ [.18*] / _.41*_	__.24*__ [.08] / _.36*_
ADHD - Total	__.28*__ [.29*] / _.01_	__.36*__ [.29*] / _.12_	__.32*__ [.23*] / _.04_	__.03__ [.07] / _-.05_	__.33*__ [.27*] / _.04_	__.02__ [.06] / _-.07_
ADHD - Inattention	__.29*__ [.24] / _.04_	__.33*__ [.26*] / _.12_	__.31*__ [.20*] / _.06_	__.06__ [.04] / _-.04_	__.31*__ [.23*] / _.04_	__.07__ [.07] / _-.03_
ADHD - Hyperactivity	__.19*__ [.23*] / _-.02_	__.29*__ [.25*] / _.08_	__.21*__ [.17] / _.02_	__-.00__ [.07] / _-.07_	__.26*__ [.22*] / _.01_	__-.04__ [.00] / _-.08_

Antisocial Int. Funct.	**.23*** [.15] / _.23*_	**.37*** [.34*] / _.39*_	**.21*** [.16*] / _.17*_	**.01** [-.01] / _-.03_	**.22*** [.14] / _.15_	**-.03** [-.05] / _.12_
Victim Int. Funct.	**.22*** [.14] / _.31*_	**.16*** [.12] / _.32*_	**.27*** [.19*] / _.31*_	**.15*** [.10] / _.11_	**.19*** [.12] / _.16*_	**.04** [.03] / _.16*_
School Grades	**-.08** [.02] / _-.01_	**-.16*** [-.13] /_-.17*_	**-.11** [-.02] / _.03_	**-.02** [.08] / _.00_	**-.04** [.05] / _.06_	**-.02** [.03] / _.05_
School Discipline	**.02** [.08] / _-.02_	**.17*** [.18*] / _.17*_	**-.01** [.09] / _-.02_	**-.10** [-.02] / _-.08_	**.07** [.07] / _.03_	**-.08** [-.01] / _-.13_

	Positive Emotionality		
	PE	Surgency	Affiliation
Depression	**.06** [.06] / _-.04_	**.06** [.03] / _-.07_	**.03** [.06] / _.05_
Anxiety	**.03** [.04] / _.15_	**-.00** [-02] / _-.22*_	**.05** [.07] / _.17*_
ADHD - Total	**.10** [.15] / _-.04_	**.12** [.14] / _.07_	**.04** [.08] / _-.13_
ADHD - Inattention	**.06** [.08] / _-.03_	**.09** [.11] / _.05_	**.01** [.02] / _-.11_
ADHD - Hyperactivity	**.11** [.16] / _-.03_	**.11** [.14] / _.11_	**.07** [.10] / _-.12_
Antisocial Interpersonal Functioning	**-.03** [.03] / _-.15_	**-.00** [.02] / _-.03_	**-.03** [.02] / _-.06_
Victim Interpersonal Functioning	**.00** [.03] / _.02_	**-.00** [.01] / _-.02_	**.01** [.03] / _.01_
School Grades	**.03** [.09] / _.11_	**-.08** [-.06] / _-.04_	**.14** [.20*] / _.15_
School Discipline	**-.03** [-.05] / _-.03_	**.04** [.00] / _.11_	**-.08** [-.08] / _-.01_

Note. Bolded values are concurrent associations from the present study. Values in brackets are predictive associations from the present study. Values in italics after the slash are concurrent coefficients from Snyder et al. (2015) for comparison. EATQ-R = Early Adolescent Temperament Questionnaire-Revised; EC = Effortful Control; NE = Negative Emotionality; PE = Positive Emotionality; ADHD = Attention deficit/hyperactivity disorder. Int. Funct. = Interpersonal Functioning. $^*p < .0005$

Table S1.9

Correlations between EC Factor Scores from the Various Models

	Common EC (BFM)	Activation-specific (BFM)	Activation Control (Traditional)	Attention (Traditional)	Inhibitory Control (Traditional)
Activation Control (CFM)	.97*	.15*	.86*	.75*	.59*
Attention (CFM)	.98*	.06	.73*	.79*	.71*
Inhibitory Control (CFM)	.96*	.01	.65*	.82*	.73*
Activation Control (Traditional)	.78*	.57*	-	-	-
Attention (Traditional)	.80*	.08	.51*	-	-
Inhibitory Control (Traditional)	.68*	.08	.42*	.52*	-

Note. BFM = factor scores from retained bifactor model for Snyder et al. (2015) fit to the present data. CFM = factor scores from correlated factors model. Traditional = manifest scores from the traditional scoring method presented in Snyder et al. (2015) fit to the present data. *$p < .0003$

Table S1.10

Correlations between NE Factor Scores from the Various Models

	Common NE (BFM)	Aggression-specific (BFM)	Depressed Mood-specific (BFM)	Fear-specific (BFM)	Frustration-specific (BFM)	Shyness-specific (BFM)	5.	6.	7.	8.	9.
1. Fear (CFM)	.81*	-.30*	.13	.65*	.01	.34*	.93*	.52*	.56*	.02	.61*
2. Frustration (CFM)	.89*	.29*	.05	-.01	.66*	-.06	.44*	.98*	.23*	.58*	.57*
3. Shyness (CFM)	.53*	-.21*	.14	.29*	-.12	.90*	.49*	.26*	.96*	-.00	.40*
4. Aggression (CFM)	.55*	.89*	.13	-.32*	.40*	-.21*	.02	.59*	-.00	.99*	.38*
5. Depressed Mood (CFM)	.95*	.10	.44*	.20*	.13	.16*	.63*	.68*	.46*	.43*	.90*
6. Fear (Traditional)	.66*	-.24*	-.02	.83*	-.10	.19*	-	-	-	-	-
7. Frustration (Traditional)	.82*	.21*	-.04	-.07	.76*	-.08	.36*	-	-	-	-
7. Shyness (Traditional)	.42*	-.17*	.12	.21*	-.13	.91*	.37*	.19*	-	-	-
8. Aggression (Traditional)	.47*	.92*	.10	-.28*	.30*	-.19*	.01	.48*	-.01	-	-
9. Depressed Mood (Traditional)	.76*	.05	.69*	.04	-.01	.05	.41*	.47*	.31*	.32*	-

Note. BFM = factor scores from retained bifactor model for Snyder et al. (2015) fit to the present data. CFM = factor scores from correlated factors model. Traditional = manifest scores from the traditional scoring method presented in Snyder et al. (2015) fit to the present data. *$p < .0003$

Table S1.11

Correlations between PE Factor Scores from the Various Models

	Surgency (BFM)	Affiliation (BFM)	Surgency (Traditional)	Affiliation (Traditional)
Surgency (CFM)	1.00*	.28*	.74*	.19*
Affiliation (CFM)	.25*	.98*	.21*	.99*
Surgency (Traditional)	.72*	.22*	-	-
Affiliation (Traditional)	.18*	.98*	.15*	-

Note. BFM = factor scores from retained bifactor model for Snyder et al. (2015) fit to the present data. CFM = factor scores from correlated factors model. Traditional = manifest scores from the traditional scoring method presented in Snyder et al. (2015) fit to the present data. *$p < .0003$

Table S1.12

Negative Emotionality Model Results Excluding Depressed Mood and Aggression

Model	χ^2 (*df*)	χ^2/df	CFI	RMSEA
Snyder et al. Bifactor Model	197.39(98)	2.01	.95	.042
Correlated Factors Model	366.66(116)	3.16	.88	.061

Note. CFI = confirmatory fit index; RMSEA = root mean square error of approximation. * $p < .001$.

Table S1.13

Correlations between Snyder et al. Bifactor Negative Emotionality excluding Depressed Mood and Aggression and Adolescent Functioning Measures

	Common NE	Fear-specific	Frustration-specific	Shyness-specific	Fear	Frustration	Shyness
Depression	**.20*** [.13]	**-.01** [-.04]	**.28*** [.16*]	**.01** [.06]	**.19*** [.11]	**.32*** [.20*]	**.11** [.11]
Anxiety	**.42*** [.21*]	**.11** [.01]	**.20*** [.10]	**.09** [.00]	**.40*** [.18*]	**.35*** [.18*]	**.29*** [.10]
ADHD - Total	**.11** [.14]	**-.03** [.03]	**.33*** [.23*]	**-.04** [-.02]	**.08** [.11]	**.31*** [.26*]	**.02** [.06]
ADHD - Inattention	**.14** [.11]	**-.02** [.01]	**.29*** [.19*]	**.00** [.02]	**.11** [.08]	**.30*** [.21*]	**.06** [.07]
ADHD - Hyperactivity	**.05** [.12]	**-.02** [.05]	**.27*** [.20*]	**-.09** [-.08]	**.04** [.10]	**.24*** [.21*]	**-.04** [.00]
Antisocial Interpersonal Funct.	**.05** [.00]	**-.02** [-.04]	**.24*** [.16*]	**-.06** [-.06]	**.06** [.00]	**.21*** [.12]	**-.01** [-.04]
Victim Interpersonal Funct.	**.21*** [.13]	**.00** [-.02]	**.13** [.08]	**-.05** [-.00]	**-.19*** [.11]	**.20*** [.12]	**.07** [.06]
School Grades	**-.00** [.10]	**-.08** [-.02]	**-.05** [.01]	**-.03** [.02]	**-.03** [.09]	**-.04** [.06]	**.01** [.07]
School Discipline	**-.08** [-.03]	**-.01** [.04]	**.13** [.10]	**-.04** [-.02]	**-.08** [-.01]	**.06** [.06]	**-.08** [-.03]

Note. Bolded values are concurrent associations from the present study. Values in brackets are predictive associations from the present study. EATQ-R = Early Adolescent Temperament Questionnaire-Revised; ADHD = Attention deficit/hyperactivity disorder. Funct. = Functioning.

**p* < .0003.

Table S1.14

Bifactor Models: Correlations and 95% Confidence Intervals for Temperament Factors and Adolescent Functioning Measures

Adolescent Functioning	EC	Activation-specific	Common NE	Aggression-specific	Depressed Mood-specific	Fear-specific	Frustration-specific	Surgency	Affiliation
Depression	**-.33*[-.40, -.25]** *-.16*[-.24, -.08]*	**-.16*[-.24, -.08]**	**.38*[.30, .44]** *.26*[.18, .33]*		**.27*[.20, .35]** *.20*[.12, .28]*		**.16*[.07, .23]**		
Anxiety			**.41*[.34, .48]** *.23*[.15, .31]*			**.18*[.10, .26]**			
ADHD	**-.39*[-.46, -.32]** *-.20*[-.29, -.10]*	**-.20*[-.28, 12.]**	**.30*[.22, .37]** *.28*[.19, .37]*	**.25*[.17, .32]** *.19*[.09, .28]*	**.19*[.11, .27]**		**.23*[.15, .31]**	*.18*[.09, .28]*	
ADHD – Inattention	**-.42*[-.49, -.35]** *-.24*[-.33, -.15]*	**-.18*[-.26, -.10]**	**.29*[.21, .36]** *.23*[.14, .32]*	**.22*[.14, .29]** *.18*[.08, .28]*	**.18*[.10, .26]**		**.20*[.12, .28]**		
ADHD – Hyperactivity	**-.26*[-.34, -.19]**	**-.17*[-.25, -.09]**	**.21*[.13, .29]** *.25*[.15, .34]*	**.22*[.14, .30]**			**.19*[.11, .27]**		

Antisocial Interpersonal Funct.	**-.22*[-.30, -.14]** *-.18*[-.26, -.10]*		**.21*[.13, .29]**	**.32*[.24, .39]** *.32*[.24, .39]*					
Victim Interpersonal Funct.			**.25*[.17, .32]** *.16*[.08, .24]*		**.20*[.12, .27]**				
School Grades	**.31*[.24, .39]** *.23*[.14, .31]*			**-.15*[-.23, -.07]** *-.17*[-.25, -.09]*					*.19 [.11, .27]*
School Behavior	**-.18*[-.26, -.10]** *-.22*[-.30, -.14]*			**.19*[.11, .26]** *.17*[.08, .25]*					

Note. Only significant correlations are depicted in this table. *$p < .0003$

Table S1.15

Correlated Factors Model: Correlations and 95% Confidence Intervals for Temperament Factors and Adolescent Functioning Measures

Adolescent Functioning	Act.	Att.	Inh.	Fear	Fru.	Shy	Agg,	Dep.	Sur.	Aff.
Depression	**-.30*[-.37, -.22]** *-.16*[-.24, -.08]]*	**-.29*[-.36, -.21]**	**-.28*[-.35, -.20]**	**.22*[.14, .30]**	**.36*[.28, .43]** *.23*[.15, .31]*		**.32*[.25, .40]** *.20*[.12, .28]*	**.41*[.34, .48]** *.29*[.22, .37]*		
Anxiety				**.41*[.34, .47]** *.20*[.11, .28]*	**.36*[.29, .43]** *.18*[.10, .26]*	**.30*[.22, .37]**	**.17*[.09, .25]**	**.41*[.34, .47]** *.22*[.14, .30]*		
ADHD	**-.38*[-.45, -.31]** *-.19*[-.29, -.10]*	**-.36*[-.43, -.29]**	**-.34*[-.42, -.27]**		**.35*[.28, .42]** *.29[.20, .38]*		**.40*[.33, .47]** *.31*[.22, .40]*	**.30*[.23, .37]** *.27*[.17, .36]*	*.19*[.09, .28]*	
ADHD – Inattention	**-.40*[-.47, -.33]** *-.23*[-.33, -.14]*	**-.39*[-.46, -.32]** *-.20*[-.30, -.11]*	**-.38*[-.44, -.30]** *-.19*[-.28, -.09]*		**.33*[.25, .40]** *.25*[.15, .34]*		**.37*[.29, .43]** *.28*[.18, .37]*	**.30*[.22, .37]** *.22*[.13, .31]*		
ADHD – Hyperactivity	**-.26*[-.33, -.18]**	**-.24*[-.31, -.16]**	**-.23*[-.30, -.15]**		**.27*[.19, .34]** *.24*[.15, .34]*		**.32*[.25, .39]** *.26*[.17, .35]*	**.21*[.13, .29]** *.22*[.12, .31]*		

Antisocial Interpersonal Funct.	**-.21*[-.28, -.13]** *-.16*[-.24, -.08]*	**-.20*[-.28, -.12]** *-.16*[-.24, -.08]*	**-.20*[-.27, -.12]** *-.16*[-.24, -.08]*		**.26*[.18, .33]** *.18*[.10, .26]*		**.37*[.30, .44]** *.33*[.26, .41]*	**.22*[.14, .29]** *.16*[.07, .24]*		
Victim Interpersonal Funct.				**.20*[.12, .28]**	**.23*[.15, .30]**		**.17*[.09, .25]**	**.27*[.19, .35]** *.18*[.10, .26]*		
School Grades	**.30*[.22, .37]** *.23*[.15, .31]*	**.30*[.22, .37]** *.23*[.15, .31]*	**.29*[.21, .37]** *.23*[.15, .31]*				**-.16*[-.24, -.08]**			*.20*[.11, .28]*
School Behavior	**-.18*[-.26, -.10]** *-.22*[-.29, -.13]*	**-.18*[-.25, -.09]** *-.21*[-.29, -.13]*	**-.17*[-.25, -.09]** *-.20*[-.28, -.12]*				**.17*[.09, .25]** *.17*[.09, .25]*			

Note. Act. = Activation Control. Attention = Attention Control. Inh. = Inhibitory Control. Fru. = Frustration. Shy = Shyness. Agg. = Aggression. Dep. = Depressed Mood. Sur. = Surgency. Aff.= Affiliation. Only significant correlations are depicted in this table. *$p < .0003$

Figure S1.1

Bifactor Model for Effortful Control

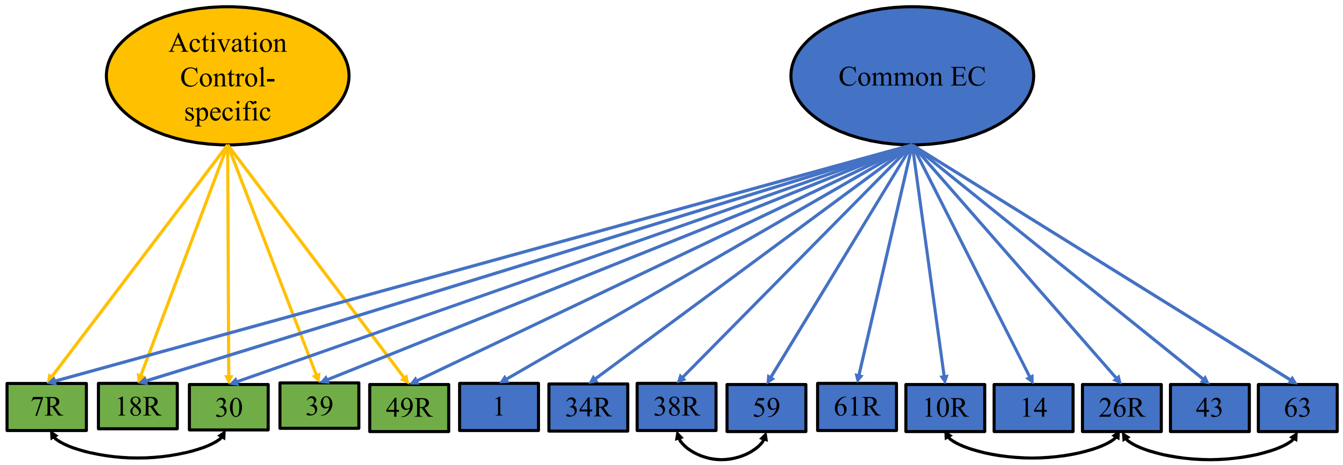


Figure S1.2

Bifactor Model for Negative Emotionality

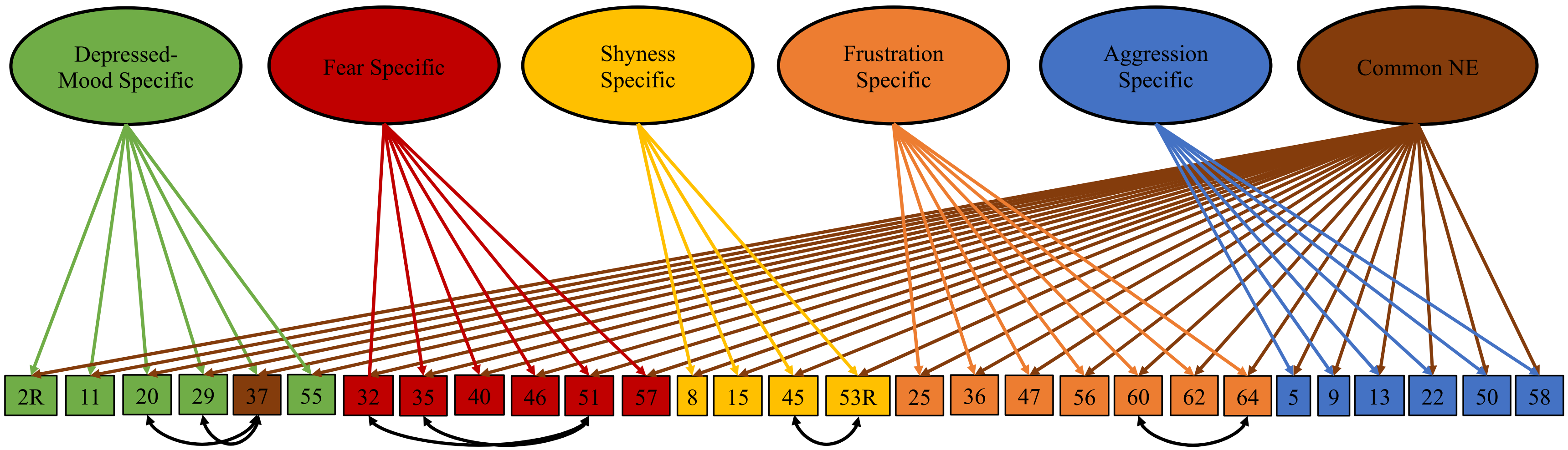


Figure S1.3

Bifactor Model for Positive Emotionality

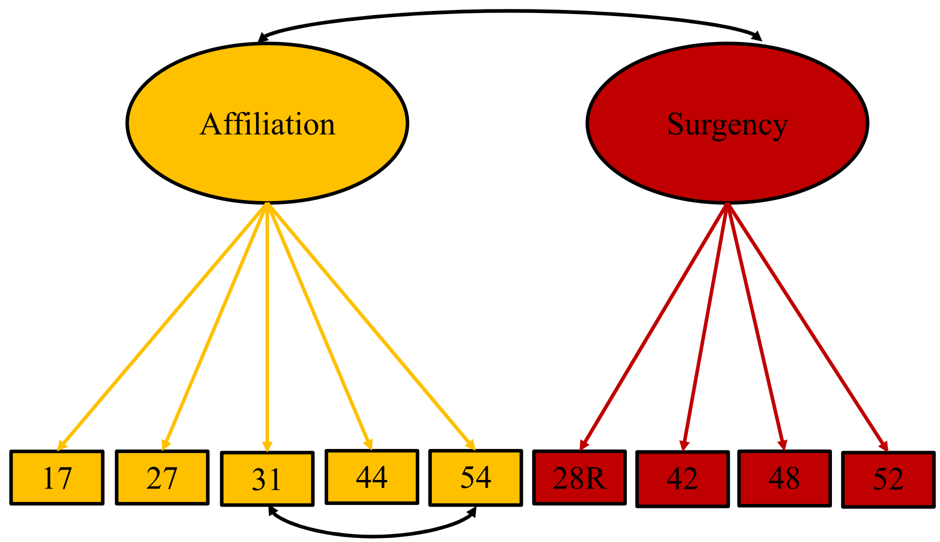


Figure S1.4

Correlated Factors Model for Effortful Control

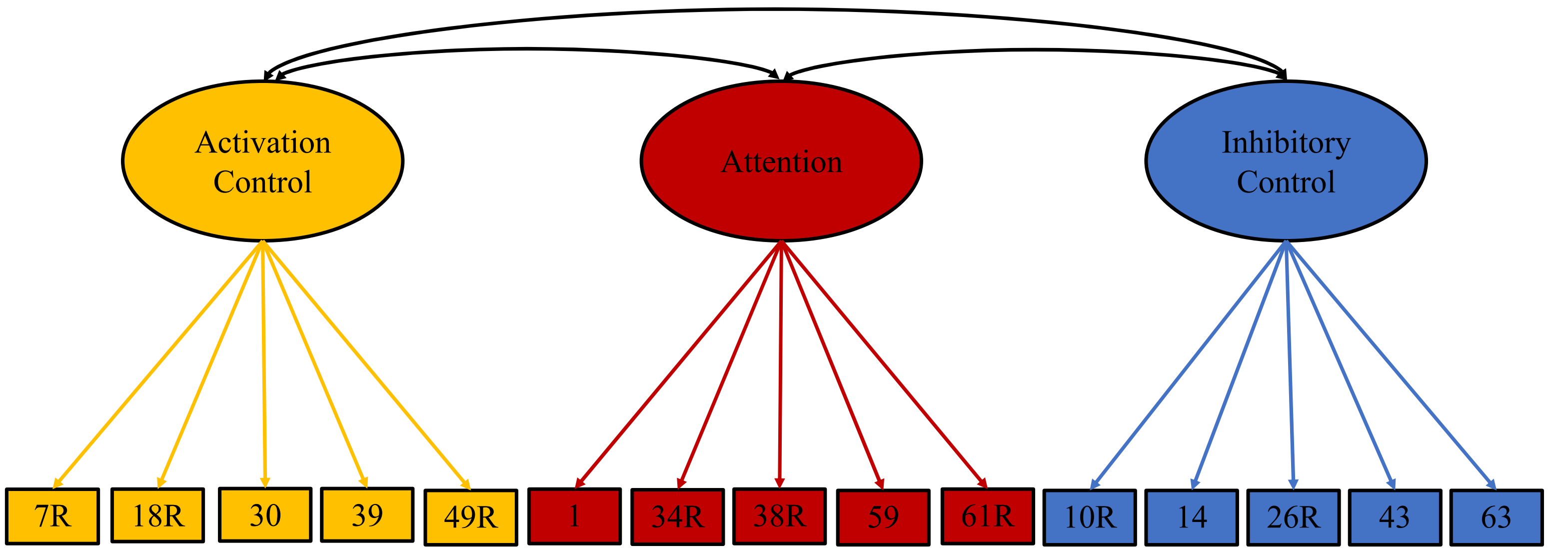


Figure S1.5

Correlated Factors Model for Negative Emotionality

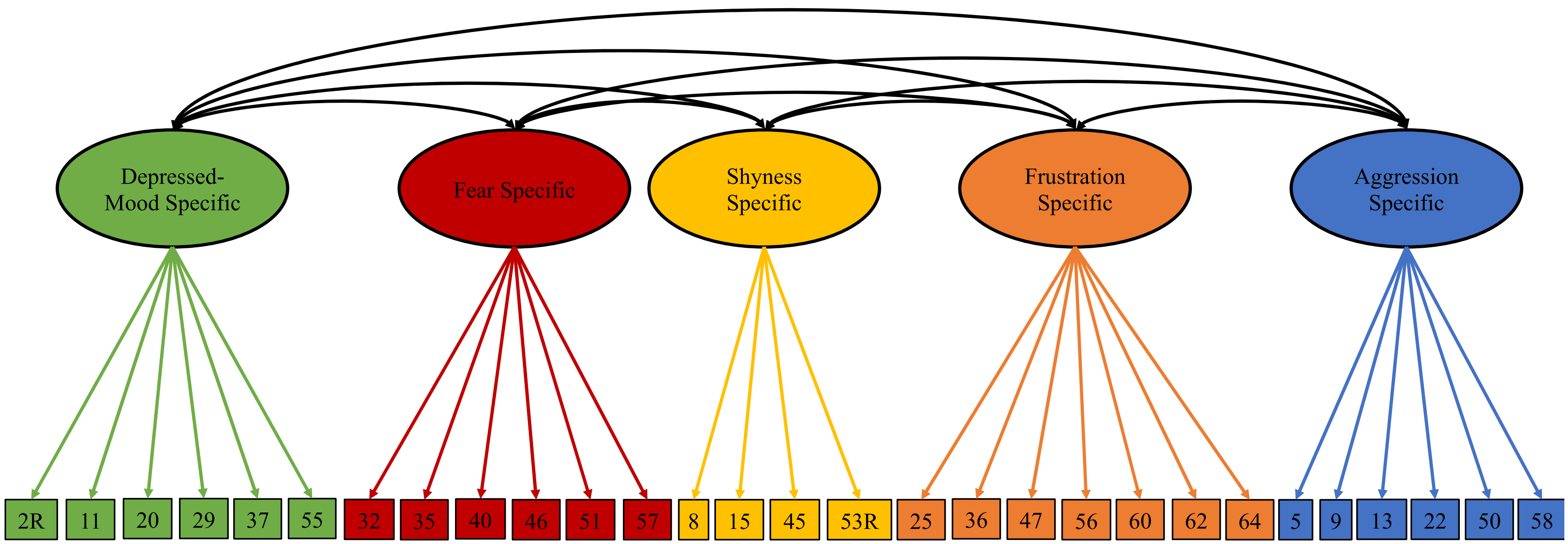
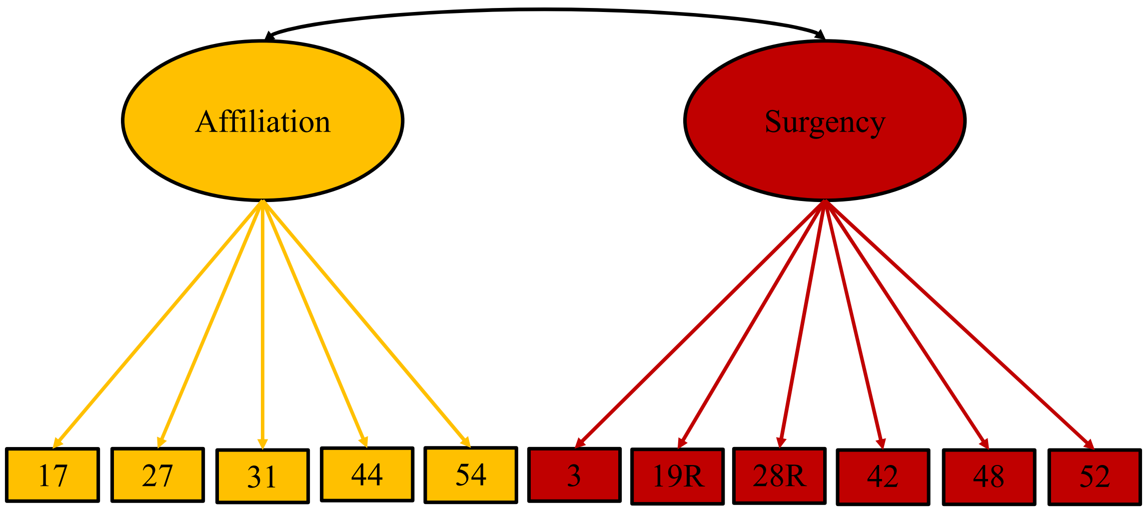


Figure S1.6

Correlated Factors Model for Positive Emotionality



Chapter 2

The Role of Temperament in the Onset of Suicidal Ideation and Behaviors across Adolescence: Findings from a 10-year Longitudinal Study of Mexican-Origin Youth

The content of this chapter has been previously published in the *Journal of Personality and Social Psychology*. I want to extend my thanks to the American Psychological Association for approving this copyrighted material to be used in my dissertation. Below is the citation for the corresponding published article.

Cite: Lawson, K. M., Kellerman, J. K., Kleiman, E. M., Bleidorn, W., Hopwood, C. J. & Robins, R. W. (2022). The role of temperament in the onset of suicidal ideation and behaviors across adolescence: Findings from a 10-year longitudinal study of Mexican-origin youth. *Journal of Personality and Social Psychology, 122*(1), 171-186.

Abstract

Suicide among young people is an increasingly prevalent and devastating public health crisis around the world. To reduce the rate of suicide, it is important to identify factors that can help us better predict suicidal ideation and behaviors. Adolescent temperament (Effortful Control, Negative Emotionality, Positive Emotionality) may be a source of risk and resilience for the onset of suicidal ideation, plans, and attempts. The present study uses longitudinal data from a large, community sample of Mexican-origin youth (N=674), assessed annually from age 12 to 21, to examine how temperament is associated with the onset of suicidal ideation and behaviors during adolescence and young adulthood. Results indicate that higher levels of Effortful Control (Activation Control, Inhibitory Control, Attention) are associated with *decreased* probability of experiencing the onset of suicidal ideation, plans, and attempts, whereas higher levels of Negative Emotionality (particularly Aggression, Frustration, and Depressed Mood) are associated with *increased* probability of experiencing the onset of suicidal ideation and behaviors. Positive Emotionality (Surgency, Affiliation) was not associated with the onset of suicidal ideation and behaviors. Supplemental analyses showed conceptually similar findings for the Big Five, with Conscientiousness associated with decreased risk, Neuroticism associated with increased risk, and the other three dimensions showing largely null results. The findings did not vary significantly for boys and girls or for youth born in the U.S. versus Mexico. Overall, these findings suggest that adolescent temperament serves as both a protective factor (via Effortful Control/Conscientiousness) and a risk factor (via Negative Emotionality/Neuroticism) for suicidal ideation and behaviors in Mexican-origin youth.

Introduction

Suicide is the second leading cause of death for Americans aged 10 to 24 and the prevalence of adolescent deaths by suicide has steadily increased during the past decade (Curtin & Heron, 2019). Consequently, it is important to understand the risk and resilience factors associated with the onset of suicidal ideation and behaviors across adolescence (Franklin et al., 2017). In particular, risk factors can identify youth who need additional support, whereas resilience factors may be leveraged to delay the onset, or minimize the severity, of suicidal behavior. Previous research on risk and resilience factors for suicide has emphasized individual factors such as depression and other mood disorders, and relational factors such as bullying and relationship problems (Shain, 2016). However, almost no longitudinal work has examined how adolescent temperament, or relatively enduring individual differences in reactivity and self-regulation that are present from an early age, might influence the onset of suicide risk. Further, few studies have examined risk and protective factors for adolescent suicide in Latinx youth, despite the fact that Latinx youth are at especially high risk for suicidal ideation and behaviors (Price & Khubchandani, 2017; Silva & Van Orden, 2018). To address these gaps, the present study examined whether three domains of adolescent temperament (Effortful Control, Negative Emotionality, Positive Emotionality) and their facets serve as risk and/or resilience factors in the onset of suicidal ideation, plans, and attempts across adolescence. To examine this question, we used data from a large, community sample of Mexican-origin youth (N = 674) assessed nine times from age 12 to 21.

Adolescent Suicide

Suicide among young people is an increasingly prevalent and devastating public health crisis around the world. Suicide is now estimated to be the second leading cause of death among adolescents and young adults and the past two decades have seen steady increases in global adolescent suicide rates, with rates in the United States increasing 24% between 1999 and 2014 (CDC, 2017; WHO, 2014). Suicide attempts and deaths are relatively rare during childhood, but dramatically increase during the transition to adolescence before decreasing and stabilizing during the transition from adolescence to adulthood (Nock et al., 2008; Nock et al., 2013). Notably, adolescents experience elevated risk of transitioning from suicidal ideation to suicidal behavior, with risk for suicidal behavior peaking at age 16 and remaining elevated into the early 20s (Nock et al., 2008). Young adults who meet risk criteria in adolescence are also more likely to engage in suicidal behavior than those who did not meet risk criteria in adolescence. Therefore, for many individuals, adolescence is the most consequential developmental period for suicide risk, and researchers should work to understand risk and resilience trajectories throughout this period.

Suicide risk encompasses multiple components, including suicidal ideation, plans, and attempts. Data from a national survey of American adolescents indicate a prevalence rate of 12.1% for suicidal ideation across youth ages 13-18 (Nock et al., 2013). Suicide plans and suicide attempts have estimated prevalence rates of 4.0% and 4.1%, respectively, among the same population. Approximately one-third of adolescents who experience suicidal ideation go on to develop a suicide plan and make a suicide attempt during adolescence, although the factors driving this transition from thought to behavior are not yet understood (Glenn et al., 2017; Nock et al., 2013). Well-studied suicide risk factors (e.g., major depressive disorder) are more predictive of suicidal ideation than they are of the transition from ideation to plans and attempts, limiting the ability to predict which youth who ideate will experience an escalation to suicidal behavior (Miranda, Ortin, Scott, & Shaffer, 2014; Nock et al., 2008). Therefore, understanding the factors underlying suicidal ideation, plans, *and* attempts is imperative to identify individuals at higher risk and to prevent suicide deaths among adolescents.

Latinx youth, and Mexican-origin youth in particular, are an understudied demographic group that are at elevated suicide risk. National statistics indicate that American adolescents who identify as Hispanic or Latinx report higher rates of suicidal ideation, plans, and attempts during adolescence than members of other ethnic groups (Carino & Roberts, 2001; Kann et al., 2014; Silva & Van Orden, 2018; Spirito & Esposito-Smythers, 2006). Further, suicide attempts among Latinx youth are generally more severe, as they are more likely to require medical attention. But, despite the increased severity of their suicidal behaviors, Latinx youth are generally less likely to receive needed mental health treatment (Goldston et al., 2008; Spirito & Esposito-Smythers, 2006). These higher rates may be due, in part, to various aspects of acculturative stress, including exposure to discrimination, greater family conflict, lower ethnic identity, and a lower sense of belonging (Fortuna et al., 2016; Gomez, Miranda, & Polanco, 2011; Silva & Van Orden, 2018). Disparities in the prevalence of suicidal ideation and behaviors are primarily driven by Latina adolescents, who consistently report the highest rates of suicidal ideation and attempts of any gender and ethnic group (Kann et al., 2014; Price & Khubchandani, 2017). Among American high school students, Latina/Hispanic girls reported higher rates of suicidal ideation (21.1%) compared to White female students (17.8%), as well as a suicide attempt rate (14.0%) nearly double the rate reported by their White female classmates (7.7%; Eaton, Koti, Brener, Crosby, et al., 2011). Finally, studies examining differences among Latinx youth have found that Mexican-origin youth are at higher risk than their other Latinx counterparts (Duarté-Vélez & Bernal, 2007). Altogether, these findings highlight that Mexican-origin youth are a particularly important demographic in which to study risk and protective factors for suicide. However, the majority of suicide research conducted with adolescents has used data from predominantly White samples.

Personality and Suicide

Past research has shown that personality disorders, in particular, borderline personality disorder (Pompili, Girardi, Ruberto, & Tatarelli, 2005; Yen et al., 2003) and antisocial personality disorder (Verona et al., 2001), are associated with increased risk for suicidal ideation and behaviors. Studies of normal-range personality traits have shown that suicidal ideation, plans, and attempts are concurrently associated with *lower* levels of Extraversion (Baertschi, Costanza, Canuto, & Weber, 2018; Brezo et al., 2006a; Cramer et al., 2012; Fino et al., 2014; Frances et al., 1986; Kerby, 2003; Tucker et al., 2014), Agreeableness (Baertschi et al., 2018; Brezo et al., 2006b; Kerby, 2003), and Conscientiousness (Brezo et al., 2006b; Kerby, 2003; Velting 1999), as well as *higher* levels of Neuroticism (Baertschi et al., 2018; Brezo et al., 2006a; Chioqueta & Stiles, 2005; Cramer et al., 2012; Fino et al., 2014; Kerby, 2003; Lolas, Gomez, & Suarez, 1991; Nordström, Schalling, & Asberg, 1995; Tanji et al., 2015; Tucker et al., 2014). However, the vast majority of these studies use cross-sectional data from relatively small samples of college students, hospitalized patients, or individuals who are incarcerated, limiting generalizability to members of the broader community. Additionally, the majority of participants from these studies are White, limiting generalizability to ethnic minority groups.

Thus, despite a plethora of concurrent associations, we have little insight into the longitudinal associations between personality traits and suicidal ideation and behaviors (Franklin et al., 2017). This is a common problem in the field of personality and psychopathology and makes it difficult to tease apart the directionality of associations between temperament and suicidal ideation and behaviors. Fortunately, longitudinal data provide an opportunity to examine evidence for various developmental models of temperament and psychopathology including the vulnerability, scar, pathoplasty, and spectrum models (Durbin & Hicks, 2014; Tackett, 2006). The vulnerability model suggests that underlying levels of temperament can increase risk for experiencing the onset of suicidal ideation and behaviors, whereas the scar model suggests that experiencing suicidal ideation and behaviors can lead to changes in temperament. Further, the pathoplasty model suggests that temperament could influence the manifestation of suicidal ideation and behaviors in terms of course, severity, presentation, or prognosis. Finally, the spectrum model suggests that temperament traits and suicidal ideation/behaviors lie on the same continuum. In the present study, we focus on the *vulnerability* model, which is especially relevant because temperament differences are present prior to the onset of suicidal ideation and behaviors and may help to identify youth who need additional support. Furthermore, only a handful of past studies have focused on adolescence, a critical, high-risk period for the onset of suicidal ideation and behaviors. Next, we review the limited cross-sectional research on adolescent temperament and suicide risk, and then turn to the small number of longitudinal studies of personality/temperament and suicidal ideation and behaviors.

Adolescent temperament. Temperament refers to individual differences in reactivity and self-regulation that are present from an early age and relatively enduring (Rothbart, Ahadi, & Evans, 2000); Rothbart, 2007; Rothbart, 2011). Researchers describing individual differences in children and adolescents often label traits as either "temperament" or "personality", but there is no clear conceptual or empirical distinction between the two (Clark & Watson, 2008; Shiner & DeYoung, 2013). In this paper, we focus on the role of individual differences through the lens of temperament, though we also include complementary analyses and interpretations using a Big Five personality framework.

Research on youth temperament is often guided by Rothbart's highly influential temperament model (Rothbart et al., 2000), which posits three key constructs: Effortful Control (EC), Negative Emotionality (NE), and Positive Emotionality (PE). According to Rothbart's model, the EC domain involves one's capacity to plan and suppress inappropriate impulses (Inhibitory Control), perform an action or pursue goals when there are competing desires (Activation Control), and focus and shift attention when needed (Attention). The NE domain involves unpleasant affect derived from anticipating distress (Fear), negative affect related to ongoing tasks being interrupted (Frustration), and behavioral inhibition to social interaction (Shyness). In some cases, the NE domain is expanded to include hostile reactivity to negative stimuli including person- and object-directed violence (Aggression) and unpleasant affect, lowered mood, and lack of enjoyment in activities (Depressed Mood). The PE domain involves pleasure derived from high intensity or novel activities (High Intensity Pleasure/Surgency; hereafter referred to as Surgency) and a desire for close, warm interpersonal connections (Affiliation). Whereas EC and NE represent broad cohesive domains, mounting theoretical and empirical evidence suggests that the PE facets do not form a cohesive domain (Lawson et al., in press; Snyder et al., 2015). Because of this, we consider both domain and facet-level scores for EC and NE but we only consider facet-level scores (i.e., separate Surgency and Affiliation) for PE.

Rothbart's temperament domains, especially low EC and high NE, are associated with numerous mental health problems, including anxiety, depression, nonsuicidal self-injury, conduct disorder, and ADHD (Atherton et al., 2020; Baetens et al., 2011; Capaldi & Rothbart, 1992; De Pauw & Mervielde, 2010; Muris & Ollendick, 2005; Snyder et al., 2015). Consequently, these temperament domains may also be related to suicidal ideation and behaviors. It is well-documented, both conceptually and empirically, that EC is comparable to Conscientiousness, NE is analogous to Neuroticism, and PE is related to Extraversion (Shiner & DeYoung, 2013; Rothbart, 2007; Rothbart & Ahadi, 1994; Rothbart et al., 2000). Given these relations and previous research on Big Five traits and suicidal ideation and behaviors, we might expect to find that higher levels of EC and PE serve as protective factors whereas NE serves as a risk factor.

Indeed, three systematic reviews found that high levels of impulsivity (associated with low levels of the Inhibitory Control facet of EC), high levels of aggression and anxiety (associated with the Aggression and Fear facets of NE, respectively), and high levels of novelty-seeking (associated with the Surgency facet of PE) were consistently associated with increased suicidal ideation, plans, attempts, and deaths (Brezo, Paris, & Turecki, 2006a; Frances, Fyer, & Clarkin, 1986; Gvion & Apter, 2011). Two studies of Japanese (Tanabe et al., 2016) and Lebanese adults (Karam et al., 2015) found that suicide attempts were associated with anxious and irritable temperament scores on the TEMPS-A (Akiskal et al., 2005). Together, these studies highlight that the temperamental domains of EC and NE, and possibly PE, are related to suicidal ideation and behaviors. However, all of these studies use cross-sectional designs, which provide little insight into *developmental* relations between temperament and suicidal ideation and behaviors. Next, we describe the few extant longitudinal studies of temperament and suicidal ideation and behaviors.

Longitudinal associations. Studies examining longitudinal relations between temperament and suicidal ideations and behaviors suggest that certain traits might predispose youth to experience later suicidal ideation and behaviors. One longitudinal study that followed American adolescents from age 10 to 25 found that individuals who attempted suicide scored six times higher in impulsivity than those who did not attempt at age 17, and showed smaller subsequent declines in impulsivity as they transitioned into young adulthood (Kasen, Cohen, & Chen, 2011). Another longitudinal study found that EC might mediate the influence of coping responses on suicide attempts, given that youth with higher levels of EC were able to cope more effectively and showed decreased likelihood of attempting suicide two years later (Piquet & Wagner, 2003). Further, a more recent study found that impulsivity was positively associated with risk of suicide attempts during adolescence and declines in impulsivity map on to declines in suicide risk during the transition from adolescence to adulthood (Thompson & Swartout, 2018). With respect to NE, in a large, long-term longitudinal study, adolescent self-reported worry and irritability predicted making a suicide plan and attempting suicide 30 years later in middle adulthood, even after accounting for adult psychopathology and Neuroticism (Pickles et al., 2010). Finally, a large, longitudinal study of adolescents in New Zealand found that Neuroticism and novelty-seeking (related to PE) assessed at age 14 and 16, respectively, were associated with the onset of suicidal ideation and attempts from age 14 to 21 (Fergusson, Beautrais, & Horwood, 2003). Thus, there is evidence that certain aspects of EC, NE, and PE might predispose youth to be more vulnerable to experience later suicidal ideation and behavior. However, these associations are not likely to manifest in the same way for all adolescents and, in particular, they may be impacted by demographic differences in gender and nativity (i.e., country of birth).

Gender and nativity. There are well-documented gender differences in the prevalence of suicidal ideation and behaviors across adolescence. In particular, adolescent boys and girls demonstrate differences in onset and trajectory of suicide risk, with girls experiencing higher overall rates of ideation compared to boys, in addition to earlier onset, peak, and decline of suicidal ideation (Kann et al., 2014; Nock et al., 2008; Reuter, Holm, McGeorge, & Conger, 2008). As previously outlined, this gender difference is exacerbated in Latinx youth, with Latinas being at especially high risk for suicidal ideation and behaviors (Price & Khubchandani, 2017). Data from the Youth Risk Behavior Surveillance study in 2013 showed that Latina high school students had the highest prevalence rates of any demographic group for suicidal ideation (26.0%), plans (20.1%), and attempts (15.6%) (Kann et al., 2014).

In addition to mean-level differences in the prevalence of suicidal ideation and behaviors, past research has documented gender differences in the association between individual-level risk and protective factors and suicidal ideation and behaviors (Edwards & Holden, 2001). This finding extends to gender differences in the relation between personality and suicide. A cross-sectional study of undergraduate students found that women higher in angry hostility and depression (i.e., NE) were more prone to suicidal ideation, whereas men with poor self-discipline (i.e., low EC) were more prone to suicidal ideation (Velting, 1999). Another cross-sectional study using a large sample of German adults found that men low in Extraversion and Conscientiousness were at increased risk for suicide, whereas women high in Neuroticism and Openness were at increased risk (Blüml et al., 2013). Together, these studies provide limited evidence that the relation between temperament and suicidal ideation and behaviors may differ for boys and girls, though much of this work has been done with adults.

In addition to gender, nativity status (i.e., whether a child was born in Mexico vs. the U.S.) may moderate the relation between temperament and suicide ideation and behaviors in Mexican-origin youth. Duarté-Vélez and Bernal (2007) highlight evidence of "within-group diversity (such as national origin, generational status in the United States)" in suicide behaviors, demonstrating the importance of studying individual differences among Mexican-origin youth

(p. 439). Indeed, there is evidence that Latinx, and particularly Mexican-origin, immigrants born in the United States may be at higher risk for suicidal behavior than Mexican-origin adolescents who were born abroad and then immigrated during childhood (Carino & Roberts, 2001; Peña et al., 2008; Silva & Van Orden, 2018; Sorenson & Shen, 1996). These findings are consistent with the "immigrant paradox", which suggests that more recent immigrants have better outcomes than more established immigrants despite facing additional barriers to social integration (Garcia-Coll & Marks, 2012). Thus, given the evidence that nativity is associated with suicidal ideation and behaviors, it may also moderate the relation between temperament and the onset of suicidal ideation and behaviors in Mexican-origin youth. In particular, temperament may be more strongly associated with the onset of suicidal ideation and behaviors for youth born in the U.S. compared to youth born in Mexico.

The Present Study

The present study examines the role of temperament in the onset of suicidal ideation and behaviors from early adolescence (age 12) to young adulthood (age 21). This research is exploratory, although prior research suggests that suicide risk will be related to higher levels of NE and lower levels of EC. Relations between PE and suicide risk are much less clear. Our primary research question concerns how temperament (i.e., EC, NE, PE) is associated with the onset of suicidal ideation, plans, and attempts from early adolescence to young adulthood. In addition, we will test whether these associations vary by gender (boys vs. girls) and nativity (born in the U.S. vs. Mexico).

The present study extends past research in several ways. First, we use long-term, longitudinal data with nine waves of data spanning ten years from early adolescence to young adulthood. This builds on previous cross-sectional work and provides a more comprehensive, fine-grained depiction of the development of suicide risk than previous longitudinal studies, which typically only included a few waves of data. Given that the onset of suicidal ideation and behaviors most frequently occurs during adolescence (Nock et al., 2008), we are likely to capture onset for the majority of participants who will experience lifetime suicidal behaviors. Second, we examined suicidal ideation and behaviors among Mexican-origin youth, a population that is historically understudied in psychological research in general and research on suicide in particular. This knowledge gap is even more striking given that Mexican-origin youth are at particularly high risk for suicidal behavior (Price & Khubchandani, 2017). These first two strengths directly address past appeals that "more within-group studies are needed, and particularly with Latino/a older adolescents" (Duarté-Vélez & Bernal, 2007, p. 445). Third, we assess multiple types of suicidal thoughts and behaviors including ideation, plans, and attempts. Though suicidal ideation, plans, and attempts have similar correlates, assessing all three provides a more nuanced depiction of the onset of suicide risk and extends past research which focuses on ideation and attempts, but less on plans (Franklin et al., 2107). Fourth, we assess both risk and protective factors, which is unique as most studies of correlates of suicidal ideation and behaviors exclusively focus on risk factors (Franklin et al., 2017). Fifth, we use multi-method temperament data including both self-reports and mother reports of adolescent temperament, which reduces problems associated with exclusive reliance on self-reports, a concern in many studies of personality and psychopathology (Durbin & Hicks, 2014). Finally, given conceptual and empirical relations between temperament and personality (Shiner & DeYoung, 2013), we also examine the role of Big Five personality traits in the onset of suicidal ideation, plans, and attempts to complement the temperament analyses.

Methods

Participants

This study uses data from the California Families Project, a longitudinal study of Mexican-origin youth and their parents (*N*=674).[15] Children were drawn at random from rosters of students from the Sacramento and Woodland, CA school districts. The focal child had to be in the 5th grade, of Mexican origin, and living with his or her biological mother in order to participate in the study. Approximately 72.6% of the eligible families agreed to participate in the California Families Project, which was granted approval by the University of California, Davis Institutional Review Board (Protocol # 217484-21). The children (50% female) were assessed annually from 5th grade to three years post-high school. The present study uses data from Waves 3-11 (collected in 2006-18), when the children were in 7th grade (M_{age} = 12.81, *SD* = 0.49) to three years post-high school (M_{age} = 21.74, *SD* =0.73). Retention rates compared to the original sample are as follows: 86% (Wave 3), 88% (Wave 4), 90% (Wave 5), 88% (Wave 6), 89% (Wave 7), 89% (Wave 8), 87% (Wave 9), 87% (Wave 10), and 80% (Wave 11).

Participants were interviewed in their homes in Spanish or English, depending on their preference. Interviewers were all bilingual and most were of Mexican heritage. Sixty-three percent of mothers and 65% of fathers had less than a high school education (median = 9th grade for both mothers and fathers); median total household income was between $30,000 and $35,000 (overall range of income = < $5,000 to > $95,000). With regard to generational status, 83.6% of mothers and 89.4% of fathers were 1st generation, and 16.4% of mothers and 10.6% of fathers were either 2nd or 3rd generation. One hundred and twenty-four of the families were single-parent households (mothers only), and 549 of the families were two-parent households. In the present study, we used data for all available participants (i.e., no exclusions were applied), and we have reported all analyses conducted to address our research questions.

Measures

Suicidal ideation and behaviors. We used two measures to assess suicidal ideation and behaviors.[16] First, adolescents responded to three questions adapted from the Youth Risk Behavior Survey annually from age 12 to 19 and at age 21 (YRBS; Brener et al., 2004). In particular, youth were asked about their experiences in the past year with suicidal ideation (i.e., “Have you thought about committing suicide?”), planning (i.e., “Have you made a plan for committing suicide?”), and attempts (i.e., “Have you attempted suicide?”). To ensure discretion and given the sensitive nature of these questions, youth completed this scale without the help of the interviewers (i.e., youth reported their responses directly on a computer that was turned away from interviewers). Additionally, youth were asked about their own ideation, plans, and attempts only *after* being asked about whether their friends had ideated, planned, or attempted (e.g., “Have your friends thought about committing suicide?”). Responses were recorded as 1 = *never*, 2 = *once*, 3 = *twice*, and 4 = *3 or more times*.

Second, adolescents responded to two items about suicide from the depression module of the NIMH Diagnostic Interview Schedule for Children-IV annually from age 12 to 18 (*DISC-IV*). The *DISC-IV* is a comprehensive, psychiatric interview that assesses mental health problems for children and adolescents using DSM-IV criteria; it is the most widely-used mental health interview that has been tested in both clinical and community populations and validated in both English and Spanish (Costello, Edelbrock, & Costello, 1985; Schwab-Stone et al., 1996; translated into Spanish by Bravo, Woodbury-Farina, Canino, & Rubio-Stipec, 1993). The two relevant depression items asked about suicidal ideation (i.e., “Was there a time when you thought seriously about killing yourself?") and suicide attempts (i.e., "During the last year, have you tried to kill yourself?") in the past year. Responses to these items were recorded dichotomously (0 = *no*, 1 = *yes*) as the symptom being present or not in the past year.

To create the most comprehensive suicide variables across assessments, we created a binary (0 = *no*, 1 = *yes*) score for each participant at each year, separately for ideation, plans, and attempts. From ages 12 to 18, the binary ideation and attempt variables included two items – one from the YRBS and one from the DISC-IV – such that participants who endorsed at least one of these items received a *yes* score. At ages 19 and 21, the suicidal ideation and attempt variables include only YRBS data. At all ages, the binary plan variable included only the item from the YRBS because there were no corresponding DISC-IV data for suicide plans.[17] Table 2.2 reports prevalence rates at each age and across the entire study period from age 12 to 21, separately for ideation, plans, and attempts.

Temperament. Adolescent temperament was measured via self-reports and mother-reports when the adolescents were 12, 14, and 16 years old using a short form of the *Early Adolescent Temperament Questionnaire – Revised* (EATQ-R; Ellis & Rothbart, 2001). In particular, adolescents and their mothers completed the EATQ-R, which measures three domains of temperament – Effortful Control (EC), Negative Emotionality (NE), and Positive Emotionality (PE). Ratings were made on a 4-point scale ranging from 1 (*not at all true of you/your child*) to 4 (*very true of you/your child*). Given variability in the way the EATQ-R temperament domains are scored (e.g., Lawson et al., in press; Snyder et al., 2015), we present all findings separately for each individual facet scale, as well as for the broad temperament domains.

For each domain and facet, we computed latent variables using both self- and mother-reports of adolescent temperament. We used parcels as indicators because parcels produce more reliable latent variables than individual items (Little et al., 2002). In particular, for the EC and NE domains, we created four parcels by randomly assigning items from *all* facets onto *each* parcel. For all facets, we created three parcels by randomly assigning items into each parcel. Additionally, to examine whether type of informant (i.e., self or mother) impacts the findings, we ran additional analyses separately for self- and mother-ratings of temperament. For all temperament data (i.e. self-and mother ratings, adolescent self-ratings, and mother-ratings), we saved factor scores of the latent variables and used these as predictors in the survival models. Descriptive statistics for the temperament domains are shown in Table 2.1.

Effortful control. The EC scale (16 items) assesses the ability to anticipate and suppress inappropriate responses, as well as the ability to perform an action despite the inclination not to do so. This scale has three facets: Activation Control (5 items), Attention (6 items), and Inhibitory Control (5 items). Activation Control assesses the ability to perform an action or pursue goals when there are competing desires. Attention assesses the ability to focus and shift attention when needed. Inhibitory Control assesses the ability to plan and suppress inappropriate impulses. Sample EC items include “It is easy for [you/your child] to really concentrate on homework problems” and “[You/your child] puts off working on projects until right before they are due.”

Negative emotionality. The NE scale (17 items) assesses the propensity to experience negative emotions. This scale has three central facets: Fear (6 items), Frustration (7 items), and Shyness (4 items). Fear assesses unpleasant affect derived from anticipating distress. Frustration assesses negative affect related to ongoing tasks being interrupted. Shyness assesses behavioral avoidance of novelty and social challenges. The Aggression and Depressed Mood scales are also conceptually related to NE. In particular, Aggression (6 items) assesses hostile reactivity to negative stimuli including person- and object-directed violence and Depressed Mood (6 items) assesses unpleasant affect, lowered mood, and lack of enjoyment in activities. Sample NE items include, "[You/your child] feel scared when entering a darkened room at home" and "It frustrates [you/your child] if people interrupt when you're talking."

Scoring of the NE domain varies widely across studies, sometimes including items from the Fear, Frustration, and Shyness subscales and other times including items from these scales plus the Aggression and Depressed Mood subscales (e.g., Snyder et al., 2015). In order to examine multiple conceptualizations of NE, we ran all analyses with both a narrow NE score (i.e., Fear, Frustration, and Shyness subscales) and a broad NE score (i.e., Fear, Frustration, Shyness, Aggression, and Depressed Mood subscales).

Positive emotionality. As detailed above, the PE scale includes two largely separate facets – Surgency and Affiliation.[18] The Surgency subscale (14 items) is the core of PE and assesses the tendency to seek out rewarding or sociable experiences. This scale includes the 6 Surgency items from the short version of the EATQ-R as well as an additional 8 items from the full-length version of the EATQ-R, which were added to improve the reliability of the Surgency scale. A sample item is, "[You/your child] enjoy[s] going places where there are big crowds and lots of excitement". The Affiliation subscale (5 items), assesses desire for warmth and closeness with others. A sample item is, "It is important to [you/your child] to have close relationships with other people."

Gender. Adolescents reported on their gender (1=girl, 2=boy).

Nativity. Participants were categorized as 1st generation if their birth country was Mexico (29%); as 2nd generation if their birth country was the U.S., and only one of their parents was reported as being born in the U.S. (62%); and as 3rd generation if their birth country and both parents were born in the U.S. (9%). Because of the low percentage of 3rd generation youth, we created a dichotomous nativity status variable comparing 1st generation (born in Mexico; 29%) to 2nd+ generation (born in U.S.; 71%) youth in all analyses.

Table 2.1
Descriptive statistics for temperament domains.

	Age 12		Age 14		Age 16	
	M (SD)	*α / ω*	*M (SD)*	*α / ω*	*M (SD)*	*α / ω*
Effortful control	0 (0.92)	.82 / .84	-0.21 (0.97)	.84 / .86	-0.20 (0.88)	.82 / .84
Negative emotionality	0 (0.92)	.85 / .87	-0.30 (1.03)	.86 / .88	-0.47 (1.00)	.83 / .86
Negative emotionality (without aggression and depressed mood)	0 (0.91)	.75 / .80	-0.43 (0.91)	.71 / .77	-0.61 (0.91)	.76 / .81
Positive emotionality	0 (0.81)	.76 / .79	1.00 (0.94)	.77 / .81	1.07 (1.00)	.81 / .84

Note. M = mean, SD = standard deviation, α = alpha reliability, ω = omega reliability. Descriptive statistics for effortful control and negative emotionality reflect the results for the broader domains, whereas the statistics for positive emotionality (PE) reflect only the results from Surgency, the core of PE. Mean and SD are provided for the saved factor scores from the latent variables. Descriptive statistics for the observed variables are shown in Table S2.1.

Procedures for the Statistical Analyses

All data cleaning was conducted in R (R Core Team, 2019) via RStudio Version 1.2.1335. All analyses were conducted in Mplus Version 8 (Muthén & Muthén, 1998-2017) using robust maximum likelihood estimation (MLR) and full information maximum likelihood (FIML) to address missing data (Allison, 2003; Schafer & Graham, 2002).[19] Chi square analyses were conducted to examine significant gender and nativity differences in suicidal ideation, plans, and attempts.

Measurement Invariance. We examined evidence for longitudinal measurement invariance of the temperament domain and facets. In particular, we compared four measurement models: (a) freely estimating the factor loadings for the latent factors at each age of assessment (i.e., configural invariance); (b) constraining the respective factor loadings to be equal at each of assessment (i.e., weak invariance); (c) constraining the factor loadings and intercepts to be equal at each age of assessment (i.e., strong invariance); and (d) constraining the factor loadings, intercepts, and residual variances to be equal at each age of assessment (i.e., strict invariance). If the more constrained models did not fit worse than the lesser constrained models, then we concluded that the structure of the latent constructs is the same over time. We assessed adequate model fit via changes in chi-square and degrees of freedom and comparative fit index (CFI) less than or equal to .01 (Cheung & Rensvold, 2002; Meade et al., 2006). We also note values of root-mean-square error of approximation (RMSEA), for which adequate fit is indicated by values less than or equal to .06. We found evidence for strict longitudinal measurement invariance for all of the temperament domains and facets except Inhibitory Control, Depressed Mood, and Surgency, for which we found evidence for weak invariance (see Table S2.2 for model comparisons). We used the retained models to create factor scores, which were then included in the models.

We also examined evidence for measurement invariance across gender (i.e., girl vs. boy) and nativity (i.e., born in Mexico vs. born in U.S.) to facilitate multiple-groups analyses. Using the retained models from the longitudinal measurement invariance analyses, we tested four measurement models – configural, weak, strong, and strict invariance – for each of the temperament domains by constraining the relevant parameters across groups in each model. We found evidence for strict measurement invariance across gender and nativity for all temperament domains (see Tables S2.3-S2.4 for model comparisons).

Discrete-Time Survival Models. We used discrete-time survival analyses to examine the probability of the onset of suicidal ideation and behaviors from age 12 to 21, and how the probability of first experiencing suicidal ideation, plans, and attempts varied as a function of temperament (Múthen & Masyn, 2005; Singer & Willett, 2003). Survival analysis is a commonly used method for studying age at first suicidal ideation (e.g., Bolger et al., 1989) and *discrete-time* survival models are most appropriate for data collected at discrete-time intervals (e.g., annually) rather than on a more continuous-time interval (e.g., hourly, daily; Masyn, 2003; Singer & Willett, 2003).

Discrete-time survival analyses convey the probability of "whether, and if so, when" an adolescent will experience a non-repeatable event (i.e., the onset of suicidal ideation, plans, or attempts), given that they have not endorsed ideating, planning, or attempting suicide previously (Singer & Willett, 2003, p. 306). Once a participant experiences an event (e.g., a suicidal thought, plan, or attempt), they are no longer "at risk" for that event and are excluded from future waves aimed at predicting the age of onset. These models accommodate data that are right-censored, that is, data from participants who do not experience the event over the course of the observation period and it is unknown whether they experience suicidal ideation or behaviors after the course of the study. Further, using a logit link, these analyses produce odds ratios, which depict the percentage of increased or decreased risk for first experiencing suicidal ideation, plans, or attempts that is associated with a one-unit change in a given predictor (Sharaf & Tsokos, 2014; Xie et al., 2003). For the present study, we examined both time-varying predictors (i.e., temperament) and time-invariant covariates (i.e., gender, nativity). For temperament domains and facets, suicidal ideation, plans, and attempts at each age were regressed on their respective age-matched measure. Each model was specified so that there was a proportional effect on the onset of suicidal ideation and behaviors at each age, given the estimated baseline thresholds. For gender and nativity, we conducted multiple group analyses, which test whether the effects differed significantly between boys versus girls or between adolescents born in Mexico versus the U.S. To determine differences across groups, we used a Wald test to examine whether the odds ratios for each temperament domain and facet differed significantly across groups (Koode & Palm, 1986) and we also compared values of Akaike information criteria (AIC) and Bayesian information criteria (BIC) from models where the parameters were freely estimated versus constrained across groups.

Results

Mean-level Changes in Prevalence Rates from Age 12 to 21

Table 2.2 shows the percentage of adolescents who endorsed suicidal ideation, plans, and attempts from age 12 to 21. On average, youth tended to report increases in suicidal ideation from age 12 to 14 and then plateaued until age 21, except for a dip at age 18. Suicide plans and attempts both showed similar patterns of mean-level increases from age 12 to 14 and then a fairly stable leveling-off, except for a dip at age 18.[20] Unlike suicidal ideation, the average level of suicide plans and attempts at ages 19-21 did not reach the levels reported before the dip at age 18.

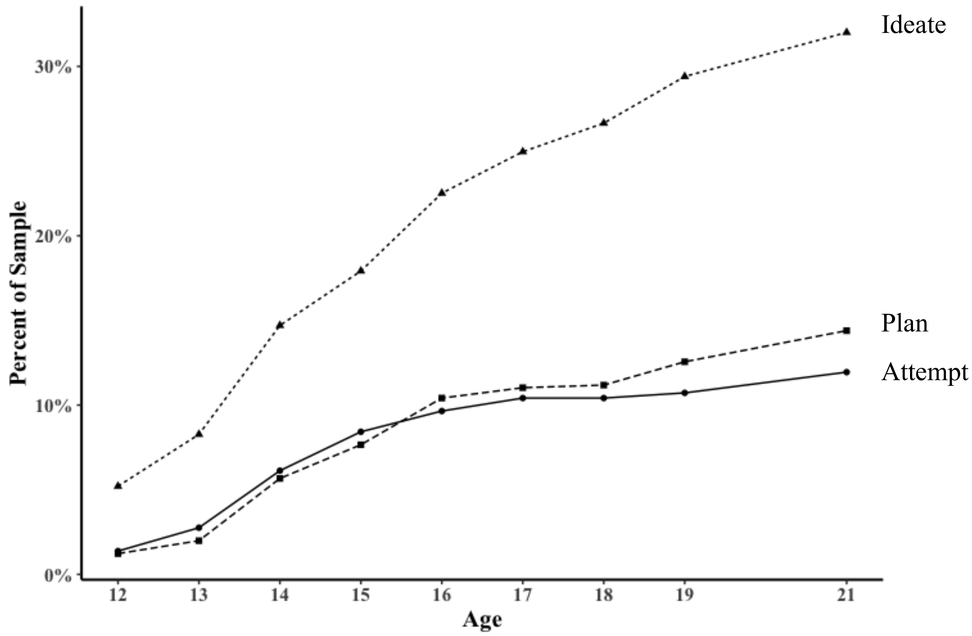
Table 2.2
Average prevalence of suicidal ideation, plans, and attempts.

	Ideation		Plans		Attempts		*N*
	No	Yes	No	Yes	No	Yes	
Age 12	94.1%	5.9%	98.6%	1.4%	98.4%	1.6%	577
Age 13	93.9%	6.1%	99.0%	1.0%	97.8%	2.2%	591
Age 14	88.9%	11.1%	95.0%	5.0%	95.4%	4.6%	605
Age 15	90.7%	9.3%	96.3%	3.7%	95.4%	4.6%	590
Age 16	88.0%	12.0%	94.2%	5.8%	95.8%	4.2%	600
Age 17	89.2%	10.8%	95.5%	4.5%	96.5%	3.5%	600
Age 18	93.7%	6.3%	98.1%	1.9%	98.8%	1.2%	587
Age 19	89.0%	11.0%	96.7%	3.3%	97.7%	2.3%	584
Age 21	89.7%	10.3%	96.8%	3.2%	97.2%	2.8%	542
Overall	68.0%	32.0%	85.6%	14.4%	88.1%	11.9%	653

Note. Percentages are rounded, so they may not add up to exactly 100%. "Overall" row includes the percentages of youth who reported ideating, planning or attempting at *any point* from age 12 to 21.

In addition to mean-level changes in suicide variables, we also examined the number of unique participants who endorsed suicidal ideation, plans, and attempts at any point across the 10 years of the study. We found that 32.0% of the participants endorsed suicidal ideation at some point from age 12 to 21; 14.4% of participants endorsed making a suicide plan; and, 11.9% of participants endorsed attempting suicide at some point (see Figure 2.1). Given the presence of a non-trivial number of unique participants endorsing suicidal ideation, plans, and attempts across adolescence, we next examined predictors of the onset of suicidal ideation and behaviors across adolescence.

Figure 2.1
Percent of sample who experienced suicide risk factors at any point from age 12 to 21, separately for ideation, plans, and attempts



Temperamental Associations with the Onset of Suicide Ideation and Behaviors

Table 2.3 shows the odds ratios (OR) and 95% confidence intervals (CI) for each of our temperament predictors on the likelihood of initiating suicidal ideation, plans, or attempts across adolescence. To assess statistical significance, we applied a Bonferroni correction to account for multiple comparisons resulting in $\alpha = .005$.[21] An OR of exactly 1.00 would imply that the covariate was not associated with the onset of the behavior, whereas an OR below 1.00 would indicate a decreased propensity to initiate and an OR above 1.00 would indicate an increased propensity to initiate. Given that temperament is a continuous predictor, odds ratios correspond to a one-unit increase in the predictor (e.g., at each age, the estimated odds of initiating suicidal ideation are about 50% lower for youth whose Effortful Control factor scores were one unit higher).[22] Notably, the temperament variables all have standard deviations around 1, so a one unit increase is approximately the same as a one standard deviation increase.

Table 2.3
Odds ratio results from survival analyses with latent time-varying covariates

	Effortful Control		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Effortful control	0.55* [0.44, 0.67]	0.55* [0.40, 0.74]	0.60* [0.43, 0.83]
Activation Control	0.59* [0.46, 0.74]	0.59* [0.41, 0.85]	0.67 [0.47, 0.97]
Inhibitory Control	0.66* [0.53, 0.83]	0.65* [0.49, 0.86]	0.71 [0.52, 0.97]
Attention	0.52* [0.42, 0.64]	0.53* [0.39, 0.73]	0.60* [0.43, 0.84]
	Negative Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Negative emotionality	2.10* [1.76, 2.51]	1.95* [1.55, 2.44]	1.67* [1.32, 2.12]
Negative emotionality (without Depressed Mood & Aggression)	1.61* [1.32, 1.96]	1.51 [1.14, 1.99]	1.27 [0.94, 1.72]
Fear	1.13 [0.87, 1.46]	1.10 [0.74, 1.62]	0.87 [0.57, 1.34]
Frustration	1.86* [1.57, 2.21]	1.74* [1.37, 2.20]	1.59* [1.23, 2.05]
Shyness	1.23 [0.97, 1.54]	1.14 [0.81, 1.60]	1.09 [0.76, 1.56]
Aggression	1.68* [1.42, 1.99]	1.56* [1.24, 1.95]	1.47* [1.18, 1.84]
Depressed Mood	2.85* [2.35, 3.46]	2.61* [2.17, 3.14]	2.27* [1.76, 2.93]
	Positive Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Surgency	1.10 [0.88, 1.37]	1.13 [0.84, 1.53]	1.09 [0.78, 1.53]
Affiliation	1.03 [0.82, 1.30]	1.10 [0.82, 1.49]	1.03 [0.73, 1.44]

Note. *$p < .005$. ORs whose 95% CI's do not include 1 are significant at $p < .05$.

In particular, we found that youth who had EC scores that were one unit *higher* were 45% *less* likely to experience the onset of suicidal ideation, 45% *less* likely to experience the onset of suicide plans, and 40% *less* likely to experience the onset of suicide attempts (all *ps* $< .001$). Further, youth who had NE scores that were one unit higher (assessed via Fear, Frustration, and Shyness facets) were 60% more likely to experience the onset of suicidal ideation ($p < .001$). However, there was no significant association between this narrow NE domain and suicide plans or attempts. Using a broader NE conceptualization including Aggression and Depressed Mood resulted in larger OR's of suicidal ideation and plans. In particular, youth who had NE scores (assessed via Fear, Frustration, Shyness, Aggression, and Depressed Mood facets) that were one unit higher were more than twice as likely to experience the onset of suicidal ideation ($p < .001$). Further, youth one unit higher on this broad conceptualization of NE were almost twice as likely to experience the onset of suicide plans ($p < .001$) and about 70% more likely to experience the onset of suicide attempts ($p = .001$). Positive emotionality, assessed via both the Surgency and Affiliation facets, was not significantly associated with the onset of suicidal ideation, plans, or attempts.[23]

Together, these results suggest that EC and NE are particularly consequential in the onset of suicidal ideation, plans, and attempts across adolescence. We also examined the facet-level results to determine which facets were driving these associations (Table 2.3). Facet results indicated that all three facets of EC (Activation Control, Inhibitory Control, and Attention) were associated with decreased probability of experiencing the onset of suicidal ideation, plans, and attempts (only for Attention) across adolescence. For the NE domain, the results show that that Frustration, Aggression, and Depressed Mood were all significantly associated with suicidal ideations, plans, and attempts, whereas Fear and Shyness were unrelated to suicidal ideation and behaviors. Noticeably, Depressed Mood had the largest association with suicidal ideation and behaviors. A one unit increase in Depressed Mood was associated with being almost three times as likely to experience the onset of suicidal ideation ($p < .001$), two and a half times as likely to experience the onset of planning ($p < .001$), and over two times as likely to experience the onset of attempting ($p < .001$).

Role of Informant. To examine the role of informants on the association between temperament and the onset of suicidal ideation, plans, and attempts, we also ran discrete-time survival analyses separately for self- and mother-reported temperament. Results from the analyses using adolescent self-report ratings of temperament parallel the results from analyses using both self- and mother-ratings. Indeed, with one exception (Frustration no longer significantly predicted suicide attempts; OR = 1.48, $p = .021$), all findings remained significant and had similar effect sizes when only using adolescent self-reported temperament (Table S2.6). Using only mother-ratings of adolescent temperament, we replicated 9 out of 23 (36%) of the effects that were significant using both self- and mother-reported temperament (Table S2.7 using $p < .005$), though the vast majority of effects (87%) were in the expected direction and 61% were replicated using $p < .05$. To contextualize these findings, we believe that the best measure of temperament is the composite measure of self- and mother-reports, so we believe that the findings in Table 2.3 are the best representation of the relation between temperament and the onset of suicidal ideation and behaviors.

Temperament and Personality. To complement our analyses of temperament, we also examined the role of Big Five personality traits in the onset of suicidal ideation, plans, and attempts (see Table S2.8 for descriptive statistics).[24] Using self-reported Big Five personality traits at ages 14 and 16 and suicide data from age 14 to 21, our results generally mirrored those from the temperament analyses (see Table S2.9). In particular, we found that, similar to EC, Conscientiousness was associated with decreased propensity to initiate suicidal ideation, plans, and attempts across adolescence. Similar to NE, Neuroticism was associated with increased propensity to initiate suicidal ideation, plans, and attempts across adolescence. Similar to PE, Extraversion, Agreeableness, and Openness were not associated with the onset of suicidal ideation, plans, and attempts (except that Extraversion was associated with a decreased propensity toward suicidal ideation and Agreeableness was associated with an increased propensity toward suicidal ideation).

Moderating Effects of Gender and Nativity

Given that there was evidence of several significant associations between temperament and the onset of suicidal ideation, plans, and attempts, we next examined whether gender or nativity moderated these associations. Before doing so, however, we first examined whether the prevalence of adolescent suicidal ideation and behaviors varied by gender and nativity.

Tables S2.10 and S2.12 show the percentage of youth exhibiting suicidal ideation, plans, and attempts separately for boys and girls and for youth born in Mexico versus the U.S. Across adolescence, girls were three to four times more likely to ideate about suicide than boys (e.g., at age 15, 13.5% of girls reported ideation compared to only 4.2% of boys). These average gender differences were similar for plans (e.g., at age 15, 4.9% of girls reported planning in the past year compared to only 1.5% of boys) and attempts (e.g., at age 15, 6.8% of girls reported attempting in the past year compared to only 1.5% of boys). Further, these gender differences bore out in the number of unique individuals who ideated, planned, and attempted across the 10 years of the study. In particular, 44.0% of the females reported suicidal ideation at some point during the study, whereas 19.9% of the males reported ideating at least once from age 12 to 21 (χ^2 = 42.46,

$p < .001$). Furthermore, 21.7% of females made suicide plans at some point during the study, compared to 7.1% of males ($\chi^2 = 27.28$, $p < .001$). Finally, 18.3% of the females reported having attempted suicide at some point during the study, compared to 5.5% of males ($\chi^2 = 24.33$, $p < .001$).

For nativity status, we did not find any consistent trends for the average levels of ideation, plans, or attempts for youth born in the U.S. versus Mexico. Additionally, we did not find significant nativity differences in the number of participants who ideated, planned, and attempted. In particular, 28.3% of the youth born in Mexico ideated at some point from age 12 to 21, whereas 33.7% of the youth born in the United States ideated during this period ($\chi^2 = 1.47$, $p = .225$). Further, 12.8% of youth born in Mexico reported making suicide plans, compared to 15.2% of youth born in the U.S ($\chi^2 = 0.45$, $p = .503$). Finally, 10.0% of the youth born in Mexico reporting attempting suicide at some point during the study, compared to 12.9% of youth born in the U.S ($\chi^2 = 0.76$, $p = .384$).

After examining differences in the prevalence of suicidal ideation and behaviors by gender and nativity, we then conducted multiple group analyses to examine whether gender and nativity moderated the individual effects of temperament on the onset of suicidal ideation and behaviors by testing whether equality constraints on the estimated thresholds were significantly different across groups. We found no evidence that gender or nativity significantly moderated any associations (see Tables S2.11 and S2.13).

Discussion

The present study examined the role of temperament in the onset of suicidal ideation, plans, and attempts across ten years, from age 12 to 21, in a large, community sample of Mexican-origin youth ($N = 674$). We used discrete-time survival analyses to assess whether three central domains of youth temperament – EC, NE, and PE – served as risk or protective factors for suicidal ideation and behaviors. Our results suggested that EC was associated with *decreased* probability of experiencing the onset of suicidal ideation and behaviors, thus serving as a protective factor, whereas NE was associated with an *increased* probability of experiencing the onset of suicidal ideation and behaviors, thus serving as a risk factor. We found no evidence suggesting that PE was related to the onset of suicidal ideation and behaviors across adolescence.

Adolescent Suicidal Ideation and Behaviors

Data on the prevalence of suicidal ideation, plans, and attempts from the present study indicate that there were relatively high levels of suicidal ideation and behaviors in this sample of Mexican-origin youth. In particular, we found that, at some point from age 12 to age 21, 32.0% of youth experienced suicidal ideation, 14.4% made a suicide plan, and 11.9% attempted suicide. These prevalence rates are approximately three times higher than the rates from a nationally representative survey of American adolescents aged 13-18, which found prevalence of 12.1%, 4.0%, and 4.1% for ideation, plans, and attempts, respectively (Nock et al., 2013).

However, the prevalence rates in the present study are comparable to those found in previous studies examining Latinx youth. In particular, the prevalence rates for plans and attempts (14.4% and 11.9%, respectively) are very similar to those found in a large sample of Latinx high school students in the Youth Risk Behavior Surveillance (YRBS) study (15.7% and 11.3%, respectively; Kann et al., 2013). We did find a higher rate of ideating in the present study (32.0%) than in the YRBS study (18.9%), but this could reflect the different ages captured by each study (i.e., age 12 to 21 years in the present study vs. 9th to 12th grade in the YRBS). When we limited the timeframe of the present study to high school, the rate of ideation (22.4%) was much closer to the 18.9% in the YRBS study. Additionally, we might be seeing slightly higher rates in the present study because our participants have been part of the California Families Project for many years and have grown comfortable reporting emotional problems and learned to trust that their responses will remain confidential, whereas participants in large national surveys are unlikely to have the same personal connection with the project.

With respect to gender, we found that girls were two to three times more likely to experience suicidal ideation, plans, and attempts than boys. This gender disparity is consistent with past research that found that suicidal ideation and behaviors are approximately three times more prevalent among girls compared to boys (Nock et al., 2013). We also found differences in the onset and trajectory of suicide risk by gender that is consistent with past research (Reuter, Holm, McGeorge, & Conger, 2008). In particular, we found that girls experienced earlier onset, such that, at age 12, 8.8% of girls ideated at age 12 versus 2.9% of boys. Additionally, we found gender differences in peak prevalence, such that ideation for girls peaked at age 14 with 17.2% of female participants endorsing suicidal ideation and it peaked for boys at age 17 with 6.9% endorsing ideation. Similar patterns were observed for gender differences in plans and attempts.

The strikingly high prevalence of suicidal ideation (44%), plans (22%), and attempts (18%) among girls in our sample corresponds to the well-documented finding that Latina adolescents consistently report the highest suicide ideation and attempt rates of any gender and ethnic group (Eaton et al., 2011; Price & Khubchandani, 2017). Indeed, when we consider the differences between the studies noted above – capturing a long period of time and having participants who are especially comfortable with interviewers – our rates are comparable with data from the YRBS study, which showed that Latina high school students had the highest prevalence rates of any demographic group for suicidal ideation (26.0%), plans (20.1%), and attempts (15.6%) (Kann et al., 2014).

For nativity, we found no evidence of a significant difference between suicidal ideation, plans, or attempts for youth born in Mexico versus those born in the United States. These results are not consistent with past research showing that youth born in Mexico were *less* likely to experience suicidal ideation and behaviors than youth born in the United States (Carino & Roberts, 2001; Silva & Van Orden, 2018; Sorenson & Shen, 1996) and that Latinx youth born in the U.S. were two to three times *more* likely to attempt suicide than youth born outside of the U.S. (Peña et al., 2008). Thus, our results are not consistent with the "immigrant paradox," which posits that U.S.-born ethnic minority youth have more emotional and behavioral problems than 1st generation ethnic minority youth who were born in another country and immigrated to the U.S. (García Coll & Marks, 2012; Marks et al., 2014).

Temperament and Suicidal Ideation and Behaviors

Supporting the vulnerability model of personality and psychopathology, we found evidence that multiple temperament domains (and their facets) were associated with the onset of suicidal ideation and behavior across adolescence in our sample of Mexican-origin youth. In almost all cases, the associations with temperament replicated across suicidal ideation, plans, and attempts, suggesting that the importance of temperament in identifying youth at higher or lower risk for suicidal ideation and behaviors is not limited to any particular aspect of suicide risk. Below, we detail the specific relations between EC, NE, and PE and the onset of suicidal ideation and behaviors.

Effortful Control. We found that youth who were higher in EC had a substantially *decreased* risk of experiencing the onset of suicidal ideation, plans, and attempts. This finding is consistent with past longitudinal work showing that impulsivity (which is related to low EC) is associated with increased suicide risk later in life (Kasen et al., 2011; Piquet & Wagner, 2003).

Notably, the association between EC and suicidal ideation and behaviors replicated across all EC facets, suggesting that not only is there an association between suicidal ideation/behaviors and Inhibitory Control (i.e., low impulsivity), but also a relation between ideation/behaviors and Activation Control and Attention. This finding is also consistent with previous claims that Conscientiousness (closely related to EC) is particularly relevant in the development of psychopathology (Tackett, 2006). Indeed, in the present study, we found that higher levels of Conscientiousness were associated with decreased propensity to experience the onset of suicidal ideation, plans, and attempts.

Negative Emotionality. We found that youth who were higher in NE, particularly the Frustration, Aggression, and Depressed Mood facets, had a substantially *increased* risk of experiencing the onset of suicidal ideation, plans, and attempts. These findings are consistent with past longitudinal work suggesting that high levels of NE are associated with later suicidal ideation and behavior (Enns et al., 2003; Fergusson et al., 2003; Pickles et al., 2010). Additionally, results suggesting a large association between Depressed Mood and suicidal ideation and behaviors are consistent with the substantial literature on relations between clinical (rather than temperament) measures of depression and suicidal ideation/behaviors (Minkoff et al., 1973). These results are especially interesting given the content captured by the Depressed Mood subscale of the EATQ-R. In particular, most EATQ-R Depressed Mood items relate to sad mood (“You get sad more than other people realize”, “You get sad when a lot of things are going wrong”, “You feel sad even when you should be enjoying yourself, like at Christmas or on a trip”, “Your friends seem to enjoy themselves more than you do”) which is the most common symptom captured on depression inventories (Fried, 2017). The other items capture crying (“It often takes very little to make you feel like crying”) and feeling happy (“You feel pretty happy most of the day" – reverse scored), which are both less commonly included depression symptoms. Notably, the EATQ-R does not include any of the other types of depression symptoms described by Fried (2017), including none of the somatic symptoms nor symptoms relating to suicidal ideation. This bolsters the importance of the *temperament* construct of Depressed Mood in the development of suicidal ideation and behaviors, and highlights the distinction between this scale and our measures of suicidal ideation, plans, and attempts. Further, the importance of NE also bolsters claims that Neuroticism (analogous to NE) is particularly relevant in the development of later psychopathology (Tackett, 2006). Indeed, the present study found that higher levels of Neuroticism were associated with increased propensity to experience the onset of suicidal ideation, plans, and attempts across adolescence.

Positive Emotionality. We did not find evidence that either the Surgency facet or the Affiliation facet of PE was associated with suicidal ideation and behaviors. These findings are not consistent with previous research showing that Extraversion is associated with reduced risk, and novelty seeking with higher risk, of suicidal ideation and behaviors (Fergusson et al, 2003). The Big Five analyses in the present study also failed to support an important role of Extraversion. In particular, Extraversion was not associated with suicidal plans or attempts, although it was associated with a decreased propensity toward suicidal ideation.

Although we believe the best measure of adolescent temperament is the composite of both the adolescent and mother perspectives, we also examined whether the findings differed when using only self-report or mother report of the adolescent's temperament. We found identical results for the adolescent self-reports that we found with the composite measure. For the mother-reports of adolescent temperament, we also saw similar patterns of effects, but they were not always significant.

Finally, we did not find evidence that gender or nativity moderated the association between temperament and suicidal ideation and behaviors. These results suggest that EC and NE have similar associations with suicide risk for boys and girls and for youth born in the U.S. and Mexico. With respect to gender, this finding is inconsistent with past research that documented gender differences in the association between individual-level risk and protective factors and suicidal ideation and behaviors (Edwards & Holden, 2001). In particular, we did not replicate the finding from a cross-sectional study of undergraduate students that women high in angry hostility and depression (i.e., NE) were more prone to suicidal ideation, whereas men with poor self-discipline (i.e., low EC) were more prone to suicidal ideation (Velting, 1999). Additionally, we did not replicate findings from a cross-sectional study of German adults that found men low in Extraversion and Conscientiousness were at increased risk for suicide, whereas women high in Neuroticism were at increased risk (Blüml et al., 2013).

Limitations

The present study has several limitations. First, our suicide risk measures were single-item assessments of suicide ideation, plans, and attempts that did not provide participants with definitions of what each of these meant, which might have impacted youth's responses to those items (Franklin et al., 2017) and could be prone to misclassification compared to multiple-item measures (Millner, Lee, & Nock, 2015). Second, we only assessed suicidal ideation and behaviors until age 21. Although this is the developmental period when most people first exhibit suicidal ideation and behaviors, some individuals exhibit these later in young adulthood and beyond. Because of this, we may not have captured all of the participants who went on to experience suicidal ideation or behaviors during their lifetimes. Third, our use of survival analyses meant that we did not examine severity or duration of suicidal ideation and behaviors.

Although survival analyses have been used with data on the onset of suicidal ideation and behaviors (e.g., Bolger et al., 1989), future studies should examine these other aspects of suicide risk. Relatedly, in order to run the survival analyses, we assumed that the censoring was noninformative, though it is possible that some cases of censoring were informative (i.e., youth stopped participating in the study because they experienced the onset of suicidal ideation but they had yet to be documented as having experienced it; Singer & Willett, 2003). This is especially relevant given that our data were "group-timed" (Allison, 1982); that is, the onset of suicidal ideation, plans, or attempts occurred at some point in the previous 12 months but we were only able to document the event occurrence happening at our annual assessment. Fourth, although we reference prevalence rates from other studies to contextualize our results, we cannot draw conclusions about whether the prevalence of suicidal ideation, plans, and attempts of Mexican-origin youth are high compared to other ethnic groups, given that we do not have a matched comparison sample. Fifth, although we found that temperament was associated with the onset of suicidal ideation and behaviors across adolescence and into young adulthood, we still have a limited understanding of the mechanisms through which temperament serves as a risk or protective factor, or the family and social-contextual processes that might moderate the influence of temperament on suicide risk. Relatedly, although our prospective longitudinal design is a strength in testing whether temperament is associated with suicidal ideation/behaviors, more evidence is needed for strong causal inferences; in particular, researchers could make use of genetically-informed designs (Durbin & Hicks, 2014).

Future Directions

In the future, researchers should aim to replicate these findings in other samples, including samples with different demographic makeups. In particular, researchers should compare Mexican-origin youth to other (especially other Latinx) subsamples. Once replicated, researchers should seek to understand possible mediators of this connection; in particular, which factors might be part of a developmental sequence, or cascade, that leads from temperament to suicidal ideation, plans, and attempts (Dumais et al., 2005; Pickles et al., 2010). Additionally, researchers should explore whether mental healthcare providers at schools and community centers could use measures of adolescent temperament to identify adolescents likely to experience suicidal ideation and behaviors before the onset actually occurs. Indeed, this is one of the most beneficial outcomes of personality and psychopathology research; that is, to use early assessments of temperament to identify individuals who are particularly at risk for suicidal ideation and behaviors (Frick, 2004). Researchers should also explore socio-cultural factors that may underlie the gender disparity documented in past studies and replicated in the present study, where Latinas are at a substantially higher risk than Latinos for experiencing the onset of suicidal ideation, plans, and attempts across adolescence. In addition, there may be endogenous biological factors that undergird gender differences in suicidal ideation and behavior; in particular, there is evidence from research on complex trait genetics that suggests pleiotropic effects on suicidal behavior and MDD, both of which tend to be higher in females compared to males (Levey et al., 2019). Finally, researchers should examine evidence for other developmental models of temperament and suicidal ideation and behaviors, including the scar, pathoplasty, and spectrum models (Durbin & Hicks, 2014; Tackett, 2006).

Supplemental Materials

Table S2.1

Descriptive statistics for observed temperament domains.

	Age 12		Age 14		Age 16	
	M (SD)	*α / ω*	*M (SD)*	*α / ω*	*M (SD)*	*α / ω*
Effortful control	3.00 (.35)	.82 / .84	2.94 (.35)	.84 / .86	2.94 (.33)	.82 / .84
Negative emotionality	2.02 (.28)	.85 / .87	1.94 (.30)	.86 / .88	1.89 (.30)	.83 / .86
Negative emotionality (without aggression and depressed mood)	2.32 (.35)	.75 / .80	2.18 (.34)	.71 / .77	2.12 (.34)	.76 / .81
Positive emotionality	2.59 (.39)	.76 / .79	2.66 (.37)	.77 / .81	2.71 (.32)	.81 / .84

Note. M = mean, SD = standard deviation, α = alpha reliability, ω = omega reliability. Descriptive statistics for effortful control and negative emotionality reflect the results for the broader domains, whereas the statistics for positive emotionality (PE) reflect only the results from Surgency, the core of PE.

Table S2.2
Model comparisons for longitudinal measurement invariance

	χ^2	*df*	CFI	RMSEA [90% CI]
Effortful Control				
Configural	163.30	43	.97	.066 [.056, .077]
Weak	175.74	49	.97	.064 [.054, .074]
Strong	222.77	55	.96	.069 [.060, .079]
Strict	235.97	63	.95	.066 [.057, .075]
Activation Control				
Configural	44.50	18	.99	.048 [.030, .066]
Weak	46.39	22	.99	.042 [.025, .058]
Strong	48.41	26	.99	.037 [.020, .053]
Strict	73.08	32	.98	.045 [.031, .059]
Attention Control				
Configural	68.03	18	.97	.066 [.050, .083]
Weak	78.42	22	.96	.063 [.049, .079]
Strong	80.11	26	.96	.057 [.043, .072]
Strict	98.15	32	.96	.057 [.044, .070]
Inhibitory Control				
Configural	138.17	18	.89	.102 [.087, .119]
Weak	141.46	22	.89	.092 [.078, .107]
Strong	224.80	26	.82	.109 [.097, .123]
Strict	240.17	32	.82	.101 [.089, .113]
Negative Emotionality				
Configural	193.07	43	.96	.074 [.064, .085]
Weak	201.08	49	.96	.070 [.060, .080]
Strong	294.38	55	.94	.083 [.073, .092]
Strict	304.38	63	.94	.077 [.069, .086]
Negative Emotionality (without Depressed Mood & Aggression)				
Configural	185.56	43	.97	.072 [.062, .083]
Weak	202.36	49	.97	.070 [.060, .080]
Strong	293.50	55	.95	.082 [.073, .092]
Strict	305.56	63	.95	.078 [.069, .087]
Fear				
Configural	216.20	18	.87	.131 [.116, .147]

Weak	226.70	22	.87	.121 [.107, .135]
Strong	249.74	26	.85	.116 [.103, .129]
Strict	272.34	32	.84	.109 [.097, .121]
Frustration				
Configural	57.92	18	.98	.059 [.042, .076]
Weak	59.38	22	.98	.052 [.036, .068]
Strong	68.26	26	.98	.050 [.036, .065]
Strict	75.29	32	.98	.046 [.033, .060]
Shyness				
Configural	52.78	18	.99	.055 [.038, .073]
Weak	57.28	22	.99	.050 [.034, .066]
Strong	61.26	26	.99	.046 [.031, .061]
Strict	91.85	32	.98	.054 [.041, .067]
Aggression				
Configural	142.94	18	.95	.104 [.089, .121]
Weak	153.37	22	.95	.097 [.083, .111]
Strong	182.11	26	.94	.097 [.084, .111]
Strict	200.39	32	.94	.091 [.079, .103]
Depressed Mood				
Configural	88.50	18	.96	.078 [.062, .095]
Weak	93.59	22	.96	.071 [.057, .087]
Strong	184.38	26	.91	.098 [.085, .111]
Strict	194.28	32	.91	.089 [.077, .101]
Positive Emotionality				
Configural	559.25	43	.78	.137 [.127, .147]
Weak	566.41	49	.78	.129 [.119, .138]
Strong	826.18	55	.67	.148 [.139, .157]
Strict	1081.80	63	.57	.159 [.151, .168]
Affiliation				
Configural	101.57	18	.94	.085 [.070, .102]
Weak	104.83	22	.94	.077 [.062, .092]
Strong	143.82	26	.91	.084 [.071, .098]
Strict	149.91	32	.91	.076 [.064, .088]

Note. χ^2 = Chi-square test statistics; df = degrees of freedom; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; 90%CI = 90% confidence interval.

Table S2.3
Results from tests of measurement invariance across gender

	χ^2	*df*	CFI	RMSEA [90% CI]
Effortful Control				
Configural	307.83	126	.95	.067 [.058, .077]
Weak	308.55	130	.95	.066 [.056, .075]
Strong	320.12	134	.95	.066 [.057, .075]
Strict	324.26	138	.95	.065 [.056, .074]
Negative Emotionality				
Configural	395.30	126	.95	.082 [.073, .091]
Weak	398.92	130	.95	.081 [.072, .090]
Strong	436.77	134	.94	.084 [.075, .093]
Strict	437.72	138	.94	.083 [.074, .091]
Negative Emotionality (without Depressed Mood & Aggression)				
Configural	405.43	126	.93	.083 [.074, .093]
Weak	407.63	130	.93	.082 [.073, .091]
Strong	443.07	134	.92	.085 [.076, .094]
Strict	445.75	138	.92	.084 [.075, .092]
Positive Emotionality				
Configural	1080.14	126	.60	.154 [.146, .163]
Weak	1142.10	130	.57	.146 [.148, .165]
Strong	1194.21	134	.55	.157 [.149, .166]
Strict	1198.56	138	.55	.155 [.147, .163]

Note. χ^2 = Chi-square test statistics; df = degrees of freedom; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; 90%CI = 90% confidence interval.

Table S2.4
Results from tests of measurement invariance across nativity.

	χ^2	*df*	CFI	RMSEA [90% CI]
Effortful Control				
Configural	314.34	126	.95	.069 [.059, .078]
Weak	316.60	130	.95	.067 [.058, .077]
Strong	327.57	134	.95	.068 [.058, .077]
Strict	333.30	138	.95	.067 [.058, .076]
Negative Emotionality				
Configural	367.30	126	.94	.078 [.069, .087]
Weak	368.90	130	.94	.076 [.067, .086]
Strong	383.54	134	.94	.077 [.068, .086]
Strict	388.47	138	.94	.076 [.067, .085]
Negative Emotionality (without Depressed Mood & Aggression)				
Configural	365.90	126	.95	.078 [.068, .087]
Weak	366.88	130	.95	.076 [.067, .085]
Strong	376.04	134	.95	.076 [.067, .085]
Strict	379.11	138	.95	.074 [.066, .083]
Positive Emotionality				
Configural	1120.72	126	.57	.158 [.150, .167]
Weak	1127.83	130	.57	.156 [.148, .164]
Strong	1131.62	134	.57	.154 [.145, .162]
Strict	1133.85	138	.57	.151 [.143, .159]

Note. χ^2 = Chi-square test statistics; df = degrees of freedom; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; 90%CI = 90% confidence interval.

Table S2.5
Odds ratio results from survival analyses with observed aggregate adolescent- and mom-report time-varying covariates

	Effortful Control		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Effortful control	0.18*[0.10, 0.32]	0.20*[0.08, 0.52]	0.25*[0.09, 0.64]
Activation Control	0.37*[0.24, 0.56]	0.39*[0.20, 0.76]	0.43*[0.21, 0.87]
Inhibitory Control	0.38*[0.23, 0.64]	0.39*[0.19, 0.79]	0.37*[0.17, 0.81]
Attention	0.20*[0.12, 0.33]	0.24*[0.11, 0.54]	0.32*[0.14, 0.72]
	Negative Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Negative emotionality	11.25*[6.08, 20.82]	8.37*[3.75, 18.70]	5.98*[2.44, 14.61]
Negative emotionality (without Depressed Mood & Aggression)	3.25*[1.88, 5.61]	2.47*[1.14, 5.36]	1.78 [0.72, 4.14]
Fear	1.23 [0.78, 1.93]	1.17 [0.57, 2.39]	0.69 [0.30, 1.57]
Frustration	3.51*[2.43, 5.06]	2.89*[1.73, 4.81]	2.87*[1.62, 5.10]
Shyness	1.37 [0.93, 2.03]	1.12 [0.60, 2.09]	1.06 [0.55, 2.03]
Aggression	3.56*[2.37, 5.36]	3.20*[1.85, 5.51]	3.26*[1.85, 5.75]
Depressed Mood	16.17*[9.71, 26.91]	12.48*[7.44, 20.94]	8.79*[4.54, 17.00]
	Positive Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Surgency	1.68 [0.95, 2.99]	1.70 [0.70, 4.12]	1.46 [0.58, 3.71]

Affiliation	1.36 [0.86, 2.16]	1.34 [0.72, 2.51]	1.38 [0.70, 2.72]

Note. * $p < .005$

Table S2.6
Odds ratio results from survival analyses with adolescent self-report latent time-varying covariates

	Effortful Control		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Effortful control	0.50* [0.39, 0.64]	0.48* [0.32, 0.72]	0.57* [0.38, 0.85]
Activation Control	0.56* [0.42, 0.74]	0.54* [0.34, 0.86]	0.65 [0.40, 1.04]
Inhibitory Control	0.74* [0.60, 0.92]	0.69* [0.52, 0.92]	0.71 [0.50, 0.99]
Attention	0.50* [0.39, 0.62]	0.49* [0.34, 0.69]	0.59* [0.42, 0.84]
	Negative Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Negative emotionality	2.22* [1.85, 2.66]	2.04* [1.60, 2.60]	1.73* [1.35, 2.21]
Negative emotionality (without Depressed Mood & Aggression)	1.75* [1.45, 2.12]	1.68* [1.28, 2.20]	1.40 [1.04, 1.87]
Fear	1.37 [1.08, 1.75]	1.34 [0.94, 1.90]	1.09 [0.71, 1.66]
Frustration	1.84* [1.54, 2.19]	1.74* [1.36, 2.23]	1.48 [1.12, 1.94]
Shyness	1.32 [1.04, 1.66]	1.22 [0.88, 1.69]	1.28 [0.92, 1.79]
Aggression	1.73* [1.47, 2.03]	1.61* [1.28, 2.04]	1.59* [1.24, 2.04]
Depressed Mood	2.88* [2.39, 3.48]	2.51* [2.02, 3.11]	2.04* [1.63, 2.54]
	Positive Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Surgency	1.06 [0.85, 1.33]	1.16 [0.83, 1.62]	1.03 [0.71, 1.49]

Affiliation	1.14 [0.93, 1.40]	1.04 [0.78, 1.39]	0.98 [0.68, 1.40]

Note. * $p < .005$

Table S2.7

Odds ratio results from survival analyses with mom-report of adolescent temperament latent time-varying covariates

	Effortful Control		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Effortful control	0.71* [0.58, 0.86]	0.70* [0.53, 0.95]	0.77 [0.55, 1.06]
Activation Control	0.75* [0.60, 0.92]	0.75 [0.55, 1.03]	0.83 [0.59, 1.15]
Inhibitory Control	0.73* [0.58, 0.92]	0.74 [0.53, 1.04]	0.87 [0.61, 1.25]
Attention	0.67* [0.54, 0.83]	0.68* [0.50, 0.92]	0.75 [0.52, 1.07]
	Negative Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Negative emotionality	1.35 [1.11, 1.64]	1.29 [0.98, 1.69]	1.23 [0.88, 1.72]
Negative emotionality (without Depressed Mood & Aggression)	1.14 [0.94, 1.40]	1.06 [0.79, 1.41]	0.98 [0.70, 1.39]
Fear	0.88 [0.69, 1.12]	0.83 [0.57, 1.21]	0.73 [0.48, 1.12]
Frustration	1.34 [1.11, 1.61]	1.22 [0.95, 1.56]	1.30 [0.96, 1.75]
Shyness	1.09 [0.87, 1.37]	1.05 [0.74, 1.49]	0.91 [0.61, 1.36]
Aggression	1.34 [1.08, 1.67]	1.27 [0.92, 1.75]	1.16 [0.85, 1.60]
Depressed Mood	1.72* [1.44, 2.06]	1.83* [1.45, 2.30]	1.87* [1.35, 2.58]
	Positive Emotionality		
	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Surgency	1.08 [0.88, 1.32]	1.07 [0.80, 1.44]	1.11 [0.78, 1.57]
Affiliation	0.92 [0.72, 1.18]	1.05 [0.74, 1.49]	1.00 [0.66, 1.50]

Note. * $p < .005$

Table S2.8

Descriptive statistics for latent Big Five personality domains using factor scores.

	Age 14		Age 16	
	M (SD)	*α* / *ω*	*M (SD)*	*α* / *ω*
Extraversion	0 (.99)	.73 / .80	-0.28 (.99)	.75 / .81
Agreeableness	0 (1.01)	.69 / .76	-0.13 (.96)	.64 / .68
Conscientiousness	0 (.99)	.73 / .80	-0.19 (.98)	.69 / .77
Neuroticism	0 (.99)	.72 / .78	0.19 (1.00)	.73 / .78
Openness	0 (.99)	.74 / .79	0.02 (1.02)	.72 / .77

Note. M = mean, SD = standard deviation,. α = alpha reliability, *ω* = omega reliability. Mean and SD are provided for the saved factor scores from the latent variables.

Table S2.9
Odds ratio results from survival analyses with Big Five personality latent time-varying covariates

	Ideation Odds Ratio [95% CI]	Plans Odds Ratio [95% CI]	Attempts Odds Ratio [95% CI]
Extraversion	0.77* [0.60, 0.98]	0.93 [0.64, 1.35]	0.82 [0.54, 1.25]
Agreeableness	0.79* [0.65, 0.95]	0.83 [0.67, 1.03]	0.83 [0.59, 1.15]
Conscientiousness	0.61* [0.48, 0.78]	0.59* [0.42, 0.84]	0.72* [0.50, 1.03]
Neuroticism	3.38* [2.64, 4.33]	2.94* [2.27, 3.80]	2.34* [1.81, 3.01]
Openness	1.17 [0.95, 1.43]	1.28 [0.96, 1.70]	1.23 [0.88, 1.72]

Note. * $p < .05$

Table S2.10

Prevalence of suicidal ideation, plans, and attempts separated by gender.

	Ideation		Plans		Attempts	
	Girls	Boys	Girls	Boys	Girls	Boys
Age 12	8.8%	2.9%	2.2%	0.4%	2.6%	0.4%
Age 13	9.1%	2.6%	2.3%	0.0%	3.4%	0.4%
Age 14	17.2%	4.8%	7.3%	2.2%	6.6%	1.8%
Age 15	13.5%	4.2%	4.9%	1.5%	6.8%	1.1%
Age 16	16.7%	6.7%	6.7%	3.7%	4.8%	2.2%
Age 17	13.8%	6.9%	5.6%	2.7%	4.5%	1.5%
Age 18	8.8%	2.8%	3.1%	0.4%	1.1%	0.4%
Age 19	14.7%	5.3%	4.6%	1.6%	2.3%	1.2%
Age 21	13.7%	6.2%	5.2%	0.5%	4.4%	0.5%
Overall	44.0%	19.9%	21.7%	7.1%	18.3%	5.5%

Note. Percentages depict the proportion of boys and girls who endorsed suicidal ideation, plans, and attempts at each age.

Table S2.11
Odds ratio results from survival analyses with gender as a moderator

Ideation					
	Girls Odds Ratio [95% CI]	Boys Odds Ratio [95% CI]	Wald Test	AIC (free) / AIC (constrained)	BIC (free) / BIC (constrained)
Effortful control	0.50* [0.39, 0.63]	0.51* [0.32, 0.82]	0.01 (*p* = .912)	3927 / 3925	4020 / 4014
Negative emotionality	1.95* [1.60, 2.39]	1.96 [1.28, 3.01]	0.00 (*p* = .988)	3918 / 3916	4012 / 4006
Negative emotionality (without Depressed Mood & Aggression)	1.44 [1.15, 1.82]	1.41 [0.89, 2.23]	0.01 (*p* = .927)	3969 / 3967	4063 / 4056
Positive emotionality	1.17 [0.90, 1.51]	1.17 [0.83, 1.63]	0.00 (*p* = .996)	3982 / 3980	4075 / 4069
Plans					
	Girls Odds Ratio [95% CI]	Boys Odds Ratio [95% CI]	Wald Test	AIC (free) / AIC (constrained)	BIC (free) / BIC (constrained)
Effortful control	0.54* [0.38, 0.76]	0.45* [0.23, 0.88]	0.22 (*p* = .638)	2324 / 2323	2418 / 2412
Negative emotionality	1.71* [1.32, 2.22]	2.23 [1.28, 3.90]	0.73 (*p* = .394)	2322 / 2321	2416 / 2410
Negative emotionality (without Depressed Mood & Aggression)	1.23 [0.87, 1.73]	1.83 [1.03, 3.27]	1.37 (*p* = .241)	2342 / 2342	2436 / 2431
Positive emotionality	1.35 [0.95, 1.93]	0.80 [0.51, 1.24]	3.33 (*p* = .068)	2344 / 2346	2438 / 2435
Attempts					
	Girls Odds Ratio [95% CI]	Boys Odds Ratio [95% CI]	Wald Test	AIC (free) / AIC (constrained)	BIC (free) / BIC (constrained)
Effortful control	0.55* [0.38, 0.79]	0.62 [0.31, 1.24]	0.09 (*p* = .759)	2207 / 2205	2300 / 2294
Negative emotionality	1.45 [1.11, 1.89]	1.82 [0.94, 3.52]	0.39 (*p* = .534)	2212 / 2210	2305 / 2300
Negative emotionality (without	0.98 [0.69, 1.40]	1.66 [0.77, 3.61]	1.46 (*p* = .227)	2220 / 2221	2314 / 2310

Depressed Mood & Aggression)					
Positive emotionality	1.24 [0.84, 1.84]	0.87 [0.57, 1.32]	1.53 (*p* =.216)	2221 / 2220	2315 / 2310

Note. * $p < .005$

Table S2.12
Prevalence of suicidal ideation, plans, and attempts separated by nativity status.

	Ideation		Plans		Attempts	
	Born in Mexico	Born in U.S.	Born in Mexico	Born in U.S.	Born in Mexico	Born in U.S.
Age 12	6.5%	5.7%	1.9%	1.1%	1.3%	1.6%
Age 13	6.0%	5.6%	0.7%	1.3%	2.7%	1.6%
Age 14	6.5%	13.0%	4.5%	4.9%	3.2%	4.7%
Age 15	7.8%	9.4%	1.3%	4.0%	3.9%	4.0%
Age 16	8.5%	13.1%	5.9%	5.0%	3.9%	3.4%
Age 17	6.8%	11.9%	4.8%	4.0%	4.8%	2.4%
Age 18	4.3%	6.5%	1.4%	1.9%	0.0%	1.1%
Age 19	7.6%	11.4%	2.8%	3.4%	1.4%	2.0%
Age 21	13.4%	9.1%	3.1%	3.0%	2.3%	2.7%
Overall	28.3%	33.7%	12.8%	15.2%	10.0%	12.9%

Note. Percentages depict the proportion of boys and girls who endorsed suicidal ideation, plans, and attempts at each age.

Table S2.13

Odds ratio results from survival analyses with nativity as a moderator

	Born in Mexico Odds Ratio [95% CI]	Born in U.S. Odds Ratio [95% CI]	Wald Test	AIC (free) / AIC (constrained)	BIC (free) / BIC (constrained)
		Ideation			
Effortful control	0.52* [0.27, 0.98]	0.56* [0.45, 0.69]	0.04 (*p* = .840)	3902 / 3900	3995 / 3989
Negative emotionality	2.28* [1.58, 3.29]	2.08* [1.70, 2.55]	0.17 (*p* = .680)	3858 / 3856	3952 / 3945
Negative emotionality (without Depressed Mood & Aggression)	1.71 [1.13, 2.58]	1.62* [1.29, 2.02]	0.05 (*p* = .818)	3916 / 3914	4009 / 4003
Positive emotionality	1.39 [0.98, 1.98]	1.04 [0.80, 1.33]	1.76 (*p* = .185)	3945 / 3945	4038 / 4034
		Plans			
	Born in Mexico Odds Ratio [95% CI]	Born in U.S. Odds Ratio [95% CI]	Wald Test	AIC (free) / AIC (constrained)	BIC (free) / BIC (constrained)
Effortful control	0.59 [0.25, 1.36]	0.53* [0.40, 0.72]	0.04 (*p* = .841)	2254 / 2252	2347 / 2341
Negative emotionality	2.01 [1.26, 3.23]	1.96* [1.51, 2.54]	0.01 (*p* = .918)	2240 / 2238	2333 / 2327
Negative emotionality (without Depressed Mood & Aggression)	1.46 [0.89, 2.41]	1.55 [1.11, 2.17]	0.04 (*p* = .844)	2264 / 2262	2357 / 2351
Positive emotionality	1.39 [0.90, 2.14]	1.06 [0.73, 1.55]	0.83 (*p* = .362)	2273 / 2272	2366 / 2361
		Attempts			
	Born in Mexico Odds Ratio [95% CI]	Born in U.S. Odds Ratio [95% CI]	Wald Test	AIC (free) / AIC (constrained)	BIC (free) / BIC (constrained)
Effortful control	0.89 [0.44, 1.79]	0.55* [0.39, 0.77]	1.44 (*p* = .231)	2139 / 2139	2233 / 2228
Negative emotionality	1.56 [0.92, 2.63]	1.71* [1.31, 2.24]	0.10 (*p* = .751)	2134 / 2134	2229 / 2223
Negative emotionality	1.40 [0.64, 3.06]	1.27 [0.91, 1.76]	0.05 (*p* = .822)	2151 / 2149	2244 / 2238

(without Depressed Mood & Aggression)					
Positive emotionality	1.23 [0.71, 2.12]	1.06 [0.71, 1.57]	0.19 (p = .663)	2153 / 2152	2247 / 2241

Note. * $p < .005$

Chapter 3

Developmental Trajectories of Temperament from Late Childhood through Adolescence and Associations with Anxiety and Depression in Young Adulthood

The content of this chapter is currently under review at the *European Journal of Personality*. Below is the citation for the corresponding manuscript.

Cite: Lawson, K. M., Bleidorn, W., Hopwood, C. J., Cheng, R., & Robins, R. W. (invited revision under review). Developmental trajectories of temperament from late childhood through adolescence and associations with anxiety and depression in young adulthood. *European Journal of Personality.*

Abstract

Anxiety and depression are pervasive and pernicious mental health problems for young adults. Developmental trajectories of adolescent temperament (Effortful Control, Negative Emotionality, Positive Emotionality) may help us predict who will experience anxiety/depression during young adulthood. The present study uses longitudinal data from a large, community sample of Mexican-origin youth ($N = 674$), to examine how temperament develops across adolescence (age 10 to 16) and whether the developmental trajectories of temperament are associated with anxiety/depression during young adulthood (ages 19 and 21). Results indicate that Effortful Control, Negatively Emotionality, and the Affiliation facet of Positive Emotionality tend to decrease across adolescence, whereas Surgency tends to increase. Greater increases in Effortful Control and Positive Emotionality across adolescence are associated with fewer anxiety/depression symptoms during young adulthood, whereas greater increases in Negative Emotionality are associated with more anxiety/depression symptoms later on. Thus, temperament development serves as both a protective factor (Effortful Control, Positive Emotionality) and a risk factor (Negative Emotionality) for later anxiety/depression in Mexican-origin youth.

Keywords: temperament, development, anxiety, depression, adolescence

Introduction

Anxiety affects about one in seven and depression one in 10 American young adults (Goodwin et al., 2020; Mojtabai et al., 2016). In addition to the immense personal suffering associated with anxiety and depression, experiencing these internalizing problems during young adulthood contributes to various negative outcomes both concurrently and prospectively (Naicker et al., 2013), including lower graduation rates (Breslau et al., 2008) and higher levels of unemployment (Kawakami et al., 2012). For Latino youth, and Mexican-origin youth in particular, dealing with acculturative stress and discrimination can lead to higher rates of anxiety and depression symptoms than youth who are members of other ethnic groups (Bridges et al., 2021; Maiya et al., 2021; Stein et al., 2019; Suarez-Morales & Lopez, 2009; Torres, 2010). Given this widespread public health problem, it is important to identify factors that contribute to anxiety and depression symptoms during young adulthood, which may then highlight potential prevention and intervention targets.

One important factor to consider is adolescent temperament (Compas et al., 2004). Numerous studies have examined the relation between temperament and anxiety/depression, but most of this work relies on cross-sectional data or examines how temperament assessed at one timepoint predicts subsequent levels of anxiety and depression. Thus, past research provides relatively little insight into how *changes* in temperament might impact anxiety and depression. Moreover, the vast majority of past research in this area has focused on samples comprised of predominantly White American and European youth. The present study used data from a long-term, longitudinal study of Mexican-origin youth living in the United States ($N = 674$) to examine the developmental trajectories of three temperament domains – Effortful Control, Negative Emotionality, and Positive Emotionality – from late childhood (age 10) through adolescence (age 16), and their associations with anxiety and depression at ages 19 and 21. Importantly, we addressed the overlap between anxiety and depression by examining how temperament trajectories are associated with symptoms they share in common (generalized distress), as well as distinct aspects of anxiety (anxious arousal) and depression (anhedonia).

Temperament

Temperament refers to individual differences in reactivity and self-regulation that are present from an early age and relatively enduring (Rothbart, 2007; Rothbart, 2011). Researchers studying individual differences during adolescence often label traits as either "temperament" or "personality", though there is no clear conceptual or empirical distinction between the two (Clark & Watson, 2008; Shiner, 2015; Shiner & DeYoung, 2013; Shiner et al., 2021; Soto & Tackett, 2015). Research on youth temperament is often guided by Rothbart's highly influential model (Rothbart et al., 2000), which posits three key constructs: Effortful Control (EC), Negative Emotionality (NEM), and Positive Emotionality (PEM). According to Rothbart's model, the EC domain involves one's capacity to plan and suppress inappropriate impulses (Inhibitory Control), perform an action or pursue goals when there are competing desires (Activation Control), and focus and shift attention when needed (Attention). The NEM domain involves unpleasant affect derived from anticipating distress (Fear) and negative affect related to ongoing tasks being interrupted (Frustration). Often, the NEM domain is expanded to include hostile reactivity to negative stimuli including person- and object-directed violence (Aggression) and unpleasant affect, lowered mood, and lack of enjoyment in activities (Depressed Mood). The PEM domain involves pleasure derived from high intensity or novel activities (High Intensity Pleasure/Surgency; hereafter referred to as Surgency) and a desire for close, warm interpersonal connections (Affiliation). Factor analytic work, including research using the same dataset as the present study (Lawson et al., 2021), suggests that EC and NEM represent broad but cohesive domains, whereas the PEM facets do not form a cohesive domain (Putnam, 2001; Snyder et al., 2015). Consequently, we consider both domain and facet-level scores for EC and NEM but only facet-level scores (i.e., separate Surgency and Affiliation) for PEM.

Despite the conceptual similarity between temperament and personality traits, studies typically focus on either one or the other. Fortunately, prior research has connected Rothbart's temperament domains to Big Five trait domains, allowing some inferences to be made about the development and prospective effects of temperament from studies of personality traits. In particular, past theoretical and empirical work has found that EC relates to Conscientiousness, NEM is closely associated with Neuroticism, the Surgency facet of PEM shares similarities with Extraversion, and the Affiliation facet of PEM is akin to Agreeableness (Evans & Rothbart, 2007; Putnam, 2001; Rothbart, 2007; Shiner & Caspi, 2003). Altogether, a substantial body of work has examined the development of temperament and personality traits from childhood through adolescence. Below, we organize our review of relevant research using Rothbart's three-factor temperament framework, but include studies based on other temperament and personality frameworks, such as the Big Five (see Table 3.1 for a summary of cited studies).

Developmental Changes in Temperament across Adolescence

Adolescence is a developmental period characterized by rapid maturational changes, shifting societal expectations, conflicting role demands, and increasingly complex peer and romantic relationships (Hill & Edmonds, 2017). Many have speculated that these dramatic biological, cognitive, and social changes may be matched by accompanying changes in temperament. Most notably, researchers have suggested that adolescent personality development follows the *disruption hypothesis*, where "biological, social and psychological transitions from childhood to adolescence are accompanied by temporary dips in some aspects of personality maturity" (Soto & Tackett, 2015, p. 360). In particular, youth may struggle to meet the increased demands and challenges of adolescence and become (temporarily) less responsible, kind, and emotionally stable during this process. In contrast, other researchers have posited that adolescents rise to the challenge of this tumultuous developmental period, raising their level of personality maturation by becoming more well-regulated, prosocial, and emotionally stable (Bleidorn et al., 2013; Roberts et al. 2006). In particular, the *maturity principle* assumes that normative personality development equips adolescents to engage in relevant educational, occupational, and interpersonal domains. However, meta-analytic findings suggest that there are only minor mean-level changes in traits from age 10 to 18, especially when compared to changes that occur during young adulthood (Bleidorn et al., 2021; Roberts et al., 2006), suggesting that neither disruption nor maturation captures the whole picture. Thus, debate persists about the normative development of temperament traits during adolescence.

Effortful Control

Several longitudinal studies have examined developmental trajectories of EC and related constructs (e.g., Conscientiousness) across adolescence. Most studies find mean-level *decreases* in both self- and parent-reported EC (De Fruyt et al., 2006; Göllner et al., 2017; Laceulle et al., 2012; van den Akker et al., 2014; Zohar et al., 2019). However, some studies have found *no mean-level changes* in EC (Göllner et al., 2017; Klimstra et al., 2009; Pullmann et al., 2006; Roberts et al., 2006) and other studies even found mean-level *increases* in Conscientiousness, but only for girls (Borghuis et al., 2017; Brandes et al., 2022; Branje et al., 2007). Despite these discrepancies, the majority of evidence supports the *disruption hypothesis* with respect to EC across adolescence. Indeed, research using data from the same sample as the present study has documented mean-level decreases in EC from age 10 to 14 followed by mean-level increases from age 14 to 19 (Atherton et al., 2020).

Negative Emotionality

Previous research has found mean-level *decreases* in NEM (and Neuroticism) during adolescence using both self- and parent-reports (Brandes et al., 2022; De Fruyt et al., 2006; Klimstra et al., 2009; Laceulle et al., 2012; Pullmann et al., 2006; Roberts et al., 2006; van den Akker et al., 2014; Zohar et al., 2019; but see Borghuis et al., 2017, Branje et al., 2007, & Göllner et al., 2017, who largely found no changes). These findings support the *maturity principle* of personality development because they suggest that youth are becoming more emotionally stable as they get older. However, research on the facets of NEM suggests a more complicated pattern; although the core facets of Fear and Frustration seem to decrease during adolescence (Brandes et al., 2022), the Depressed Mood facet tends to increase, especially for girls (Soto et al., 2011).

Positive Emotionality

Findings on mean-level changes in PEM during adolescence are mixed. For Surgency (and the related construct of Extraversion), some longitudinal studies have found mean-level *increases* in both self- and parent-reports (Göllner et al., 2017; Klimstra et al., 2009; Laceulle et al., 2012; Pullmann et al., 2006; Roberts et al., 2006; Zohar et al., 2018), whereas others have found *no mean-level changes* (Borghuis et al., 2017; De Fruyt et al., 2006) or even mean-level *decreases* in Surgency across this period (Branje et al., 2007; Brandes et al., 2022; van den Akker et al., 2014). The longitudinal evidence is similarly mixed for the other facet of PEM, Affiliation. Some studies have found mean-level *increases* in Affiliation (Borghuis et al., 2017; Brandes et al., 2020; Branje et al., 2007; Klimstra et al., 2009), whereas others have found mean-level *decreases* (Göllner et al., 2017; Laceulle et al., 2012; van den Akker et al., 2014) or *no mean-level changes* across adolescence (De Fruyt et al., 2006; Pullmann et al., 2006; Roberts et al., 2006).

Together, previous studies suggest normative decreases in EC (aligned with the *disruption hypothesis*), normative decreases in NEM (aligned with the *maturity principle*), and mixed evidence with respect to PEM development across adolescence. However, the empirical basis for these findings is relatively weak, given that many studies used a single informant to assess temperament, included only a few measurement occasions, examined only domain and not facet-level changes, and/or followed youth across only part of the adolescence. Thus, more research is needed to better understand the trajectory of temperament across the adolescent years. Next, we review past research and theory on the association between temperament trajectories and anxiety/depression symptoms in young adulthood.

Table 3.1

Summary of Previous Research Examining Temperament and Personality Development during Adolescence

Study	Measure	Assessment method	Language	Age range	Measurement occasions	Sample size	Summary of findings[d]
Borghuis et al. (2017)	QBF	Self-report	Dutch	12 to 22	7 times across 11 years	*N* = 2,230	Increases in A (all), C (girls only); Dips in E (girls only), C (boys only), ES (girls only); No change in E or ES (boys only)
Branje et al. (2007)	QBF	Self-report, parent-report	Dutch	11 to 17	3 times across 3 years	*N = 288*	Increases in A, C; Decreases in E; No change in N
Brandes et al. (2021)	ICID	Parent-report	English	10 to 13	4 times across 4 years	*N* = 440	Increases in A (all), C (girls only); Decreases in E (all), N (boys only)
De Fruyt et al. (2006)[a]	HiPIC	Parent-report	Dutch	5 to 14	2 times across 1.5 years	*N* = 1,046	Decreases in C, N; No change in E, A
Göllner et al. (2017)[b]	BFI[a]	Self-report, parent-report	German	10 to 13	4 times across 4 years	*N* = 2,761	Increases in E; Decreases in A, C; No change in N
Klimstra et al. (2009)[c]	QBF	Self-report	Dutch	12 to 16	5 times across 5 years	*N* = 923	Increases in E (boys only), A (all); Decreases in N (boys only); No change in C (all)

Laceulle et al. (2012)	EATQ-R	Parent-report	Dutch	11 to 16	2 times across 6 years	$N = 1{,}197$	Increases in Surgency; Decreases in EC, NEM, Affiliation
Pullman et al. (2006)	NEO-FFI	Self-report	Estonian	12 to 18	6 times across 7 years	$N = 876$	Increases in E; Decreases in N; No changes in A, C
van den Akker et al. (2014)	HiPIC	Self-report, parent-report	Flemish	6 to 20	5 times across 15 years	$N = 596$	Dips in A, C; Decreases in E, N
Zohar et al. (2018)	JTCI	Self-report	English	12 to 16	3 times across 5 years	$N = 752$	Increases in Surgency; Decreases in EC, NEM[e]

Note. Laceulle et al. (2012) and Zohar et al. (2018) used a measure of temperament; all other studies used a measure of personality. Previous studies suggest at least moderate convergence among many of these measures (e.g., Tackett et al., 2013). BFI = Big Five Inventory. EATQ-R = Early Adolescent Temperament Questionnaire-Revised. HiPIC = Hierarchical Personality Inventory for Children. ICID = Inventory of Child Individual Differences. JTCI = Junior Temperament and Character Inventory. NEO-FFI = NEO Five-Factor Inventory. QBF = Goldberg's Big Five Questionnaire. ES = Emotional Stability, the opposite pole of Neuroticism. Dips imply U-shaped mean-level change.

[a] DeFruyt et al. (2006) used data from two samples (a representative community sample and a twin sample); both samples are combined in this table.

[b] For Göllner et al. (2017), youth completed the full BFI and their parents completed a short (10-item) informant report version.

[c] Klimstra et al. (2009) had an older adolescent cohort (i.e., 17-years-old at initial assessment) and details are only reported for the younger adolescent cohort.

[d] Whenever possible, summarized findings are collapsed across gender and assessment method. For studies with both self- and parent-reports, findings for self-reported traits are presented in the table. Please see the original studies for more details.

[e] In Zohar et al. (2018), Surgency maps onto their measure of Novelty Seeking and Reward Dependence; EC maps onto Self-directedness and Persistence; NEM maps onto Harm Avoidance.

Temperament as a Risk Factor for Anxiety and Depression

The vulnerability model of personality and psychopathology posits that certain temperament traits place an adolescent at greater risk for, or protect them from, developing later psychopathology (Tackett, 2006). For example, youth who have lower attention control early in life may be predisposed to depression later on, whereas youth who are more fearful early on may later experience more anxiety symptoms (Muris & Ollendick, 2005). Consistent with the vulnerability model, previous theory and research supports the idea that certain traits serve as risk or protective factors for adolescent anxiety and depression (De Pauw & Mervielde, 2010; Nigg, 2006; Pérez-Edgar & Fox, 2005).

Effortful Control

Lower levels of EC, especially the Attention facet, are conceptually related to higher levels of depression (Compas et al., 2004; Muris et al., 2007) and cross-sectional studies have confirmed these associations (e.g., Oldehinkel, 2004). Longitudinally, lower levels of EC during adolescence may predispose individuals to experience depression during young adulthood. For example, youth who are lower in EC are likely to have difficulty regulating their behaviors and emotions, so they may be more likely to engage in rumination and have trouble disrupting maladaptive cycles or following through with problem-solving tactics related to depressive thoughts (Aldao et al., 2010; Van Beveren et al., 2016). These dysfunctional processes may contribute to problems in peer and romantic relationships, academic failure, and other negative life outcomes that can trigger depression. Consistent with these ideas, several two-wave studies found associations between lower levels of EC and higher levels of depression one to four years later (Crockett et al., 2013; Loukas & Roalson, 2006; Ormel et al., 2005; Verstraeten et al., 2009). However, in two of these studies, the effect of EC was no longer significant after controlling for depression at the first wave (Loukas & Roalson, 2006; Verstraeten et al., 2009). In contrast to depression, there is little theoretical justification for a relation between EC and anxiety, and past research has rarely examined these specific associations, instead grouping internalizing symptoms together (e.g., Crockett et al., 2013). In summary, there is limited evidence that EC predisposes youth to developing depression symptoms later on and no evidence that EC contributes to anxiety problems.

Negative Emotionality

Numerous studies have shown that higher levels of NEM are concurrently related to both anxiety and depression (e.g., Anthony et al., 2002). Developmentally, the presence of NEM in an adolescent's thoughts, feelings, and behaviors may contribute to an increased risk of experiencing adverse life events, which then increases risk of anxiety and depression in response to these negative events (Klein et al., 2011). More specifically, youth who have tendencies toward fear and frustration are likely to engage in maladaptive avoidance of their problems, suppressing thoughts and feelings rather than adaptively coping with them (Van Beveren et al., 2016). Further, tendencies toward sadness may make youth less likely to attempt to reappraise situations to try to see the positive, leading to downward spirals. Indeed, empirical studies have documented that NEM is longitudinally associated with later internalizing problems. Research suggests that prior levels of both self- and parent-reported NEM predispose youth to later anxiety and depression (Bould et al., 2014; Bouma et al., 2008; Brendgen et al., 2005; Caspi et al., 1996; Crockett et al., 2013; Davis et al., 2015; Lonigan et al., 2003; Mezulis et al., 2011; Ormel et al., 2005; Verstraeten et al., 2009; Wetter & Hankin, 2009). However, interpreting these findings is complicated by the conceptual overlap between measures of NEM and measures of anxiety and depression, making it important to explore findings at the facet level to determine whether the associations at the superordinate level are driven entirely by conceptually analogous scales (e.g., Fear with anxiety; Depressed Mood with depression) that have some content overlap.

Positive Emotionality

Low levels of Surgency, indicating the absence of feelings of high energy positive affect (e.g., enthusiasm, excitement), is conceptually similar to anhedonia, a core feature of depression. Indeed, Surgency is concurrently negatively associated with depression, but not anxiety (e.g., Phillips et al., 2002). Some longitudinal studies suggest that lower self- and parent-reported Surgency are prospectively associated with increased risk of developing depression and anxiety (Lonigan et al., 2003; Ormel et al., 2005; Verstraeten et al., 2009), but other studies found no prospective association (Bould et al., 2014; Mezulis et al., 2011).

The association between Affiliation and internalizing problems is even less clear. One theory is that youth who are higher in Affiliation, and thus desire close connections with others, may be more sensitive to interpersonal stressors and more anxious later on, especially during young adulthood when life circumstances could lead to difficulty establishing and maintaining consistent social support. The limited empirical research on the Affiliation facet of PEM suggests that there is no association between self- and parent-reported Affiliation and internalizing problems, either concurrently (Oldehinkel, 2004) or two to three years later (Ormel et al., 2005). However, one study found that higher levels of Affiliation were associated concurrently with higher levels of anxiety, but not depression (Kushner et al., 2012). Altogether, little is known about the prospective effects of Surgency and Affiliation on later anxiety and depression.

Temperament Change

Above and beyond initial or average levels, *changes* in temperament across adolescence may be especially relevant for adjustment later on. In particular, youth who are experiencing disruption (vs. maturation) with respect to various temperament traits during adolescence may struggle more with their mental health in young adulthood. For example, an adolescent who is increasing in their Fear, Frustration, and Depressed Mood and decreasing in EC may have more difficulties coping with the many challenges of adolescence, selecting themselves into more dysfunctional environments (e.g., choosing friends and romantic partners who are prone to conflict), and evocating changes in and reacting to their environment in ways that further increase their NEM over time and lead to more negative life outcomes. Together, these transactional processes could create a vicious cycle of declining levels of functioning, contributing to the development of anxiety and depression (Hopwood et al., 2022; Klein et al., 2011).

The handful of previous studies that have examined how changes in temperament relate to later internalizing problems have found results consistent with this pattern. In a three-year longitudinal study of 190 children aged 8- to 12-years-old at baseline, youth who experienced smaller decreases in EC and smaller increases in fear and irritability (assessed via the EATQ) had fewer internalizing problems at the final assessment (Lengua, 2006). Further, a longitudinal study of 290 8- to 9-year-old Belgian children followed across six years found that child personality trajectories, measured via parent- and teacher-reports on the HiPIC, were related to self- and parent-reported internalizing problems at the final assessment (van den Akker et al., 2010). In particular, youth who had lower Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness intercepts *and* those who became less extraverted and emotionally stable over time (i.e., those with more negative slopes) experienced more internalizing problems later on. A third study of 1,195 Dutch adolescents assessed at ages 11, 16, and 19 found that temperament development (assessed via the EATQ-R) from age 11 to 16 predicted internalizing problems at age 19 (Laceulle et al., 2014). In particular, youth who showed smaller decreases or greater increases than their peers in Fear and Frustration, respectively, were more likely to exhibit internalizing symptoms between age 16 to 19. Therefore, prior research indicates that not only initial levels, but also changes in temperament across adolescence, may play a role in the development of internalizing problems.

Summary

Together, these findings suggest robust concurrent and prospective associations between temperament traits and anxiety/depression, but they also highlight numerous gaps and inconsistencies in previous literature. Indeed, despite the many strengths of past research, most longitudinal support for the vulnerability model relies on studies with only one or two waves of temperament data, leaving open the question of whether *changes* in temperament might predispose youth to later anxiety and depression. The three studies that examined changes in temperament and later anxiety/depression (Lengua, 2006; Laceulle et al., 2014; van den Akker et al., 2010) had relatively few measurement occasions and/or covered only part of adolescence, leaving open questions about the precise trajectory of temperament across adolescence and association with internalizing problems in young adulthood. Further, with one exception (Loukas & Roalson, 2006), the vast majority of previous work uses samples of predominantly White American and European youth, limiting the generalizability of these findings to other racial and ethnic groups. Moreover, most of these studies examined only depression, and those that examined both rarely attempted to disentangle the two to examine how temperament relates to their unique and common components. This is important given the well-established comorbidity between anxiety and depression, and the likelihood that certain temperament traits might be distinctly associated with anxiety or depression, but not both. Notably, one study that did examine anxiety and depression separately found that high NEM was associated with both, whereas high Affiliation was uniquely associated with anxiety and low EC uniquely associated with depression (Kushner et al., 2012). This suggests that the tripartite model of anxiety and depression (Anderson & Hope, 2008; Clark & Watson, 1991), which posits that there are both overlapping (i.e., general distress) and distinct aspects of anxiety (i.e., anxious arousal) and depression (i.e., anhedonia), may be an especially useful framework for understanding the temperamental precursors of anxiety and depression.

The Present Study

The present study used data from the California Families Project (CFP), an ongoing longitudinal study of 674 Mexican-origin youth and their parents, to examine the developmental trajectories of EC, NEM, and PEM from late childhood (age 10) through adolescence (age 16), and their associations with anxiety and depression in young adulthood (ages 19 and 21). In particular, we examined the following research questions:

1. What is the average developmental trajectory of EC, NEM, and PEM, and their respective facets from age 10 to 16?
2. Are the EC, NEM, and PEM trajectories (i.e., level and slope) associated with anxiety and depression at ages 19 and 21? Further, do these associations replicate for (a) the facets of each temperament domain, (b) age 19 and age 21 anxiety/depression, and (c) different components of anxiety/depression, including *general distress* (symptoms related to anxiety & depression), *anxious arousal* (symptoms unique to anxiety), and *anhedonic depression* (symptoms unique to depression)? Further, do the associations hold when controlling for (d) prior levels of anxiety and depression (age 16) and (e) parent anxiety and depression? Parent anxiety/depression may serve as a common

cause and thus confound the relation between child temperament and anxiety/depression (Bould et al., 2014; Bouma et al., 2008; Degnan et al., 2010; Durbin et al., 2005; Ormel et al., 2005; Pérez-Edgar & Fox, 2005)

3. Is gender associated with the temperament trajectories and does gender moderate the associations between temperament development and anxiety/depression?[25] Prior research has documented gender differences in temperament development (Brandes et al., 2022; Borghuis et al., 2017; Göllner et al., 2017; Klimstra et al., 2009) and anxiety/depression (McLean et al., 2011; Nolen-Hoesksema, 2001). We examined gender as a moderator to explore the generalizability of our findings across boys and girls.

We made several predictions based on prior research and theory (Table 3.2).[26] With respect to research question 1, we expected to find (a) mean-level decreases in NEM and all of its facets (Fear, Frustration, Depressed Mood, Aggression), (b) mean-level increases in Surgency, and (c) mean-level decreases in Affiliation. We did not make hypotheses about the developmental trajectory of EC or its facets because prior research using the same dataset had already examined the development of the EC domain and found mean-level decreases across this period (Atherton et al., 2020; Damian et al., 2020).

With respect to research question 2, we predicted that: (a) individuals with lower initial levels and more negative slopes of EC (and each of its facets) would have higher levels of later general distress and anhedonic depression, but not anxious arousal; (b) individuals with higher initial levels and more positive slopes of NEM (and each of its facets) would have higher levels of later general distress, anxious arousal, and anhedonic depression; (c) individuals with lower initial levels and more negative slopes of Surgency would have higher levels of later general distress and anhedonic depression, but not anxious arousal; and (d) individuals with higher initial levels and more positive slopes of Affiliation would have higher levels of later general distress and anxious arousal, but not anhedonic depression. Additionally, we expected that the Fear facet of NEM would have the strongest association with later anxious arousal and the Depressed Mood facet of NEM have the strongest association with later anhedonic depression. We expected that all of the predicted associations would hold after controlling for prior levels of anxiety and depression and after controlling for parent anxiety and depression.

Finally, with respect to research question 3, we did not make specific predictions about the effects of gender on mean-level temperament trajectories across adolescence, or the potential moderating effect of gender on the associations between temperament and anxiety/depression. These analyses were purely exploratory.

Table 3.2
Summary of Hypotheses

Temperament	Expected pattern of mean-level change	Anxiety/Depression		
		General Distress	Anxious Arousal	Anhedonic Depression
Effortful Control	decrease[a]	-	0	-
Activation Control		-	0	-
Inhibitory Control		-	0	-
Attention		-	0	-
Negative Emotionality	decrease	+	+	+
Fear	decrease	+	+	+
Frustration	decrease	+	+	+
Depressed Mood	decrease	+	+	+
Aggression	decrease	+	+	+
Positive Emotionality				
Surgency	increase	-	0	-
Affiliation	decrease	+	+	0

Note. The three rightmost columns (i.e., those under Anxiety/Depression) indicate hypothesized associations with both the levels and slopes of the temperament trajectories.
[a] The results for the mean-level trajectory of the EC domain from age 10 to 16 have already been examined in a previous study using the same data (Damian et al., 2020). The mean-level trajectories of the EC facets (Activation

Control, Inhibitory Control, Attention) have been examined from age 10 to 19 (Atherton et al., 2020), but not from age 10 to 16.

The present study extends past research in several ways. First, our use of four waves of temperament data provides a more nuanced understanding of the development of temperament from late childhood through adolescence, including an examination of nonlinear trajectories. Second, our long-term, longitudinal data, spanning from 10- to 21-years-old, allows us to test the vulnerability model of temperament and psychopathology by examining prospective associations between adolescent temperament trajectories and young adult anxiety/depression. This builds on previous shorter-term longitudinal studies that found associations of EC and NEM with anxiety/depression. Third, we examined these associations in a sample of Mexican-origin youth, which contributes to the generalizability of previous work in this area that has focused on predominantly White adolescents. This is especially important given the higher rates of anxiety/depression in ethnic minority youth, which are due, at least in part, to experiences of acculturative stress and discrimination (Bridges et al., 2021; Maiya et al., 2021; Stein et al., 2019; Suarez-Morales & Lopez, 2009; Torres, 2010). Fourth, we used multimethod assessments of temperament (i.e., self- and parent-report), which helps capture temperament constructs more validly and adds to past longitudinal work that relied on only a single assessment method. Multimethod assessments also help alleviate concerns that associations between temperament and anxiety/depression are due solely to shared method variance (Compas et al., 2004; Klein et al., 2011; Wilson & Olino, 2021). Fifth, we used a well-validated measure of temperament to examine both domain and facet-level trajectories, which is important because mean-level changes in facets can be more heterogeneous than mean-level changes in domains (Brandes et al., 2022; Klimstra et al., 2018; Ormel et al., 2005; Schwaba et al., 2022). Finally, we used a measure of anxiety and depression based on the tripartite model, which allowed us to examine how temperament is associated with the common core of anxiety and depression (general distress), as well as with the unique aspects of anxiety (anxious arousal) and depression (anhedonia).

Methods

Participants

The present study used data from the California Families Project, a longitudinal study of Mexican-origin youth and their parents (N = 674).[27] Children were drawn at random from rosters of students from the Sacramento and Woodland, CA school districts, in 2006-07. The focal child had to be in the 5th grade, of Mexican origin, and living with their biological mother in order to participate in the study. Approximately 72.6% of the eligible families agreed to participate in the study, which was granted approval by the [ANONYMIZED] Institutional Review Board (Protocol # ANONYMIZED). The children (50% female) were assessed annually from age 10 to 19, and then two years later at age 21. The present study will use data from when the children were ages 10 (M_{age} = 10.86, SD = 0.50), 12 (M_{age} = 12.81, SD = 0.49), 14 (M_{age} = 14.75, SD = 0.49), 16 (M_{age} = 16.80, SD = 0.51), 19 (M_{age} = 19.86, SD = 0.52), and 21 years old (M_{age} = 21.74, SD = 0.73). Retention rates compared to the original sample are as follows: 86% (age 12), 90% (age 14), 89% (age 16), 87% (age 19), and 80% (age 21).

Participants were interviewed in their homes in Spanish or English, depending on their preference. Interviewers were all bilingual and most were of Mexican heritage. Sixty-three percent of mothers and 65% of fathers had less than a high school education (median = 9th grade for both mothers and fathers); median total household income was between $30,000 and $35,000 (overall range of income = < $5,000 to > $95,000). With regard to generational status, 83.6% of mothers and 89.4% of fathers were 1st generation, and 16.4% of mothers and 10.6% of fathers were either 2nd or 3rd generation.

Measures

Child Temperament

EC, NEM, and PEM were assessed at ages 10, 12, 14, and 16 via self- and mother-reports using the short version of the *Early Adolescent Temperament Questionnaire – Revised* (EATQ-R; Ellis & Rothbart, 2001). Ratings were made for each item on a 4-point scale ranging from 1 (*not at all true of you/your child*) to 4 (*very true of you/your child*). Descriptive statistics including means, standard deviations, and alpha and omega reliabilities for the observed variables are shown in Table S3.1. For all domains and facets, we created latent variables using parcels including both self- and mother-reported data.

Effortful control. The EC scale has three facets: Activation Control (5 items), Attention (6 items), and Inhibitory Control (5 items). Activation Control assesses the ability to perform an action or pursue goals when there are competing desires. Attention assesses the ability to focus and shift attention when needed. Inhibitory Control assesses the ability to plan and suppress inappropriate impulses.

Negative emotionality. The NEM scale has two facets: Fear (6 items) and Frustration (7 items). Fear assesses unpleasant affect derived from anticipating distress. Frustration assesses negative affect related to ongoing tasks being interrupted. Depressed Mood and Aggression also load onto the NEM domain. Aggression (6 items) assesses hostile reactivity to negative stimuli including person- and object-directed violence. Depressed Mood (6 items) assesses unpleasant affect, lowered mood, and lack of enjoyment in activities. To account for different scoring methods for the NEM domain, we ran analyses examining a broader (including Fear, Frustration, Aggression, and Depressed Mood) and a narrower (including only Fear and Frustration) conceptualization of NEM (Lawson et al., 2021).

Positive emotionality. The PEM scale has two facets: Surgency (6 or 14 items) and Affiliation (5 items).[28] Surgency assesses pleasure derived from activities involving high intensity or novelty. At age 10, the Surgency scale included the 6 items from the short version of the EATQ-R. Because of low reliability of the 6-item scale at age 10 (see Table S3.1), eight items were added from the full-length version of the EATQ-R Surgency scale at ages 12, 14, and 16. Given the low reliability at age 10, and the change in item content, we report main Surgency analyses using all of the items from age 12 to 16, and report analyses from age 10 to 16 in the Supplemental Materials. Affiliation assesses the desire for warmth and closeness with others. Surgency and Affiliation do not form a coherent superordinate factor (Lawson et al., 2021; Snyder et al., 2015); consequently, we did not compute a PEM domain score and instead examined these scales separately.

Child Anxiety and Depression

Child anxiety and depression symptoms were assessed at ages 16, 19, and 21 via self-report using the *Mini-Mood and Anxiety Symptom Questionnaire* (MASQ; Watson, Clark et al., 1995; Watson, Weber et al., 1995). This 26-item measure assesses three components of internalizing problems: general distress, anxious arousal, and anhedonic depression (Corral-Frías et al., 2019). General distress (8 items) includes symptoms that are experienced by both anxious and depressed individuals (e.g., “felt worthless”, “felt uneasy”). Anxious arousal (10 items) includes symptoms specific to anxiety, including somatic tension or physiological hyperarousal (e.g., “was short of breath”, “hands were shaky”). Anhedonic depression (8 items) includes symptoms specific to depression, including anhedonia or the absence of pleasurable experience (e.g., “felt like nothing was very enjoyable”, “felt really lively, up” – reverse-scored). Ratings were made on a 4-point scale ranging from 1 (*not at all*) to 4 (*very much*). We created separate latent variables for each component (i.e., general distress, anxious arousal, anhedonic depression) at each assessment (age 16, 19, and 21) and then saved the factor scores to use in all analyses. In particular, scores from ages 19 and 21 were used as outcomes in all analyses and scores from age 16 were used as a control variable in a subset of the analyses. In addition to examining scores from ages 19 and 21 separately, we also computed a composite “young adulthood” score for each of the components (i.e., general distress, anxious arousal, anhedonic depression), which is an average of scores from ages 19 and 21. Descriptive statistics for the three components at each age are shown in Table S3.2 and correlations among the components at each age are shown in Table S3.3.

Mother Anxiety and Depression

Maternal anxiety and depression symptoms were assessed via self-report when the children were 10-years-old using the MASQ (described above). Descriptive statistics are as follows: general distress (M = 1.62, SD = .58, α / ω = .88/.91); anxious arousal (M = 1.23, SD = .36, α / ω = .84/.89); anhedonic depression (M = 1.78, SD = .59, α / ω = .87/.91). We created separate latent variables for each component (i.e., general distress, anxious arousal, anhedonic depression) at each assessment and then saved the factor scores to use in all analyses.

Statistical Analyses

All data cleaning and analyses were conducted in R (R Core Team, 2019) via RStudio Version 1.2.1335 using full information maximum likelihood estimation (FIML) to account for missing data (Allison, 2003; Schafer & Graham, 2002). We used the following R packages to analyze the data and visualize the results: *psych* (Revelle, 2018), *lavaan* (Rosseel, 2012), *ggplot2*

(Wickham, 2016), *semTools* (Jorgensen et al., 2019), and *lm.beta* (Behrendt, 2014). Model fit was assessed via comparative fit index (CFI) and root-mean-square error of approximation (RMSEA). We interpreted good fit as values greater than or equal to .95 for CFI and less than or equal to .06 for RMSEA (Hu & Bentler, 1999). We assessed differences in model fit via change in CFI less than or equal to .01 (Chen, 2007) and change in chi-square and degrees of freedom (Cheung & Rensvold, 2002; Meade et al., 2006).

Latent Variables and Measurement Invariance

We created latent variables for each temperament domain and facet using both self- and mother-reported items to create parcels to use as indicators (the same items across self- and mother-reports were placed in the same parcels). Parcels typically produce more stable solutions, are less likely to share specific sources of variance, and reduce the likelihood of spurious correlations (Little et al., 2002). Indicators based on the same items were allowed to correlate across waves. We conducted tests of longitudinal measurement invariance for all domains and facets. In particular, we compared three measurement models: (a) freely estimating the factor loadings for the latent factors at each age of assessment (i.e., configural invariance); (b) constraining the respective factor loadings to be equal at each of assessment (i.e., weak invariance); and (c) constraining the factor loadings and intercepts to be equal at each age of assessment (i.e., strong invariance). If the more constrained models did not fit worse than the lesser constrained models, then we concluded that the structure of the latent construct was the same over time. Because we did not find evidence of strong measurement invariance for the majority of the domains/facets, we also examined evidence of partial strong measurement invariance (i.e., freeing the intercepts of one of the parcels from the strong invariance model).

We found evidence for partial strong longitudinal measurement invariance for all domains and facets, except Fear where we found strong invariance (Table S3.4).[29]

For child anxiety and depression, we created latent variables for each component of the MASQ (i.e., general distress, anxious arousal, anhedonic depression) using parcels created with self-reported items. Given that we were not interested in mean-level change in anxious and depressive tendencies, we did not test for longitudinal measurement invariance. We saved factor scores to use in the subsequent analyses examining associations between temperament development and anxiety/depression. We followed the same process for mother anxiety and depression when the children were 10-years-old. Factor loadings of parcels for all latent variables are shown in Table S3.5.

Research Question 1: Mean-Level Change in Temperament

To examine mean-level change in the temperament domains and facets from age 10 to 16, we ran univariate latent growth curve (LGC) models with four timepoints (ages 10, 12, 14, and 16). To guide our selection of a growth trajectory, we conducted a series of LGC model comparisons and evaluated changes in fit indices. Specifically, we compared three models: (1) no growth model, where only an intercept (no slope) was estimated; (2) linear growth model, where the slope increased linearly over time; and (3) latent basis model, where the first and last time points of the slopes were fixed to zero and six, respectively, and the middle time points were freely estimated in order to detect nonlinearities in the growth model.[30] We considered model fit, as well as parsimony, when selecting a LGC model for each domain and facet. For the best-fitting model, we tested for individual differences in the level and slope by examining the variance of the average level and slope. We then saved factor scores of the level/slope for each participant using the retained model for use in subsequent analyses, rather than attempting to fit latent variable models that include estimates of both growth in temperament and latent outcomes. This helped facilitate consistency across analyses by minimizing problems associated with model non-convergence for more complex analyses.

Research Question 2: Temperament and Anxiety/Depression

To examine whether individuals' developmental trajectories predicted their subsequent anxiety/depression, we tested whether the level and slope of each temperament domain/facet from the retained LGC model was significantly correlated with anxiety/depression at age 19, 21, and the young adult composite. In particular, we ran multiple regressions where we regressed the factor scores of each anxiety/depression component (i.e., general distress, anxious arousal, anhedonic depression) at each age on the factor scores of both the level and slope from the retained LGC model, separately for each temperament domain and facet. For these analyses, we set the alpha level to .006 (two-tailed) to account for multiple comparisons; we report exact *p*-values.[31] Then, for the significant associations, we examined whether the findings held when controlling for prior levels of anxiety/depression by adding the corresponding component from age 16 as a predictor in the multiple regression (e.g., when examining the association between EC and anhedonic depression at age 19, we included anhedonic depression at age 16). Next, we examined whether the original findings (i.e., not controlling for age 16 anxiety/depression) held when the corresponding mother's anxiety/depression was included as a control variable.

Research Question 3: Gender

Finally, we tested whether gender was associated with the temperament trajectories and whether gender moderated the associations between temperament development and anxiety/depression. To examine the association between gender (a dichotomized variable) and temperament trajectories, we ran multiple group models and compared models that constrained the means, variances, and covariance of the level and slope of girls and boys to be equal across groups to models that allowed these parameters to be freely estimated across groups. If the constrained model did not fit significantly worse than the freely estimated model, then we concluded that the developmental trajectory is the same across girls and boys. To examine whether gender moderated the association between temperament development and anxiety/depression, we included the main effect of gender in the regression model as well as the interaction term between gender and the level/slope of temperament (separately for the level and slope).[32]

Results

Mean-Level Change in Temperament

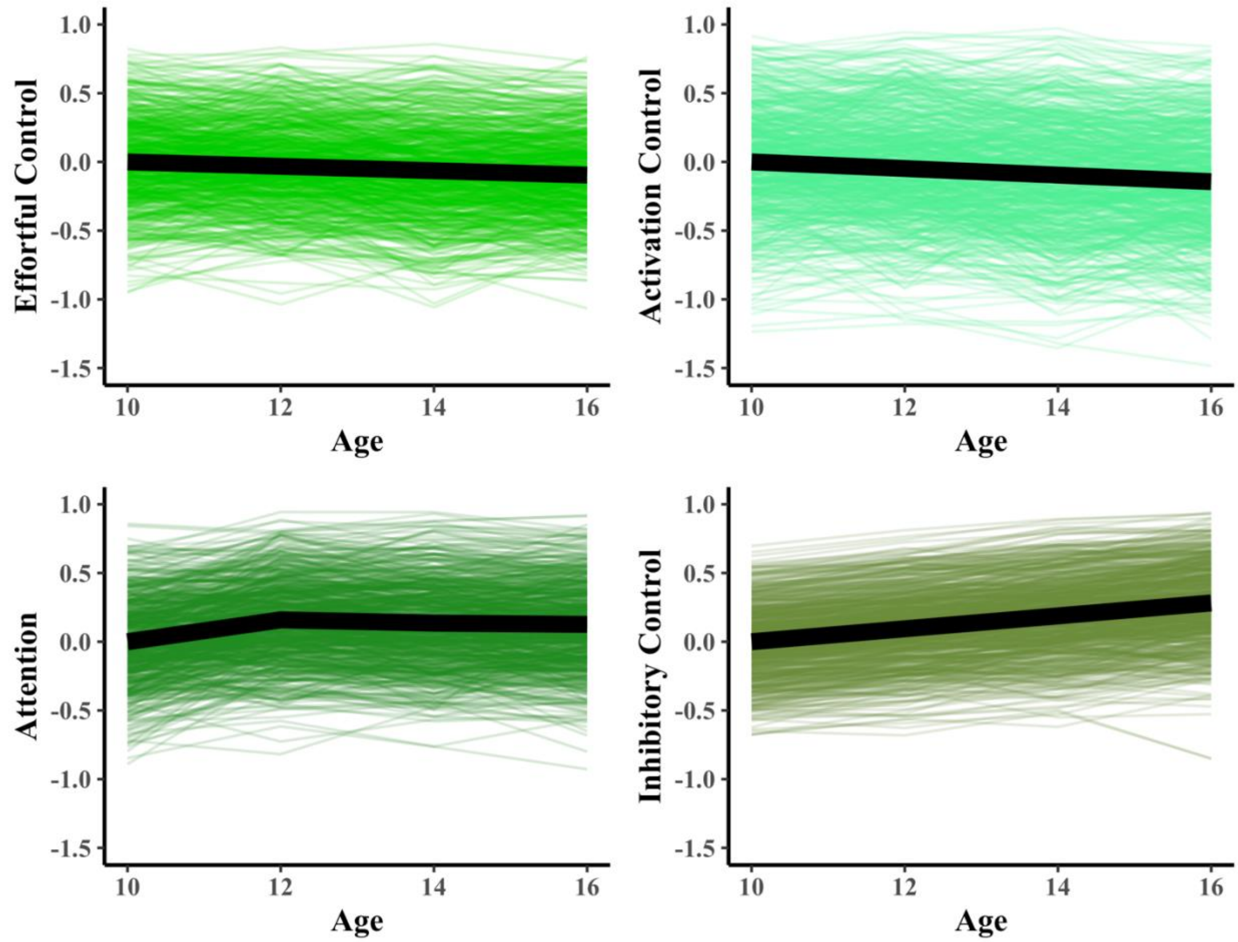
First, we examined mean-level change in all of the temperament domains and facets by comparing no growth, linear growth, and latent basis models (Table S3.6). Intercorrelations among the temperament trajectory levels and slopes are shown in Table S3.7.

Effortful Control

The individual and best-fitting average trajectories for the EC domain and facets are shown in Figure 3.1 (see Table S3.6 for model parameters). For both the EC domain and Activation Control facet, we found that youth tended to decrease linearly, on average, from age 10 to 16. For Attention, youth increased, on average, from age 10 to 12, followed by slight decreases from age 12 to 16. Finally, youth linearly increased, on average, in Inhibitory Control from age 10 to 16. There were substantial individual differences in the level and slope for EC and its facets, except for Attention and Inhibitory Control whose slope variances were not significant.

When we examined gender differences in the trajectories of EC and its facets, we found no significant differences in the means of the levels or slopes (see Table S3.8 for model fit comparisons of multiple-group models).[33]

Figure 3.1
Individual and Average Trajectories for the Effortful Control Domain and Facets

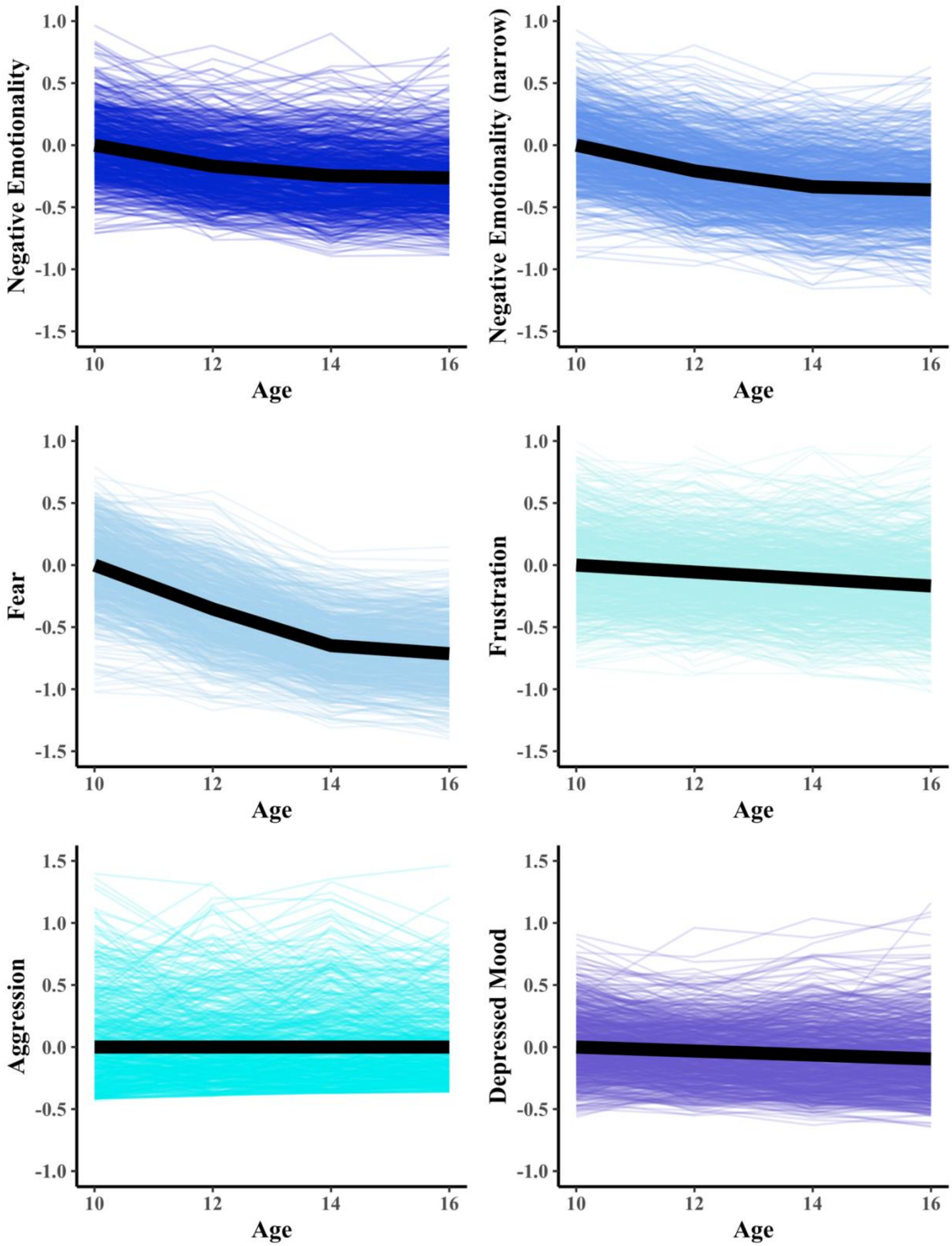


Negative Emotionality

The individual and best-fitting average trajectories for the NEM domain and facets are shown in Figure 3.2 (see Table S3.6 for model parameters). For both the broad and narrow NEM domains and the Fear facet, we found nonlinear mean-level decreases across adolescence, with the greatest decreases taking place from age 10 to 12. Both Frustration and Depressed Mood decreased linearly from age 10 to 16. Aggression – a facet where most of the youth scored very low at every assessment – showed no mean-level changes across adolescence. For NEM and all of its facets, there were substantial individual differences in the estimated levels and slopes. Additionally, none of the NEM trajectories differed for boys versus girls (Table S3.8).

Figure 3.2

Individual and Average Trajectories for the Negative Emotionality Domain and Facets

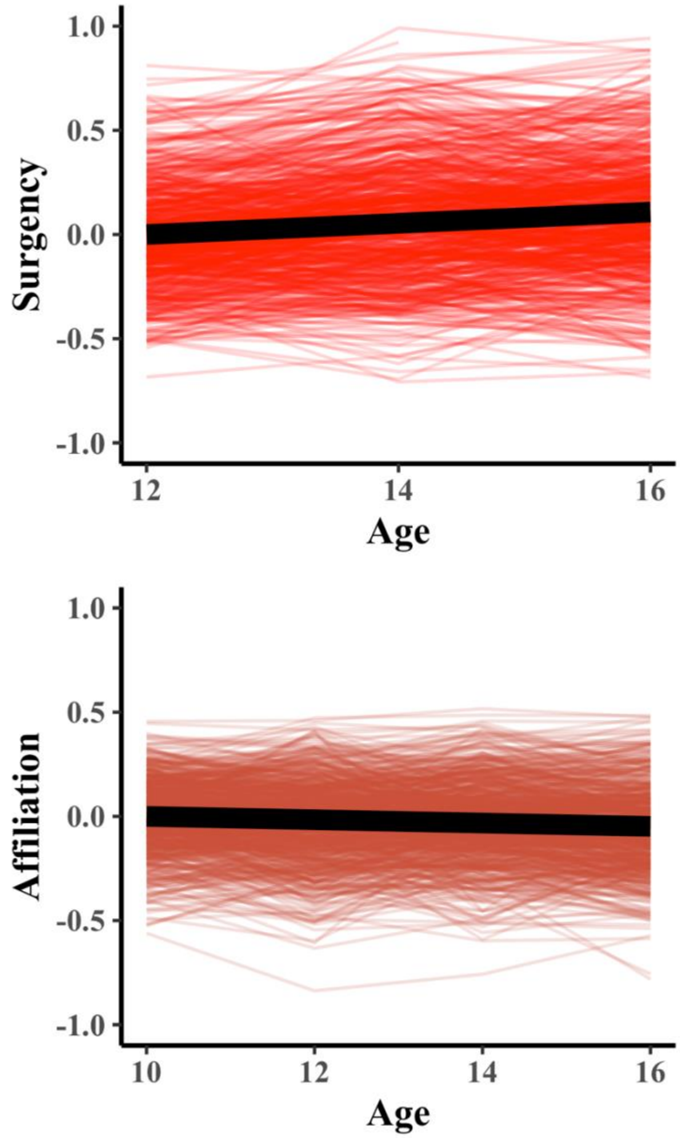


Positive Emotionality

The individual and best-fitting average trajectories for the PEM facets are shown in Figure 3.3 (see Table S3.6 for model parameters). For Surgency, we found that youth tended to increase from age 12 to 16.[34] Affiliation showed slight decreases from age 10 to 16. For both Surgency and Affiliation, there were substantial individual differences in the estimated levels and slopes. The Surgency and Affiliation trajectories did not differ significantly for boys versus girls (Table S3.8).

Figure 3.3

Individual and Average Trajectories for the Positive Emotionality Facets



Temperament Development and Anxiety/Depression

Next, we examined whether youths' individual temperament trajectories during adolescence were associated with their anxiety/depression symptoms during young adulthood. Table 3.2 shows the standardized coefficients from multiple regression analyses with the level and slope of each temperament domain/facet jointly predicting anxiety/depression at ages 19 and 21 (and their composite). General distress, anxious arousal, and anhedonic depression were each regressed separately on both the intercept and slope for each temperament domain/facet (i.e., intercepts and slopes pairings were included jointly). In most cases, the effect sizes were larger for the more proximal measure of anxiety/depression (age 19), although the vast majority of results at age 19 were in the same direction at age 21. Given this pattern, we focus below on the results for the composite scores, but present exact results at ages 19 and 21 in all tables.

Effortful Control

The slope of the EC domain was significantly negatively associated with general distress during young adulthood (β = -.15, p < .001), which suggests that youth who experienced greater decreases in EC from age 10 to 16 felt more worthless and uneasy later on. This association was largely driven by Activation Control (β = -.11, p = .005). For anhedonic depression, the slopes of the EC domain (β = -.18, p < .001) and all facets were significantly negatively correlated with unique depression symptoms during young adulthood. Neither the EC domain nor its facets were associated with anxious arousal during young adulthood. Notably, the slopes, but not the levels, of EC were most often associated with general distress and anhedonic depression symptoms during young adulthood. This suggests that it is the way an adolescent *develops* from age 10 to 16, rather than where their EC begins at age 10, that is more relevant to their later depression symptoms.

When we controlled for prior anxiety/depression symptoms (assessed at age 16) in the models, none of the associations between the EC trajectories and anxiety/depression remained significant using our preregistered alpha level of .006 (Table S3.9). This suggests that, although adolescent EC development is associated with later depression, it does not predict over and above prior symptoms of depression. Conversely, when we controlled for mother anxiety/depression symptoms assessed when the youth were 10-years-old, most (i.e., 75%) of the associations remained significant (Table S3.10). This suggests that adolescent EC development remains a robust predictor of young adulthood depression even when mother depression is taken into account.

Negative Emotionality

For the broad NEM domain, both the level ($\beta = .19$, $p < .001$) and slope ($\beta = .22$, $p < .001$) were positively associated with general distress during young adulthood, suggesting that youth who are higher in NEM at age 10 and those who experience greater increases in NEM from age 10 to 16 are more distressed later on. These associations were driven by the Fear, Frustration, and Depressed Mood facets. Similarly, both the level and slope of the broad NEM domain were positively associated with anxious arousal and anhedonic depression during young adulthood. These associations with specific anxiety and depression symptoms were largely driven by the Depressed Mood facet, and were not significant with the narrower assessment of NEM that only included the Fear and Frustration facets.

When we controlled for anxiety/depression symptoms at age 16, only the level of Depressed Mood remained a significant predictor of later general distress and anhedonic depression (Table S3.9). As was the case with EC, the vast majority of associations between NEM trajectories and anxiety/depression (i.e., 84%) remained significant when we controlled for mother anxiety/depression (Table S3.10).

Positive Emotionality

The slope of Surgency was negatively correlated with anhedonic depression during young adulthood ($\beta = -.13$, $p = .001$), such that youth who experience greater decreases in Surgency from age 12 to 16 report higher levels of anhedonia during young adulthood. For Affiliation, both the level ($\beta = -.14$, $p < .001$) and slope ($\beta = -.22$, $p < .001$) were negatively associated with anhedonic depression in young adulthood, indicating that youth with lower Affiliation at age 10 and those who experience greater decreases from age 10 to 16 have higher levels of anhedonia in young adulthood. The trajectories of Surgency and Affiliation were not significantly related to general distress or anxious arousal during young adulthood.

When we controlled for specific depression symptoms at age 16, the slopes of both Surgency and Affiliation remained significant predictors of anhedonic depression at age 19 (Table S3.9). Further, all of the PEM associations (i.e., 100%) remained significant when we controlled for mother anxiety/depression (Table S3.10).

Gender as a Moderator

To examine whether gender moderated the association between adolescent temperament development and anxiety/depression during young adulthood, we included gender and its interaction with temperament level and slope into multiple regressions predicting EC, NEM, and Surgency (the core of PEM; Table S3.11). We found one significant interaction ($\beta = -.16$, $p = .005$), suggesting that the association between the Surgency level and anhedonic depression is stronger for girls than for boys.

Table 3.3
Associations Between Temperament Trajectories and Young Adult Anxiety/Depression

	General Distress			Anxious Arousal			Anhedonic Depression		
	Age 19	Age 21	Age 19 and 21 Composite	Age 19	Age 21	Age 19 and 21 Composite	Age 19	Age 21	Age 19 and 21 Composite
EC									
Level	-.07 (*p* = .076)	.00 (*p* = .969)	-.05 (*p* = .212)	-.10 (*p* = .011)	-.04 (*p* = .403)	-.09 (*p* = .020)	**-.12* (*p* = .004)**	-.01 (*p* = .841)	-.08 (*p* = .048)
Slope	**-.17* (*p* < .001)**	-.10 (*p* = .016)	**-.15* (*p* < .001)**	**-.12* (*p* = .005)**	-.05 (*p* = .225)	-.10 (*p* = .010)	**-.15* (*p* < .001)**	**-.17* (*p* < .001)**	**-.18* (*p* < .001)**
Activation Control									
Level	-.07 (*p* = .111)	.02 (*p* = .716)	-.03 (*p* = .441)	**-.14* (*p* < .001)**	-.00 (*p* = .914)	-.09 (*p* = .032)	-.05 (*p* = .230)	.00 (*p* = .933)	-.02 (*p* = .592)
Slope	**-.15* (*p* < .001)**	-.06 (*p* = .158)	**-.11* (*p* = .005)**	-.10 (*p* = .015)	-.02 (*p* = .718)	-.07 (*p* = .081)	**-.12* (*p* = .006)**	-.09 (*p* = .036)	**-.12* (*p* = .003)**
Attention									
Level	-.07 (*p* = .189)	.04 (*p* = .471)	-.03 (*p* = .573)	-.09 (*p* = .098)	.06 (*p* = .324)	-.02 (*p* = .703)	-.09 (*p* = .100)	.08 (*p* = .150)	-.01 (*p* = .832)
Slope	-.08 (*p* = .134)	-.15 (*p* = .008)	-.13 (*p* = .017)	-.05 (*p* = .342)	-.14 (*p* = .014)	-.12 (*p* = .021)	-.12 (*p* = .024)	**-.20* (*p* < .001)**	**-.19* (*p* < .001)**
Inhibitory Control									
Level	-.02 (*p* = .583)	.03 (*p* = .467)	-.00 (*p* = .969)	-.01 (*p* = .744)	-.06 (*p* = .174)	-.06 (*p* = .140)	-.08 (*p* = .048)	-.02 (*p* = .609)	-.07 (*p* = .078)
Slope	-.06 (*p* = .163)	-.05 (*p* = .241)	-.05 (*p* = .207)	-.03 (*p* = .484)	-.04 (*p* = .326)	-.04 (*p* = .341)	-.09 (*p* = .025)	**-.13* (*p* = .003)**	**-.11* (*p* = .005)**
NEM									
Level	**.20* (*p* < .001)**	**.14* (*p* = .005)**	**.19* (*p* < .001)**	**.17* (*p* < .001)**	.09 (*p* = .080)	**.17* (*p* < .001)**	**.18* (*p* < .001)**	.12 (*p* = .012)	**.17* (*p* < .001)**
Slope	**.23* (*p* < .001)**	**.16* (*p* < .001)**	**.22* (*p* < .001)**	**.16* (*p* < .001)**	.11 (*p* = .023)	**.19* (*p* < .001)**	**.18* (*p* < .001)**	**.18* (*p* < .001)**	**.20* (*p* < .001)**

NEM (narrow)									
Level	**.14* ($p = .002$)**	.07 ($p = .129$)	.12 ($p = .008$)	.10 ($p = .036$)	.07 ($p = .119$)	.12 ($p = .008$)	.11 ($p = .020$)	.04 ($p = .399$)	.08 ($p = .065$)
Slope	**.19* ($p < .001$)**	.07 ($p = .136$)	**.16* ($p < .001$)**	.11 ($p = .022$)	.04 ($p = .391$)	.12 ($p = .010$)	.09 ($p = .053$)	.06 ($p = .183$)	.09 ($p = .049$)
Fear									
Level	**.12* ($p = .005$)**	.04 ($p = .322$)	.09 ($p = .034$)	.02 ($p = .704$)	.03 ($p = .485$)	.04 ($p = .302$)	.07 ($p = .103$)	.02 ($p = .649$)	.05 ($p = .279$)
Slope	**.16* ($p < .001$)**	.05 ($p = .286$)	**.13* ($p = .003$)**	.11 ($p = .015$)	-.03 ($p = .506$)	.06 ($p = .144$)	.07 ($p = .090$)	.04 ($p = .406$)	.06 ($p = .132$)
Frustration									
Level	**.12* ($p = .005$)**	.08 ($p = .069$)	**.11* ($p = .006$)**	**.12* ($p = .003$)**	.09 ($p = .031$)	**.14* ($p < .001$)**	.11 ($p = .012$)	.05 ($p = .227$)	.09 ($p = .022$)
Slope	**.17* ($p < .001$)**	.11 ($p = .014$)	**.16* ($p < .001$)**	.05 ($p = .276$)	.07 ($p = .120$)	.09 ($p = .028$)	.07 ($p = .079$)	.10 ($p = .027$)	.10 ($p = .019$)
Aggression									
Level	.06 ($p = .156$)	.07 ($p = .126$)	.08 ($p = .042$)	**.12* ($p = .004$)**	.03 ($p = .557$)	.10 ($p = .010$)	.07 ($p = .115$)	**.12* ($p = .006$)**	.11 ($p = .007$)
Depressed Mood									
Level	**.18* ($p < .001$)**	**.16* ($p < .001$)**	**.19* ($p < .001$)**	**.17* ($p < .001$)**	.07 ($p = .109$)	**.15* ($p < .001$)**	**.19* ($p < .001$)**	.11 ($p = .009$)	**.16* ($p < .001$)**
Slope	**.23* ($p < .001$)**	**.17* ($p < .001$)**	**.21* ($p < .001$)**	.11 ($p = .007$)	.11 ($p = .012$)	**.14* ($p < .001$)**	**.20* ($p < .001$)**	**.20* ($p < .001$)**	**.22* ($p < .001$)**
Surgency									
Level	.02 ($p = .573$)	-.02 ($p = .699$)	-.02 ($p =.686$)	.10 ($p = .015$)	-.03 ($p = .479$)	.03 ($p = .486$)	-.06 ($p = .157$)	-.01 ($p = .817$)	-.06 ($p = .125$)
Slope	-.10 ($p = .021$)	-.02 ($p = .632$)	-.06 ($p = .127$)	-.07 ($p = .104$)	.01 ($p = .862$)	-.04 ($p = .336$)	**-.18* ($p < .001$)**	-.06 ($p = .180$)	**-.13 ($p = .001$)**
Affiliation									
Level	.03 ($p = .475$)	.03 ($p = .440$)	.01 ($p = .735$)	.03 ($p = .402$)	-.02 ($p = .696$)	.01 ($p = .897$)	**-.14* ($p < .001$)**	-.07 ($p = .084$)	**-.14* ($p < .001$)**

Slope	-.05 (*p* = .243)	-.06 (*p* = .163)	-.07 (*p* = .108)	-.03 (*p* = .510)	-.08 (*p* = .051)	-.09 (*p* = .037)	**-.20* (*p* < .001)**	**-.17* (*p* < .001)**	**-.22* (*p* < .001)**

Note. Values are standardized coefficients from multiple regressions where anxiety/depression were regressed on the level and slope of each temperament domain (i.e., intercept and slope were entered jointly). The Surgency associations are shown for age 12 to 16 trajectory, but they replicate for the age 10 to 16 trajectory. * $p < .006$

Discussion

The present study examined the development of EC, NEM, and PEM from late childhood (age 10) through adolescence (age 16) and associations between these developmental trajectories and anxiety/depression in young adulthood (ages 19 and 21). For the development of adolescent temperament, we found evidence of both personality maturation (NEM, Surgency) and disruption (EC) from age 10 to 16. Additionally, we found numerous predicted associations between the developmental trajectories of temperament and anxiety/depression during young adulthood. Below, we situate our results in the existing literature and discuss the findings more generally, separately for each temperament domain.

Mean-Level Change in Temperament

Consistent with the disruption hypothesis (Soto & Tackett, 2015) and with the majority of prior empirical research (De Fruyt et al., 2006; Göllner et al., 2017; Laceulle et al., 2012; van den Akker et al., 2014; Zohar et al., 2019), we found mean-level decreases in the EC domain from age 10 to 16. These previous studies span a wide range of measures, assessment methods, languages, age ranges, and measurement occasions (see Table 3.1), suggesting that this finding is robust and generalizable across various samples and methods. At the facet level, results for Activation Control mirrored those of the EC domain; however, for Attention, youth increased, on average, from age 10 to 12 but then decreased from age 12 to 16. Conversely, youth tended to increases in Inhibitory Control from age 10 to 16. Overall, these findings suggest that, on average, youth tend to become worse at regulating their impulses and paying close attention as they go through adolescence, although there is substantial variability across individuals.

Aligned with the maturity principle (Roberts et al., 2006) and consistent with our hypotheses, we found mean-level decreases in NEM and all of its facets, except Aggression, from age 10 to 16. These findings converge with prior research examining NEM development across adolescence (Brandes et al., 2022; De Fruyt et al., 2006; Göllner et al., 2017; Klimstra et al., 2009; Laceulle et al., 2012; Pullmann et al., 2006; Roberts et al., 2006; van den Akker et al., 2014; Zohar et al., 2019). Similar to EC, the breadth of measures and longitudinal designs used in previous studies (see Table 3.1) suggests that this normative pattern is generalizable to adolescents from a wide range of populations. Although we expected to find linear decreases in NEM and its facets, we found nonlinear trajectories for the NEM domain and Fear facet; in particular, decreases were greater in magnitude from age 10 to 12 than from age 12 to 16. This pattern suggests that the majority of personality maturation with respect to NEM happens earlier on in adolescence and then continues at a slower rate from mid- to late-adolescence. This may be due to the fact that the challenges of adolescence ramp up substantially after puberty and the transition to high school, with increasing autonomy from parents, more complex peer and romantic relationships, and an increasingly competitive school environment, all of which demand more emotional stability. With respect to Aggression, we found that a no-growth model fit the data best due, in large part, to the fact that most youth reported very low levels of Aggression at all ages. Therefore, we do not interpret this finding as evidence that adolescents do not change in their aggression across adolescence, but rather that the EATQ-R did not do a good job measuring aggressive tendencies in this sample.

Consistent with our hypotheses, we found increases in Surgency from age 12 to 16 (and replicated these associations from age 10 to 16), which is consistent with many previous studies (Göllner et al., 2017; Klimstra et al., 2009; Laceulle et al., 2012; Pullmann et al., 2006; Roberts et al., 2006). Also consistent with our hypotheses, we found small mean-level decreases in Affiliation from age 10 to 16. The decrease is consistent with previous research on the Affiliation facet (Göllner et al., 2017; Laceulle et al., 2012; Pullmann et al., 2006; van den Akker et al., 2014), although the small magnitude of the change helps explain why some studies have found no mean-level changes in Affiliation across adolescence (De Fruyt et al., 2006; Pullman et al., 2006; Roberts et al., 2006).

Across all of the temperament domains, the findings from the present study were largely consistent with previous studies that used many different measures of both temperament and Big Five personality traits (see Table 3.1). Given that relatively few studies used a temperament measure across this same age range (e.g., Laceulle et al., 2012; Zohar et al., 2018), it is difficult to draw strong conclusions about the reasons underlying small discrepancies from past findings. However, together, this suggests that, in addition to similar conceptualizations and well-documented moderate concurrent correlations, temperament traits and personality traits develop across adolescence in much the same way, highlighting the similarity between these constructs (Lawson & Robins, 2021) and supporting evidence that the distinction between them is somewhat arbitrary (Clark & Watson, 2008; Shiner et al., 2021). Furthermore, in our sample, we did not find any significant gender differences in the levels and slopes of the temperament domains and facets, which diverges from previous studies including Brandes et al. (2020), Borghuis et al. (2017), Göllner et al. (2017), and Klimstra et al. (2009) where many of the domains showed pronounced gender differences. The discrepant findings may reflect differences between the studies in the samples (Mexican-origin youth vs. majority White youth), the measures, the assessment methods, the developmental periods examined, or a combination of these factors (see Table 3.1 for a detailed comparison of previous studies).

Temperament Development and Anxiety/Depression

After estimating the developmental trajectories of the temperament domains and facets, we examined whether these trajectories predicted anxiety and depression in young adulthood. Notably, this design allowed for "taking into consideration developmentally normative…change in personality and psychopathology [which] is critical for understanding their links at any given time over the course of development" (Wilson & Olino, 2021, p. 920). Our prospective, longitudinal design, which allowed us to examine associations between temperament *change* and anxiety/depression, has several characteristics that were previously identified as important directions for future studies (Klein et al., 2011). For example, we examined multimethod assessments of EC, NEM, and PEM, as well as their facets, helping to "determine whether a more specific level of analysis will yield more powerful effects and increase the specificity of associations between personality constructs and particular forms of psychopathology" (Klein et al., 2011, p. 19).

As predicted, changes in EC were associated with symptoms related to depression (i.e., general distress and anhedonic depression), but not those uniquely related to anxiety (i.e., anxious arousal). This finding is consistent with prior longitudinal research that found significant associations between earlier levels of EC and later depression, but not anxiety (Crockett et al., 2013; Loukas & Roalson, 2006; Ormel et al., 2005; Verstraeten et al., 2009). It is also consistent with Lengua's (2006) finding that youth who experienced smaller decreases in EC across three years had fewer internalizing problems at the final temperament assessment (i.e., in Year 3). For the facets of EC, we found that associations with Activation Control largely paralleled the EC domain, whereas the slopes of Attention and Inhibitory Control were correlated only with anhedonic depression, but not general distress. This pattern is not consistent with theoretical work suggesting that the Attention facet should be most strongly associated with anxiety and depression (Compas et al., 2004; Muris et al., 2007). However, these findings may reflect the fact that decreases in Attention and Inhibitory Control are related to emotion regulation, including reward and threat sensitivity, which is related to anhedonic depression but not general distress (Young et al., 2022). Inhibitory Control may be especially relevant to the emotion regulation strategy of suppression, or inhibiting emotionally expressive behavior, whereas youth with better Attention Control may be able to shift their attention away from what is making them sad and selectively attend to positive thoughts (Compas et al., 2004; Nigg, 2006). Notably for the developmental trajectories of the EC domain and facets, individual differences in the slopes, but not the levels, were most often associated with depression during young adulthood, which suggests that the way youth develop across adolescence is more relevant to young adulthood psychopathology than their initial EC levels at age 10. This finding supports the idea that adolescents who are getting worse at regulating themselves may have even more trouble dealing with changes during adolescence and the transition to young adulthood, given that they have had fewer opportunities to develop effective strategies for coping with negative person-environment transactions, such as dysfunctional relationships with parents, teachers, peers, and romantic partners.

As predicted, higher initial levels and greater increases in the broad NEM domain across adolescence were associated with higher general distress, anxious arousal, and anhedonic depression scores during young adulthood. These findings converge with previous research that has consistently documented earlier levels of both self- and parent-reported NEM predisposing youth to later anxiety and depression (Bould et al., 2014; Bouma et al., 2008; Brendgen et al., 2005; Caspi et al., 1996; Crockett et al., 2013; Davis et al., 2015; Lonigan et al., 2003; Mezulis et al., 2011; Ormel et al., 2005; Verstraeten et al., 2009). They also converge with prior studies showing that changes in NEM were related to later internalizing symptoms (Laceulle et al., 2014; Lengua, 2006; van den Akker et al. 2010). Not surprisingly, the strongest associations were with the Depressed Mood facet; in particular, both the levels and slopes were associated with general distress, anxious arousal, and anhedonic depression during young adulthood. This suggests that adolescents with a temperamental disposition toward depressed mood may spiral downward over time in their anxiety and depression symptoms, perhaps due to difficulties coping with academic, familial, and interpersonal problems that occur frequently during adolescence. There were also clear associations between Frustration and both general distress and anxious arousal. Adolescents who get irritated and annoyed when dealing with daily problems or when their progress toward reaching their goals is thwarted may feel more stressed and helpless in the face of future problems, contributing to later distress and anxiety. Further, our findings suggest that the associations between NEM and anxiety/depression are not exclusively due to conceptual overlap between measures. In particular, we saw associations between the narrow NEM domain (including only Fear and Frustration) and general distress, suggesting that key aspects of temperament besides conceptually overlapping Depressed Mood are prospectively associated with anxiety/depression. Surprisingly, the Fear domain was not associated with anxious arousal, despite the fact that fear is a core aspect of what is unique to anxiety.

Partially consistent with our hypothesis, the individual differences in changes in Surgency were negatively associated with anhedonic depression, but not general distress, during young adulthood. This suggests that youth who become more novelty seeking and reward sensitive across adolescence are less likely to experience anhedonia later on. This finding connects to previous longitudinal research that has found that higher Surgency is associated with fewer internalizing problems later on (Lonigan et al., 2003; Ormel et al., 2005; Verstraeten et al., 2009). The fact that Surgency development relates to anhedonic depression, but not general distress, is likely due to the complementary relationship between anhedonia (lack of pleasure) and Surgency (high-intensity pleasure; Klein et al., 2011). Further, contrary to our hypothesis, individual differences in the levels and slopes of Affiliation did not predict either general distress or anxious arousal. However, Affiliation was associated with anhedonic depression, such that youth who have lower initial levels and greater decreases in Affiliation from age 10 to 16 tend to have higher levels of anhedonic depression during young adulthood. These findings add to the limited and inconsistent research on Affiliation and anxiety/depression (Kushner et al., 2012; Oldehinkel, 2004; Ormel et al., 2005). Overall, the PEM domain (both levels and changes over time) showed inconsistent associations with later anxiety/depression symptoms, potentially because PEM may serve as a risk or a protective factor for internalizing problems depending on the context. For example, high levels of Surgency may be protective in higher risk, but not lower risk, environments (Nigg, 2006; Werner & Smith, 1992).

Across all three temperament domains, we found evidence consistent with our hypotheses about the vulnerability model of personality and psychopathology; in particular, certain temperament traits place an adolescent at greater risk for, or protect them from, later psychopathology (Tackett, 2006). These findings were stronger for general distress and anhedonic depression than for anxious arousal, which is consistent with prior research suggesting that temperament has a tighter connection with depression than anxiety (Laceulle et al., 2014). Further, extending previous research, we found that the way temperament develops across adolescence (i.e., individual differences in change) predicts experiences of anxiety/depression over and above initial temperament levels. This suggests that youth who are experiencing personality disruption across adolescence may struggle more with mental health problems in young adulthood, compared to youth who show normative personality maturation during adolescence. Importantly, we replicated prior results in our sample of Mexican-origin youth, providing initial evidence for the generalizability of these patterns. Of course, future research should continue to examine the generalizability of these results to other samples of Mexican-origin adolescents, Latinx adolescents more generally, and adolescents of other races, ethnicities, and nationalities.

Our associations between temperament and anxiety/depression are relatively robust for several reasons. First, our measure of anxiety/depression asked about symptoms during the previous week. This means that adolescent temperament (age 10 to 16) was significantly associated with anxiety/depression symptoms during a particular week three (for age 19) to five (for age 21) years later. Second, we used a fairly strict Bonferroni correction to adjust our alpha level to .006. Therefore, the significant associations we interpret have p-values lower than .006 and, in a world where there is no significant association between temperament and anxiety/depression, such findings would be infrequent. Third, the majority of the significant temperament-anxiety/depression associations remained significant when we controlled for mother anxiety/depression. However, when we controlled for prior anxiety/depression symptoms (assessed at age 16), the majority of the associations between temperament and anxiety/depression were no longer significant. This suggests that adolescent temperament is significantly related to anxiety/depression during young adulthood, but primarily via its effect on anxiety/depression in late adolescence.

Limitations and Future Directions

The present findings should be considered in the context of several limitations that suggest directions for future research. First, the present study does not examine any of the various sociocultural factors that may influence associations between temperament and anxiety/depression. Given previously documented associations between internalizing problems and stressors impacting Mexican-origin youth, including acculturative stress and discrimination, (Bridges et al., 2021; Maiya et al., 2021; Suarez-Morales & Lopez, 2009; Torres, 2010), future research should examine whether experiences with these stressors moderate the associations between temperament and anxiety/depression. Second, the present study focused on the vulnerability model, but we were unable to distinguish between the predisposition model (i.e., personality and psychopathology are causally related) and precursor models (i.e., personality and psychopathology are associated due to shared etiology); to address this issue, future research should use twin designs or quasi-experimental interventions to better elucidate the nature of the temperament-psychopathology association (Wilson & Olino, 2021). Further, our results do not necessarily undermine other developmental models of temperament and anxiety/depression, including the scar, pathoplasty, and spectrum models (Durbin & Hicks, 2014; Tackett, 2006). Future research should continue to examine all of these models and, particularly, whether certain models better capture the developmental associations for certain combinations of temperament traits and mental health problems. For example, previous research suggests that Affiliation has a pathoplastic relationship with anxiety/depression (Cain et al., 2012; Dawood et al., 2013; Przeworski et al., 2011; Salzer et al., 2008). Third, anxiety and depression may be changing in tandem with temperament traits across adolescence and into young adulthood, so future studies should examine these co-developmental patterns. For example, youth showing increases (decreases) over time in EC may show corresponding decreases (increases) over time in depression across the same period. Fourth, future research should further explore gender differences in temperament development and its associations with anxiety/depression. For example, it is possible that the well-documented gender difference in depression is due, in part, to gender differences in the temperamental traits associated with the development of depression. Fifth, researchers should explore whether mental healthcare providers at schools and community centers could use measures of adolescent temperament to identify adolescents likely to experience anxiety and depression in young adulthood and connect them with resources to mitigate the negative impact.

Conclusion

The present study provides a nuanced depiction of the development of EC, NEM, and PEM (and their facets) from age 10 to 16 in a large sample of Mexican-origin youth. Additionally, our findings suggest robust associations between the way that temperament develops across adolescence (i.e., level and slope of the trajectory) and anxiety/depression in early adulthood, with EC and PEM serving as protective factors and NEM as a risk factor.

Supplemental Materials

Table S3.1

Descriptive Statistics for Observed Temperament Domains/Facets

	Age 10		Age 12		Age 14		Age 16	
	M (SD)	*α / ω*	*M (SD)*	*α / ω*	*M (SD)*	*α / ω*	*M (SD)*	*α / ω*
Effortful Control	2.96 (.34)	.78 / .81	3.00 (.35)	.82 / .84	2.94 (.35)	.84 / .86	2.94 (.33)	.82 / .84
Activation Control	2.96 (.47)	.63 / .69	2.88 (.49)	.69 / .75	2.77 (.48)	.72 / .76	2.75 (.45)	.68 / .74
Attention	2.95 (.39)	.60 / .67	3.03 (.37)	.61 / .67	2.98 (.38)	.64 / .70	2.98 (.37)	.66 / .72
Inhibitory Control	2.99 (.40)	.46 / .52	3.08 (.38)	.51 / .56	3.06 (.39)	.56 / .65	3.10 (.38)	.55 / .61
Negative emotionality	2.18 (.31)	.84 / .86	1.99 (.29)	.84 / .87	1.91 (.32)	.82 / .86	1.87 (.30)	.84 / .87
Negative emotionality (narrow)	2.61 (.38)	.77 / .80	2.35 (.37)	.78 / .82	2.20 (.37)	.66 / .74	2.14 (.37)	.78 / .82
Fear	2.83 (.47)	.67 / .72	2.47 (.49)	.70 / .75	2.19 (.43)	.61 / .68	2.12 (.43)	.64 / .71
Frustration	2.41 (.46)	.73 / .79	2.25 (.44)	.74 / .81	2.21 (.45)	.76 / .82	2.17 (.45)	.77 / .83
Aggression	1.42 (.41)	.78 / .83	1.40 (.38)	.79 / .84	1.41 (.40)	.82 / .87	1.36 (.33)	.75 / .81
Depressed Mood	2.01 (.41)	.62 / .69	1.79 (.36)	.57 / .67	1.78 (.40)	.72 / .78	1.79 (.41)	.73 / .78
Surgency	2.53 (.36)	.34 / .45	2.59 (.39)	.76 / .79	2.66 (.37)	.77 / .81	2.65 (.37)	.79 / .82
Affiliation	2.91 (.38)	.57 / .64	2.88 (.41)	.56 / .67	2.94 (.40)	.65 / .75	2.89 (.41)	.66 / .73

Note. M = mean, SD = standard deviation, α = alpha reliability, *ω* = omega reliability.

Table S3.2

Descriptive Statistics for Anxiety/Depression

	Age 16		**Age 19**		**Age 21**		**Age 19 and 21 Composite**	
	M (SD)	*α / ω*	*M (SD)*	*α / ω*	*M (SD)*	*α / ω*	*M (SD)*	*α / ω*
General Distress	1.48 (.48)	.87 / .92	1.41 (.49)	.88 / .93	1.50 (.62)	.91 / .95	1.45 (.48)	.91 / .94
Anxious Arousal	1.15 (.25)	.81 / .84	1.14 (.27)	.85 / .89	1.18 (.35)	.87 / .90	1.16 (.24)	.86 / .89
Anhedonic Depression	1.86 (.53)	.83 / .89	1.84 (.60)	.87 / .91	1.79 (.64)	.88 / .92	1.81 (.52)	.89 / .93

Note. M = mean, SD = standard deviation, α = alpha reliability, ω = omega reliability.

Table S3.3

Correlations Between Anxiety/Depression Latent Variables

Variable	1	2	3	4	5	6	7	8	9	10	11
1. GD (16)											
2. GD (19)	.34 [.26, .41]										
3. GD (21)	.28 [.20, .35]	.43 [.36, .50]									
4. GD (YA)	.34 [.26, .41]	.83 [.80, .85]	.89 [.87, .90]								
5. AA (16)	.53 [.47, .59]	.16 [.07, .24]	.21 [.12, .29]	.19 [.11, .27]							
6. AA (19)	.22 [.14, .29]	.53 [.47, .58]	.29 [.21, .37]	.47 [.41, .53]	.22 [.14, .30]						
7. AA	.15	.17	.53	.44	.21	.19					

(21)											
	[.07, .24]	[.08, .25]	[.47, .59]	[.37, .50]	[.13, .29]	[.11, .27]					
8. AA (YA)	.23	.43	.56	.58	.27	.72	.84				
	[.15, .31]	[.36, .49]	[.50, .62]	[.53, .63]	[.19, .34]	[.68, .76]	[.81, .86]				
9. AD (16)	.52	.23	.22	.25	.30	.15	.09	.16			
	[.46, .57]	[.15, .31]	[.14, .30]	[.18, .33]	[.23, .37]	[.06, .23]	[.01, .18]	[.08, .24]			
10. AD (19)	.24	.58	.32	.51	.08	.40	.12	.32	.37		
	[.16, .32]	[.52, .63]	[.24, .39]	[.45, .57]	[-.01, .16]	[.33, .47]	[.04, .21]	[.24, .39]	[.29, .44]		
11. AD (21)	.25	.30	.61	.56	.18	.23	.37	.41	.33	.40	
	[.17, .33]	[.22, .37]	[.55, .66]	[.50, .61]	[.09, .26]	[.15, .31]	[.30, .44]	[.34, .48]	[.26, .41]	[.33, .47]	
12. AD (YA)	.27	.51	.57	.62	.13	.37	.31	.43	.40	.84	.85
	[.19, .34]	[.45, .57]	[.51, .62]	[.57, .67]	[.04, .20]	[.30, .44]	[.23, .38]	[.36, .49]	[.33, .46]	[.81, .86]	[.83, .87]

Note. GD = General distress. AA = Anxious arousal. AD = Anhedonic depression. YA = Young adult. Ages (16, 19, 21, YA) are in brackets. Values in square brackets indicate the 95% confidence interval for each correlation. All correlations are significant at $p < .05$

Table S3.4
Model Comparisons for Longitudinal Measurement Invariance

	χ^2	*df*	CFI	RMSEA [90% CI]
Effortful Control				
Configural	98.37	74	1.00	.022 [.007, .033]
Weak	117.66	80	.99	.026 [.015, .036]
Partial Strong	**163.86**	**89**	**.99**	**.035 [.027, .044]**
Strong	272.00	92	.96	.054 [.047, .061]
Activation Control				
Configural	26.10	30	1.00	.000 [.000, .024]
Weak	28.84	33	1.00	.000 [.000, .023]
Partial Strong	**30.72**	**39**	**1.00**	**.000 [.000, .017]**
Strong	148.99	42	.95	.061 [.051, .072]
Attention				
Configural	32.86	30	1.00	.012 [.000, .032]
Weak	45.22	33	.99	.023 [.000, .039]
Partial Strong	**46.88**	**39**	**.99**	**.017 [.000, .033]**
Strong	142.41	42	.95	.060 [.049, .070]
Inhibitory Control				
Configural	40.33	30	.99	.023 [.000, .039]
Weak	46.66	33	.99	.025 [.000, .040]
Partial Strong	**209.49**	**39**	**.88**	**.081 [.070, .091]**
Strong	360.53	42	.78	.106 [.096, .116]
Negative Emotionality				
Configural	103.24	74	1.00	.024 [.011, .035]
Weak	134.87	80	.99	.032 [.022, .041]
Partial Strong	**189.40**	**89**	**.98**	**.041 [.033, .049]**
Strong	316.31	92	.97	.060 [.053, .067]
Negative Emotionality (narrow)				
Configural	110.71	74	.99	.027 [.016, .037]
Weak	143.06	80	.99	.034 [.025, .043]
Partial Strong	**182.99**	**89**	**.98**	**.040 [.031, .048]**
Strong	378.84	92	.93	.068 [.061, .075]
Fear				
Configural	37.07	30	1.00	.019 [.000, .036]
Weak	62.36	33	.99	.036 [.022, .050]

Partial Strong	75.98	39	.98	.038 [.025, .050]
Strong	**79.71**	**42**	**.98**	**.036 [.024, .049]**
Frustration				
Configural	20.25	30	1.00	.000 [.000, .012]
Weak	26.69	33	1.00	.000 [.000, .020]
Partial Strong	**29.11**	**39**	**1.00**	**.000 [.000, .014]**
Strong	74.12	42	.99	.034 [.021, .046]
Aggression				
Configural	39.68	30	1.00	.022 [.000, .039]
Weak	50.13	33	1.00	.028 [.009, .043]
Partial Strong	**84.70**	**39**	**.99**	**.042 [.030, .054]**
Strong	133.22	42	.98	.057 [.046, .068]
Depressed Mood				
Configural	39.41	30	1.00	.022 [.000, .038]
Weak	49.30	33	.99	.027 [.008, .042]
Partial Strong	**192.56**	**39**	**.94**	**.076 [.066, .087]**
Strong	255.74	42	.91	.087 [.077, .097]
Surgency (age 12 to 16)				
Configural	39.09	15	.99	.050 [.031, .070]
Weak	43.77	17	.99	.050 [.032, .068]
Partial Strong	**87.31**	**21**	**.97**	**.070 [.055, .086]**
Strong	168.71	23	.93	.100 [.086, .114]
Surgency (age 10 to 16)				
Configural	52.85	30	.99	.034 [.018, .048]
Weak	59.53	33	.99	.035 [.020, .048]
Partial Strong	**155.46**	**39**	**.95**	**.067 [.056, .078]**
Strong	289.35	42	.89	.093 [.083, .104]
Affiliation				
Configural	32.54	30	1.00	.011 [.000, .032]
Weak	45.92	33	.99	.024 [.000, .040]
Partial Strong	**190.75**	**39**	**.91**	**.076 [.065, .087]**
Strong	217.93	42	.90	.079 [.069, .089]

Note. χ^2 = Chi-square test statistics; df = degrees of freedom; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; 90%CI = 90% confidence interval.

Table S3.5
Factor Loadings for Parcels for Latent Variables

	Parcel 1	Parcel 2	Parcel 3	Parcel 4
Temperament				
Effortful Control	1.00	0.86	0.73	0.99
Activation Control	1.00	0.79	0.79	-
Attention	1.00	0.76	0.85	-
Inhibitory Control	1.00	0.43	0.49	-
Negative Emotionality	1.00	0.94	0.95	1.05
Negative Emotionality (narrow)	1.00	0.94	0.98	1.14
Fear	1.00	0.96	1.01	-
Frustration	1.00	1.19	1.09	-
Depressed Mood	1.00	1.05	1.27	-
Aggression	1.00	0.80	0.82	-
Surgency	1.00	1.00	1.04	-
Surgency (age 10 to 16)	1.00	0.95	1.01	-
Affiliation	1.00	1.68	1.36	-
Anxiety/Depression				
General Distress	0.50	0.46	0.46	-
Anxious Arousal	0.21	0.23	0.25	-
Anhedonic Depression	0.48	0.50	0.53	-

Note. The first parcel for all temperament factor loadings was constrained to be 1. All temperament domains/facets were at least partially strong invariant across time; therefore, factor loadings were constrained to be equal across waves. Factor loadings for young adult anxiety/depression scores are reported.

Table S3.6

Model Statistics for Best-Fitting Second-Order LGC Models for Temperament Domains/Facets

	No Growth	**Linear Growth**	Latent Basis
Effortful Control			
Slope			
β_1	0	**0**	0
β_2	0	**2**	1.35
β_3	0	**4**	5.64
β_4	0	**6**	6
Means			
Level	.00	**.00**	.00
Slope	-	**-.02***	-.02*
Variances			
Level	.08*	**.09***	.08*
Slope	-	**.002***	.001*
Covariance $_{\text{Level, Slope}}$	-	**-.003***	-.001
Goodness-of-Fit			
χ^2 (df)	273.66 (97)	**198.04 (94)**	191.78 (92)
RMSEA [90% CI]	.052 [.045, .059]	**.041 [.033, .048]**	.040 [.032, .048]
CFI	.96	**.98**	.98
Fit changes, $\Delta\chi^2/\Delta df$	-	**75.62/3**	6.26/2
Activation Control			
	No Growth	**Linear Growth**	Latent Basis
Slope			
β_1	0	**0**	0
β_2	0	**2**	0.74
β_3	0	**4**	3.87
β_4	0	**6**	6
Means			
Level	.00	**.00**	.00
Slope	-	**-.02***	-.02*
Variances			
Level	.13*	**.15***	.15*
Slope	-	**.003***	.00
Covariance $_{\text{Level, Slope}}$	-	**-.01***	-.00

Goodness-of-Fit			
χ^2 (df)	114.36 (47)	**60.68 (44)**	59.34 (42)
RMSEA [90% CI]	.046 [.035, .057]	**.024 [.002, .037]**	.025 [.006, .038]
CFI	.97	**.99**	.99
Fit changes, $\Delta\chi^2/\Delta df$	-	**53.68/3**	1.34/2
Attention			
	No Growth	Linear Growth	**Latent Basis**
Slope			
β_1	0	0	**0**
β_2	0	2	**7.58**
β_3	0	4	**6.45**
β_4	0	6	**6**
Means			
Level	.00	.00	**.00**
Slope	-	.01*	**.02***
Variances			
Level	.09*	.08*	**.08***
Slope	-	.002*	**.00**
Covariance $_{\text{Level, Slope}}$	-	-.00	**.00**
Goodness-of-Fit			
χ^2 (df)	131.36 (47)	97.84 (44)	**68.81 (43)**
RMSEA [90% CI]	.052 [.041, .062]	.043 [.031, .054]	**.030 [.016, .043]**
CFI	.96	.97	**.99**
Fit changes, $\Delta\chi^2/\Delta df$	-	33.52/3	**29.02/1**
Inhibitory Control			
	No Growth	**Linear Growth**	Latent Basis
Slope			
β_1	0	**0**	0
β_2	0	**2**	3.60
β_3	0	**4**	4.72
β_4	0	**6**	6
Means			
Level	.00	**.00**	.00
Slope	-	**.05***	.06*
Variances			

Level	.04*	**.10***	.14*
Slope	-	**.00**	.00
Covariance Level, Slope	-	**-.00**	-.01
Goodness-of-Fit			
χ^2 (df)	293.07 (47)	**228.62 (44)**	209.90 (42)
RMSEA [90% CI]	.088 [.079, .098]	**.079 [.069, .089]**	.077 [.067, .088]
CFI	.83	**.87**	.88
Fit changes, $\Delta\chi^2/\Delta df$	-	**64.45/3**	18.72/2

Negative Emotionality			
	No Growth	Linear Growth	**Latent Basis**
Slope			
β_1	0	0	**0**
β_2	0	2	**3.83**
β_3	0	4	**5.60**
β_4	0	6	**6**
Means			
Level	.00	.00	**.00**
Slope	-	-.04*	**-.04***
Variances			
Level	.05*	.06*	**.08***
Slope	-	.001*	**.002***
Covariance Level, Slope	-	-.003*	**-.01***
Goodness-of-Fit			
χ^2 (df)	677.00 (97)	300.61 (94)	**213.61 (92)**
RMSEA [90% CI]	.094 [.088, .101]	.057 [.050, .064]	**.044 [.037, .052]**
CFI	.91	.97	**.98**
Fit changes, $\Delta\chi^2/\Delta df$	-	376.39/3	**87.00/2**

Negative Emotionality (narrow)			
	No Growth	Linear Growth	**Latent Basis**
Slope			
β_1	0	0	**0**
β_2	0	2	**3.45**
β_3	0	4	**5.58**
β_4	0	6	**6**
Means			

Level	.00	.00	**.00**
Slope	-	-.06*	**-.06***
Variances			
Level	.06*	.07*	**.09***
Slope	-	.002*	**.002***
Covariance $_{\text{Level, Slope}}$	-	-.003*	**-.01***
Goodness-of-Fit			
χ^2 (df)	740.32 (97)	272.25 (94)	**192.13 (92)**
RMSEA [90% CI]	.099 [.093, .106]	.053 [.046, .060]	**.040 [.032, .048]**
CFI	.85	.96	**.98**
Fit changes, $\Delta\chi^2/\Delta df$	-	468.07/3	**80.12/2**
Fear			
	No Growth	Linear Growth	**Latent Basis**
Slope			
β_1	0	0	**0**
β_2	0	2	**2.97**
β_3	0	4	**5.45**
β_4	0	6	**6**
Means			
Level	.00	.00	**.00**
Slope	-	-.12*	**-.12***
Variances			
Level	.07*	.08*	**.10***
Slope	-	.00	**.001***
Covariance $_{\text{Level, Slope}}$	-	-.00	**-.01***
Goodness-of-Fit			
χ^2 (df)	1158.16 (50)	234.40 (47)	**90.04 (45)**
RMSEA [90% CI]	.181 [.172, .190]	.077 [.067, .087]	**.039 [.027, .050]**
CFI	.49	.91	**.98**
Fit changes, $\Delta\chi^2/\Delta df$	-	923.76/3	**144.36/2**
Frustration			
	No Growth	**Linear Growth**	Latent Basis
Slope			
β_1	0	**0**	0
β_2	0	**2**	4.03

β_3	0	**4**	5.75
β_4	0	**6**	6
Means			
Level	.00	**.00**	.00
Slope	-	**-.03***	-.03*
Variances			
Level	.08*	**.10***	.15*
Slope	-	**.003***	.003*
Covariance $_{\text{Level, Slope}}$	-	**-.01***	-.01*
Goodness-of-Fit			
χ^2 (df)	169.59 (47)	**57.21 (44)**	39.25 (42)
RMSEA [90% CI]	.062 [.052, .072]	**.021 [.000, .035]**	.000 [.000, .024]
CFI	.96	**1.00**	1.00
Fit changes, $\Delta\chi^2/\Delta df$	-	**112.37/3**	17.96/2
Aggression			
	No Growth	Linear Growth	Latent Basis
Slope			
β_1	**0**	0	0
β_2	**0**	2	-
β_3	**0**	4	-
β_4	**0**	6	6
Means			
Level	**.00**	.00	.00
Slope	**-**	-.01	-
Variances			
Level	**.09***	.12*	-
Slope	**-**	.002*	-
Covariance $_{\text{Level, Slope}}$	**-**	-.01*	-
Goodness-of-Fit			
χ^2 (df)	**146.96 (47)**	124.92 (44)	-
RMSEA [90% CI]	**.056 [.046, .067]**	.052 [.042, .063]	-
CFI	**.97**	.98	-
Fit changes, $\Delta\chi^2/\Delta df$	**-**	22.04	-
Depressed Mood			
	No Growth	**Linear Growth**	Latent Basis

Slope			
β_1	0	**0**	0
β_2	0	**2**	-
β_3	0	**4**	-
β_4	0	**6**	6
Means			
Level	.00	**.00**	-
Slope	-	**-.02***	-
Variances			
Level	.06*	**.06***	-
Slope	-	**.002***	-
Covariance $_{\text{Level, Slope}}$	-	**-.003***	-
Goodness-of-Fit			
χ^2 (df)	340.56 (47)	**283.47 (44)**	-
RMSEA [90% CI]	.096 [.087, .106]	**.090 [.080, .100]**	-
CFI	.88	**.90**	-
Fit changes, $\Delta\chi^2/\Delta df$	-	**57.09/3**	-
Surgency (age 12 to 16)			
	No Growth	**Linear Growth**	Latent Basis
Slope			
β_1	0	**0**	-
β_2	0	**2**	-
β_3	0	**4**	-
Means			
Level	.00	**.00**	-
Slope	-	**.03***	-
Variances			
Level	.07*	**.07***	-
Slope	-	**.003***	-
Covariance $_{\text{Level, Slope}}$	-	**-.00**	-
Goodness-of-Fit			
χ^2 (df)	168.15 (25)	**123.84 (22)**	-
RMSEA [90% CI]	.095 [.081, .109]	**.085 [.071, .100]**	-
CFI	.93	**.95**	-
Fit changes, $\Delta\chi^2/\Delta df$	-	**44.31/3**	-

	Surgency (age 10 to 16)		
	No Growth	**Linear Growth**	Latent Basis
Slope			
β_1	0	**0**	-
β_2	0	**2**	-
β_3	0	**4**	-
β_4	0	**6**	
Means			-
Level	.00	**.00**	-
Slope	-	**.03***	
Variances			-
Level	.05*	**.05***	-
Slope	-	**.002***	-
Covariance $_{\text{Level, Slope}}$	-	**-.00**	
Goodness-of-Fit			-
χ^2 (df)	349.99 (47)	**204.61 (44)**	-
RMSEA [90% CI]	.098 [.088, .108]	**.074 [.064, .084]**	-
CFI	.87	**.93**	-
Fit changes, $\Delta\chi^2/\Delta$df	-	**145.38/3**	-
	Affiliation		
	No Growth	**Linear Growth**	Latent Basis
Slope			
β_1	0	**0**	0
β_2	0	**2**	2.13
β_3	0	**4**	2.88
β_4	0	**6**	6
Means			
Level	.00	**.00**	.00
Slope	-	**-.01***	-.01*
Variances			
Level	.03*	**.03***	.03*
Slope	-	**.001***	.001*
Covariance $_{\text{Level, Slope}}$	-	**-.002***	-.002*
Goodness-of-Fit			
χ^2 (df)	244.23 (47)	**209.32 (44)**	204.33 (42)

RMSEA [90% CI]	.079 [.069, .089]	**.075 [.065, .085]**	.076 [.066, .086]
CFI	.89	**.91**	.91
Fit changes, $\Delta\chi^2/\Delta df$	-	**34.91/3**	4.99/2

Note. Values are unstandardized coefficients for the models. χ^2 = Chi-square. df=degrees of freedom. RMSEA=Root-mean-square-error of approximation. CI=90% confidence interval. CFI=Comparative fit index. For the Attention latent basis model to converge, the covariance between the level and slope had to be set to 0. The latent basis models for Aggression and Depressed Mood did not converge. * $p < .05$

Table S3.7

Correlations Between Temperament Trajectories

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. EC L																						
2. EC S	-.08																					
3. Act. L	.86	-.04																				
4. Act. S	-.12	.75	-.13																			
5. Att. L	.89	.15	.71	.02																		
6. Att. S	.48	.49	.40	.29	.63																	
7. Inh. L	.73	-.01	.47	-.02	.58	.32																
8. Inh. S	-.12	.66	-.06	.30	.06	.23	-.13															
9. NEM L	-.44	.07	-.31	.08	-.43	-.15	-.34	.05														
10. NEM S	.11	-.29	.07	-.23	.02	-.25	.04	-.18	-.53													
11. NEM (N) L	-.22	.04	-.10	.05	-.25	-.08	-.20	.03	.89	-.42												
12. NEM (N) S	.03	-.18	.00	-.15	-.02	-.18	-.02	-.08	-.44	.88	-.48											
13. Agg. L	-.58	-.16	-.51	-.08	-.53	-.40	-.47	-.12	.52	.05	.28	.06										
14. Dep. L	-.45	-.01	-.32	.03	-.47	-.28	-.29	-.01	.74	-.12	.55	-.03	.43									
15. Dep. S	-.03	-.39	-.02	-.27	-.15	-.27	-.03	-.25	-.08	.61	.01	.40	.18	.07								
16. Fear L	.02	.00	.12	.05	-.05	-.03	-.04	.01	.57	-.13	.74	-.17	-.01	.43	.12							
17. Fear S	-.06	-.06	-.09	-.03	-.07	-.10	-.07	-.00	-.34	.62	-.41	.74	.09	-.06	.29	-.34						
18. Fru. L	-.40	.04	-.30	.03	-.40	-.20	-.33	.03	.82	-.25	.83	-.25	.49	.55	.05	.36	-.11					
19. Fru. S	.10	-.30	.07	-.27	.01	-.20	.04	-.15	-.29	.75	-.28	.79	.07	.01	.49	.00	.35	-.22				
20. Sur. L	.02	.01	-.07	-.08	.06	.05	.04	.01	-.01	.04	-.05	.06	.17	-.06	-.01	-.24	.06	.12	.07			
21. Sur. S	.06	.09	.02	.00	.06	.05	.08	.05	-.03	.02	-.01	.01	-.00	-.06	-.04	-.06	-.03	.03	.06	.09		
22. Aff. L	.38	.03	.30	-.01	.33	.19	.33	-.04	.10	-.07	.23	-.10	-.19	.04	-.08	.31	-.16	.07	-.04	.20	.07	
23. Aff. S	.02	.31	-.04	.15	.09	.17	.07	.28	-.14	.13	-.15	.19	-.07	-.08	.02	-.12	.21	-.08	.10	.17	.23	-.16

Note. L = Level. S Slope. EC = Effortful Control. NEM = Negative Emotionality. NEM (N) = Negative Emotionality (Narrow). The Surgency trajectory is from age 12 to 16. Values more extreme than +/- .08 are significant at $p < .05$

Table S3.8
Model Fit Indices for Gender Multiple Group Models (Boys vs. Girls)

	χ^2 (df)	CFI	RMSEA [90% CI]
Effortful Control			
Parameters free	308.72(194)	.98	.042 [.033, .050]
Parameters constrained	314.48(198)	.98	.042 [.033, .050]
Activation Control			
Parameters free	121.11(92)	.99	.031 [.012, .045]
Parameters constrained	125.12(96)	.99	.030 [.011, .044]
Attention			
Parameters free	122.98(90)	.98	.033 [.016, .047]
Parameters constrained	124.03(93)	.99	.031 [.014, .045]
Inhibitory Control			
Parameters free	283.34(92)	.87	.079 [.068, .089]
Parameters constrained	377.62(96)	.80	.093 [.084, .103]
Negative Emotionality			
Parameters free	367.99(190)	.97	.053 [.045, .061]
Parameters constrained	404.13(194)	.97	.057 [.049, .064]
Negative Emotionality (Narrow)			
Parameters free	316.01(190)	.97	.044 [.036, .053]
Parameters constrained	352.92(194)	.96	.049 [.041, .057]
Fear			
Parameters free	188.42(95)	.96	.054 [.043, .065]
Parameters constrained	222.87(99)	.94	.061 [.050, .072]
Frustration			
Parameters free	113.48(92)	.99	.026 [.000, .041]
Parameters constrained	143.95(96)	.99	.038 [.025, .051]
Aggression			
Parameters free	239.73(98)	.96	.066 [.055, .076]
Parameters constrained	240.64(99)	.96	.065 [.055, .076]
Depressed Mood			
Parameters free	361.11(92)	.89	.093 [.083, .103]
Parameters constrained	387.83(96)	.88	.095 [.085, .105]
Surgency			
Parameters free	175.35(48)	.94	.091 [.077, .106]
Parameters constrained	179.29(52)	.94	.088 [.074, .102]
Affiliation			
Parameters free	281.44(92)	.89	.078 [.068, .089]
Parameters constrained	292.67(96)	.88	.078 [.068, .088]

Note. χ^2 = Chi-square test statistic; df = degrees of freedom; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; 90%CI = 90% confidence interval. Parameters that varied between models are means and variances of the level and slope. The Surgency trajectory is from age 12 to 16.

Table S9

Associations Between Temperament Trajectories and Young Adult Anxiety/Depression Controlling for Earlier Anxiety/Depression

	General Distress			**Anxious Arousal**			**Anhedonic Depression**		
	Age 19	Age 21	Young Adult	Age 19	Age 21	Young Adult	Age 19	Age 21	Young Adult
EC									
Level	-.04 (p = .326)		-.01 (p = .766)	-.07 (p = .111)			-.07 (p = .086)	.05 (p = .253)	-.02 (p = .608)
Slope	-.09 (p = .028)		-.07 (p = .075)	-.10 (p = .023)			-.04 (p = .292)	-.07 (p = .132)	-.07 (p = .070)
Activation Control									
Level	-.03 (p = .424)		.01 (p = .882)	-.11 (p = .013)			-.01 (p = .765)		.02 (p = .552)
Slope	-.09 (p = .032)		-.05 (p = .219)	-.08 (p = .057)			-.04 (p = .350)		-.04 (p = .318)
Attention									
Level								.10 (p = .072)	.01 (p = .862)
Slope								-.11 (p = .056)	-.09 (p = .073)
Inhibitory Control									
Level								.01 (p = .801)	-.03 (p = .382)
Slope								-.07 (p = .117)	-.05 (p = .184)
NEM									
Level	.10 (p = .047)	.06 (p = .261)	.10 (p = .047)	.11 (p = .031)		.10 (p = .036)	.10 (p = .028)	.05 (p = .267)	.09 (p = .051)
Slope	.09 (p = .067)	.05 (p = .355)	.09 (p = .061)	.10 (p = .051)		.11 (p = .018)	.08 (p = .103)	.08 (p = .092)	.10 (p = .039)
NEM (narrow)									

Level	.05 (p = .244)		.03 (p = .443)						
Slope	.08 (p = .096)		.05 (p = .291)						
Fear									
Level	.04 (p = .343)		.01 (p = .757)						
Slope	.11 (p = .011)		.08 (p = .071)						
Frustration									
Level	.07 (p = .088)		.07 (p = .084)	.09 (p = .036)		.11 (p = .010)			
Slope	.06 (p = .138)		.05 (p = .204)	-.01 (p = .773)		.02 (p = .583)			
Aggression									
Level				.06 (p = .142)				.07 (p = .104)	
Depressed Mood									
Level	**.14* (p = .001)**	**.13* (p = .004)**	**.15* (p < .001)**	**.13* (p = .003)**		.10 (p = .017)	**.14* (p < .001)**	.06 (p = .148)	**.11* (p = .005)**
Slope	**.14* (p = .003)**	.07 (p = .150)	.10 (p = .021)	.08 (p = .047)		.09 (p = .029)	.09 (p = .035)	.09 (p = .049)	.10 (p = .020)
Surgency									
Level							-.03 (p = .430)		-.03 (p = .510)
Slope							**-.13* (p= .001)**		-.08 (p= .049)
Affiliation									
Level							-.08 (p = .033)	-.01 (p = .812)	-.08 (p = .054)
Slope							**-.13* (p = .001)**	-.09 (p = .033)	**-.15* (p <.001)**

Note. Values are standardized coefficients from multiple regressions where anxiety/depression were regressed on the level and slope of each temperament domain. * $p < .006$

Table S10

Associations Between Temperament Trajectories and Young Adult Anxiety/Depression Controlling for Mother Anxiety/Depression

	General Distress			**Anxious Arousal**			**Anhedonic Depression**		
	Age 19	Age 21	Young Adult	Age 19	Age 21	Young Adult	Age 19	Age 21	Young Adult
<u>EC</u>									
Level	-.06 ($p = .134$)		-.04 ($p = .308$)	-.09 ($p = .028$)			-.12 ($p = .007$)	.01 ($p = .834$)	-.07 ($p = .109$)
Slope	**-.17* ($p < .001$)**		**-.15* ($p < .001$)**	**-.12* ($p = .004$)**			**-.14* ($p < .001$)**	**-.16* ($p < .001$)**	**-.18* ($p < .001$)**
Activation Control									
Level	-.06 ($p = .149$)		-.03 ($p = .499$)	**-.13* ($p = .002$)**			-.05 ($p = .257$)		-.01 ($p = .750$)
Slope	**-.15* ($p < .001$)**		**-.11* ($p = .006$)**	-.10 ($p = .016$)			-.11 ($p = .008$)		**-.12* ($p = .006$)**
Attention									
Level								.09 ($p = .125$)	.00 ($p = .961$)
Slope								**-.19* ($p < .001$)**	**-.19* ($p < .001$)**
Inhibitory Control									
Level								-.01 ($p = .884$)	-.06 ($p = .164$)
Slope								-.11 ($p = .012$)	-.11 ($p = .010$)
<u>NE</u>									
Level	**.19* ($p < .001$)**	.14 ($p = .008$)	**.19* ($p < .001$)**	**.15* ($p = .003$)**		**.16* ($p < .001$)**	**.16* ($p = .001$)**	.11 ($p = .028$)	**.15* ($p = .002$)**
Slope	**.23* ($p < .001$)**	**.16* ($p = .002$)**	**.22* ($p < .001$)**	**.18* ($p < .001$)**		**.20* ($p < .001$)**	**.17* ($p < .001$)**	**.16* ($p = .002$)**	**.19* ($p < .001$)**
<u>NE (narrow)</u>									

Level	**.14* (p = .004)**		.12 (p = .014)						
Slope	**.20* (p < .001)**		**.16* (p < .001)**						
Fear									
Level	.12 (p = .008)		.08 (p = .056)						
Slope	**.17* (p < .001)**		**.13* (p = .002)**						
Frustration									
Level	.11 (p = .014)		.11 (p = .012)	.10 (p = .019)		**.14* (p = .001)**			
Slope	**.17* (p < .001)**		**.16* (p < .001)**	.05 (p = .286)		.09 (p = .024)			
Aggression									
Level				.11 (p = .010)				.11 (p = .017)	
Depressed Mood									
Level	**.17* (p < .001)**	**.16* (p < .001)**	**.19* (p < .001)**	**.14* (p < .001)**		**.13* (p = .001)**	**.17* (p < .001)**	.11 (p = .014)	**.15* (p < .001)**
Slope	**.24* (p < .001)**	**.17* (p < .001)**	**.21* (p < .001)**	**.12* (p = .003)**		**.15* (p < .001)**	**.20* (p < .001)**	**.19* (p < .001)**	**.21* (p < .001)**
Surgency									
Level							-.06 (p = .156)		-.06 (p = .177)
Slope							**-.16* (p < .001)**		**-.12 (p = .004)**
Affiliation									
Level							**-.14* (p < .001)**	-.05 (p = .235)	**-.13* (p = .001)**
Slope							**-.19* (p < .001)**	**-.16*(p < .001)**	**-.21* (p < .001)**

Note. Values are standardized coefficients from multiple regressions where anxiety/depression were regressed on the level and slope of each temperament domain. * p < .006

Table S11

Associations Between Temperament Domain Trajectories and Young Adult Anxiety/Depression with Gender as Moderator

	General Distress			**Anxious Arousal**			**Anhedonic Depression**		
	Age 19	Age 21	Young Adult	Age 19	Age 21	Young Adult	Age 19	Age 21	Young Adult
<u>EC</u>									
Level: Level	-.01 (p = .848)	.07 (p = .226)	.03 (p = .553)	-.13 (p = .030)	-.01 (p = .838)	.09 (p = .118)	-.13 (p = .028)	.05 (p = .437)	-.05 (p = .402)
Slope	**-.18* (p < .001)**	-.11 (p = .009)	**-.16* (p < .001)**	**-.12* (p = .005)**	-.05 (p = .205)	.11 (p = .008)	**-.15* (p < .001)**	**-.17*(p < .001)**	**-.19* (p < .001)**
Gender	-.11 (p = .010)	-.10 (p = .022)	-.11 (p = .008)	-.05 (p = .222)	-.03 (p = .471)	.05 (p = .259)	**-.11* (p = .005)**	-.10 (p = .021)	**-.12* (p = .003)**
Level*Gender	-.10 (p = .074)	-.12 (p = .052)	-.14 (p = .016)	.02 (p = .680)	-.04 (p = .520)	.01 (p = .803)	-.00 (p = .937)	-.09 (p = .116)	-.06 (p = .253)
Slope: Level	-.08 (p = .054)	-.01 (p = .867)	-.06 (p = .149)	-.11 (p = .010)	-.04 (p = .397)	-.10 (p = .018)	**-.13* (p = .002)**	-.02 (p = .665)	-.09 (p = .025)
Slope	**-.22* (p < .001)**	-.12 (p = .031)	**-.20* (p < .001)**	**-.16* (p = .005)**	-.09 (p = .134)	-.15 (p = .008)	**-.18* (p = .001)**	**-.17* (p = .003)**	**-.21* (p < .001)**
Gender	-.07 (p = .144)	-.09 (p = .089)	-.08 (p = .111)	-.02 (p = .633)	-.01 (p = .879)	-.02 (p = .720)	-.10 (p = .050)	-.10 (p = .054)	-.11 (p = .027)
Slope*Gender	.08 (p = .191)	.03 (p = .658)	.07 (p = .241)	.06 (p = .324)	.05 (p = .386)	.07 (p = .285)	.04 (p = .469)	.00 (p = .952)	.03 (p = .600)
<u>NEM</u>									
Level: Level	.16 (p = .011)	.12 (p = .072)	.16 (p = .008)	**.20* (p = .002)**	.08 (p = .210)	**.19* (p = .002)**	.15 (p = .017)	.08 (p = .195)	.14 (p = .027)
Slope	**.21* (p < .001)**	**.15* (p = .003)**	**.21* (p < .001)**	**.16* (p < .001)**	.11 (p = .028)	**.20* (p < .001)**	**.17* (p < .001)**	**.17* (p < .001)**	**.19* (p = .000)**
Gender	-.04 (p = .310)	-.06 (p = .141)	-.05 (p = .225)	.00 (p = .949)	-.00 (p = .934)	.01 (p = .794)	-.06 (p = .177)	-.06 (p = .165)	-.06 (p = .129)
Level*Gender	.04 (p = .465)	.01 (p = .831)	.02 (p = .695)	-.04 (p = .544)	.01 (p = .916)	.02 (p = .669)	.02 (p = .712)	.04 (p = .503)	.02 (p = .677)

Slope: Level	**.19* (*p* < .001)**	.12 (*p* = .015)	**.18* (*p* < .001)**	**.17* (*p* < .001)**	.08 (*p* = .097)	**.17* (*p* < .001)**	**.16* (*p* < .001)**	.11 (*p* = .033)	**.15* (*p* = .001)**
Slope	**.25* (*p* < .001)**	**.19* (*p* = .003)**	**.25* (*p* < .001)**	.16 (*p* = .013)	.14 (*p* = .027)	**.20* (*p* = .001)**	**.21* (*p* < .001)**	.22 (*p* = .000)	**.24* (*p* < .001)**
Gender	-.09 (*p* = .202)	-.13 (*p* = .077)	-.11 (*p* = .125)	.01 (*p* = .834)	-.05 (*p* = .513)	.01 (*p* = .939)	-.11 (*p* = .115)	-.15 (*p* = .049)	-.14 (*p* = .044)
Slope*Gender	-.07 (*p* = .420)	-.10 (*p* = .262)	-.08 (*p* = .318)	.02 (*p* = .848)	-.07 (*p* = .458)	-.01 (*p* = .914)	-.08 (*p* = .346)	-.12 (*p* = .156)	-.11 (*p* = .172)
<u>PEM</u>									
Level: Level	.10 (*p* = .086)	.10 (*p* = .112)	.09 (*p* = .105)	.15 (*p* = .009)	-.01 (*p* = .848)	.07 (*p* = .217)	.05 (*p* = .363)	.09 (*p* = .148)	.06 (*p* = .268)
Slope	-.11 (*p* = .010)	-.03 (*p* = .440)	-.08 (*p* = .065)	-.07 (*p* = .077)	.01 (*p* = .905)	-.04 (*p* = .282)	**-.19* (*p* < .001)**	-.07 (*p* = .108)	**-.15* (*p* < .001)**
Gender	-.10 (*p* = .018)	-.09 (*p* = .038)	-.10 (*p* = .018)	-.05 (*p* = .247)	-.02 (*p* = .637)	-.04 (*p* = .395)	-.09 (*p* = .029)	-.09 (*p* = .039)	-.10 (*p* = .018)
Level*Gender	-.09 (*p* = .123)	-.14 (*p* = .016)	-.14 (*p* = .018)	-.06 (*p* = .282)	-.02 (*p* = .696)	-.05 (*p* = .349)	-.14 (*p* = .015)	-.12 (*p* = .040)	**-.16* (*p* = .005)**
Slope: Level	.04 (*p* = .391)	-.01 (*p* = .822)	-.01 (*p* = .893)	.11 (*p* = .009)	-.03 (*p* = .428)	.03 (*p* = .463)	-.05 (*p* = .245)	-.01 (*p* = .900)	-.05 (*p* = .198)
Slope	-.08 (*p* = .139)	.04 (*p* = .524)	-.02 (*p* = .722)	-.09 (*p* = .128)	.09 (*p* = .109)	.01 (*p* = .886)	-.15 (*p* = .009)	.03 (*p* = .610)	-.07 (*p* = .185)
Gender	-.08 (*p* = .132)	-.03 (*p* = .570)	-.05 (*p* = .335)	-.06 (*p* = .244)	.06 (*p* = .272)	.01 (*p* = .833)	-.06 (*p* = .282)	-.00 (*p* = .969)	-.04 (*p* = .506)
Slope*Gender	-.03 (*p* = .603)	-.11 (*p* = .100)	-.09 (*p* = .173)	.03 (*p* = .694)	-.15 (*p* = .024)	-.09 (*p* = .173)	-.06 (*p* = .313)	-.16 (*p* = .015)	-.12 (*p* = .067)

Note. Values are standardized coefficients from multiple regressions where anxiety/depression were regressed on the level and slope of each temperament domain, gender, and the interaction between temperament and gender (separately for level/gender interaction and slope/gender interaction). * $p < .006$

References

Aizpitarte, A., Atherton, O. E., & Robins, R. W. (2017). The co-development of relational aggression and disruptive behavior symptoms from late childhood through adolescence. *Clinical Psychological Science*, *5*, 866–873. https://doi.org/10.1177/2167702617708231

Akiskal, H. S., Mendlowicz, M. V., Jean-Louis, G., Rapaport, M. H., Kelsoe, J. R., Gillin, J. C., & Smith, T. L. (2005). TEMPS-A: Validation of a short version of a self-rated instrument designed to measure variations in temperament. *Journal of Affective Disorders*, *85*(1–2), 45–52. https://doi.org/10.1016/j.jad.2003.10.012

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237. https://doi.org/10.1016/j.cpr.2009.11.004

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, *13*, 61. https://doi.org/10.2307/270718

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545–557. https://doi.org/10.1037/0021-843X.112.4.545

Anderson, E., & Hope, D. (2008). A review of the tripartite model for understanding the link between anxiety and depression in youth. *Clinical Psychology Review*, *28*(2), 275–287. https://doi.org/10.1016/j.cpr.2007.05.004

Anthony, J. L., Lonigan, C. J., Hooe, E. S., & Phillips, B. M. (2002). An affect-based, hierarchical model of temperament and its relations with internalizing symptomatology. *Journal of Clinical Child and Adolescent Psychology, 31*(4), 480-490. https://doi.org/10.1207/S15374424JCCP3104_7

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology: Recommendations for increasing replicability. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Atherton, O. E., Lawson, K. M., & Robins, R. W. (2020). The development of effortful control from late childhood to young adulthood. *Journal of Personality and Social Psychology, 119*(2), 417-456.

Atherton, O. E., Lawson, K. M., Ferrer, E., & Robins, R. W. (2020). The role of effortful control in the development of ADHD, ODD, and CD symptoms. *Journal of Personality and Social Psychology, 118*(6), 1226–1246. https://doi.org/10.1037/pspp0000243

Atherton, O. E., Schofield, T. J., Sitka, A., Conger, R. D., & Robins, R. W. (2016). Unsupervised self-care predicts conduct problems: The moderating roles of hostile aggression and gender. *Journal of Adolescence*, *48*, 1–10. https://doi.org/10.1016/j.adolescence.2016.01.001

Atherton, O. E., Tackett, J. L., Ferrer, E., & Robins, R. W. (2017). Bidirectional pathways between relational aggression and temperament from late childhood to adolescence. *Journal of Research in Personality*, *67*, 75–84. https://doi.org/10.1016/j.jrp.2016.04.005

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

Baertschi, M., Costanza, A., Canuto, A., & Weber, K. (2018). The function of personality in suicidal ideation from the perspective of the interpersonal-psychological theory of suicide. *International Journal of Environmental Research and Public Health*, *15*(4), 636. https://doi.org/10.3390/ijerph15040636

Baetens, I., Claes, L., Willem, L., Muehlenkamp, J., & Bijttebier, P. (2011). The relationship between non-suicidal self-injury and temperament in male and female adolescents based on child- and parent-report. *Personality and Individual Differences*, *50*(4), 527–530. https://doi.org/10.1016/j.paid.2010.11.015

Behrendt, S. (2014). lm.beta: Add Standardized Regression Coefficients to lm-Objects. R package version 1.5-1. https://CRAN.R-project.org/package=lm.beta.

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*(12), 2530–2540. https://doi.org/10.1177/0956797613498396

Bleidorn, W., Schwaba, T., Zheng, A., Hopwood, C.J., Sosa, S., Roberts, B.W., & Briley, D.A. (2021). Personality stability and change: A meta-analysis of longitudinal studies. [Manuscript submitted for publication.]

Blüml, V., Kapusta, N. D., Doering, S., Brähler, E., Wagner, B., & Kersting, A. (2013). Personality factors and suicide risk in a representative sample of the German general population. *PLoS ONE*, *8*(10), e76646. https://doi.org/10.1371/journal.pone.0076646

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 5*3, 605-634.

Borghuis, J., Denissen, J. J. A., Oberski, D., Sijtsma, K., Meeus, W. H. J., Branje, S., Koot, H. M., & Bleidorn, W. (2017). Big Five Personality stability, change, and codevelopment across adolescence and early adulthood. *Journal of Personality and Social Psychology, 113*(4), 641-657.

Borsboom, D. (2008). Latent variable theory. *Measurement, 6*, 25-53.

Bould, H., Araya, R., Pearson, R. M., Stapinski, L., Carnegie, R., & Joinson, C. (2014). Association between early temperament and depression at 18 years. *Depression and Anxiety*, *31*(9), 729–736. https://doi.org/10.1002/da.22294

Bouma, E. M. C., Ormel, J., Verhulst, F. C., & Oldehinkel, A. J. (2008). Stressful life events and depressive problems in early adolescent boys and girls: The influence of parental depression, temperament and family environment. *Journal of Affective Disorders*, *105*(1–3), 185–193. https://doi.org/10.1016/j.jad.2007.05.007

Brandes, C. M., Herzhoff, K., Smack, A. J., & Tackett, J. L. (2019). The *p* factor and the *n* factor: Associations between the general factors of psychopathology and neuroticism in children. *Clinical Psychological Science, 7*(6), 1266-1284.

Brandes, C. M., Kushner, S. C., Herzhoff, K., & Tackett, J. L. (2022). Facet-level personality development in the transition to adolescence: Maturity, disruption, and gender differences. *Journal of Personality and Social Psychology*.

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

Bravo, M., Woodbury-Farina, M., Canino, G.J., & Rubio-Stipec, M. (1993). The Spanish translation and cultural adaptation of the Diagnostic Interview Schedule for Children (DISC) in Puerto Rico. *Culture, Medicine, & Psychiatry, 17*, 329-344.

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*(5), 579–594. https://doi.org/10.1007/s10802-005-6739-2

Brener, N. D., Kann, L., Kinchen, S. A., Grunbaum, J. A., Whalen, L., Eaton, D., Hawkins, J., & Ross, J. G. (2004). Methodology of the youth risk behavior surveillance system. 53, 1-13.

Breslau, J., Lane, M., Sampson, N., & Kessler, R. C. (2008). Mental disorders and subsequent educational attainment in a US national sample. *Journal of Psychiatric Research*, *42*(9), 708–716. https://doi.org/10.1016/j.jpsychires.2008.01.016

Brezo, J., Paris, J., Tremblay, R., Vitaro, F., Zoccolillo, M., Hébert, M., & Turecki, G. (2006a). Personality traits as correlates of suicide attempts and suicidal ideation in young adults. *Psychological Medicine*, *36*(2), 191–202. https://doi.org/10.1017/S0033291705006719

Brezo, J., Paris, J., & Turecki, G. (2006b). Personality traits as correlates of suicidal ideation, suicide attempts, and suicide completions: A systematic review. *Acta Psychiatrica Scandinavica*, *113*(3), 180–206. https://doi.org/10.1111/j.1600-0447.2005.00702.x

Bridges, A. J., Ledesma, R. J., Guzman, L. E., Berman, I. S., & Diaz Benitez, D. E. (2021). A systematic review and meta-analysis of the relation between acculturation and depression in Latinx adults. *Journal of Latinx Psychology*, *9*(3), 232–257. https://doi.org/10.1037/lat0000189

Cain, N. M., Ansell, E. B., Wright, A. G. C., Hopwood, C. J., Thomas, K. M., Pinto, A., Markowitz, J. C., Sanislow, C. A., Zanarini, M. C., Shea, M. T., Morey, L. C., McGlashan, T. H., Skodol, A. E., & Grilo, C. M. (2012). Interpersonal pathoplasticity in the course of major depression. *Journal of Consulting and Clinical Psychology*, *80*(1), 78–86. https://doi.org/10.1037/a0026433

Canino, G., & Roberts, R. E. (2001). Suicidal behavior among Latino youth. *Suicide and Life-Threatening Behavior*, *31*, 122–131. https://doi.org/10.1521/suli.31.1.5.122.24218

Capaldi, D. M. & Rothbart, M. K. (1992). Development and validation of an early adolescent temperament measure. *Journal of Early Adolescence, 12,* 153-173.

Carver, C. S. (2005). Impulse and constraint: Perspectives from personality psychology, convergence with theory in other areas, and potential for integration. *Personality and Social Psychology Review*, *9*(4), 312–333. https://doi.org/10.1207/s15327957pspr0904_2

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p Factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, *53*(11), 1033. https://doi.org/10.1001/archpsyc.1996.01830110071009

Centers for Disease Control and Prevention (CDC). (2017). Web-based injury statistics query and reporting system (WISQARS). Available from: https://www.cdc.gov/injury/wisqars/fatal.html

Champely, S. (2018). pwr: Basic Functions for Power Analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chioqueta, A. P., & Stiles, T. C. (2005). Personality traits and the development of depression, hopelessness, and suicide ideation. *Personality and Individual Differences*, *38*(6), 1283–1291. https://doi.org/10.1016/j.paid.2004.08.010

Clark, D. A., Donnellan, M. B., & Robins, R. W. (2018). Personality traits and parent–adolescent interactions: An observational study of Mexican origin families. *Journal of Family Psychology*, *32*(4), 544–551. https://doi.org/10.1037/fam0000408

Clark, D. A., Donnellan, M. B., Durbin, C. E., Nuttall, A. K., Hicks, B. M., & Robins, R. W. (2020). Sex, drugs, and early emerging risk: Examining the association between sexual debut and substance use across adolescence. *PLOS ONE, 15*(2). doi: 10.1371/journal.pone.0228432

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. https://doi.org/10.1016/j.adolescence.2015.02.010

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316-336.

Clark, L. A., & Watson, D. (2008). *Temperament: An organizing paradigm for trait psychology.* In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 265–286). The Guilford Press.

Compas, B. E., Connor-Smith, J., & Jaser, S. S. (2004). Temperament, stress reactivity, and coping: Implications for depression in childhood and adolescence. *Journal of Clinical Child & Adolescent Psychology*, *33*(1), 21–31. https://doi.org/10.1207/S15374424JCCP3301_3

Coplan, R. J., & Bullock, A. (2012). *Temperament and peer relationships.* In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (p. 442–461). The Guilford Press.

Corral-Frías, N. S., Velardez Soto, S. N., Frías-Armenta, M., Corona-Espinosa, A., & Watson, D. (2019). Concurrent validity and reliability of two short forms of the Mood and Anxiety Symptom Questionnaire in a student sample from northwest Mexico. *Journal of Psychopathology and Behavioral Assessment*, *41*(2), 304–316. https://doi.org/10.1007/s10862-019-09738-x

Costello, E. J., Edelbrock, C. S., & Costello, A. J. (1985). Validity of the NIMH diagnostic interview schedule for children: A comparison between psychiatric and pediatric referrals. *Journal of Abnormal Child Psychology, 13*, 579-595. http://dx.doi.org/10.1007/BF00923143

Cramer, R. J., Garza, M. J., Henderson, C. E., Ribeiro, J. D., Silva, C., Smith, A. R., Joiner, T. E., & White, J. (2012). A trait-interpersonal perspective on suicide risk in criminal offenders. *Archives of Suicide Research*, *16*(4), 334–347. https://doi.org/10.1080/13811118.2013.722057

Crockett, L. J., Carlo, G., Wolff, J. M., & Hope, M. O. (2013). The role of pubertal timing and temperamental vulnerability in adolescents' internalizing symptoms. *Development and Psychopathology*, *25*(2), 377–389. https://doi.org/10.1017/S0954579412001125

Curtin, S. C., & Heron, M. (2019). Death rates due to suicide and homicide among persons aged 10-24: United States, 2000-2017. National Center for Health Statistics (NCHS) Data Brief No. 352.

Damian, R. I., Atherton, O. E., Lawson, K. M., & Robins, R. W. (2020). The co-development of chores and effortful control among Mexican-origin youth and prospective work outcomes. *Journal of Research in Personality*, *84*, 103883. https://doi.org/10.1016/j.jrp.2019.103883

Davis, S., Votruba-Drzal, E., & Silk, J. S. (2015). Trajectories of internalizing symptoms from early childhood to adolescence: Associations with temperament and parenting. *Social Development*, *24*(3), 501–520. https://doi.org/10.1111/sode.12105

Dawood, S., Thomas, K. M., Wright, A. G. C., & Hopwood, C. J. (2013). Heterogeneity of interpersonal problems among depressed young adults: Associations with substance abuse and pathological personality traits. *Journal of Personality Assessment*, *95*(5), 513–522. https://doi.org/10.1080/00223891.2013.781031

De Bolle, M., & Tackett, J. L. (2013). Anchoring bullying and victimization in children within a five-factor model-based person-centred framework: Bullying, victimization and personality: a person-centered approach. *European Journal of Personality*, *27*(3), 280–289. https://doi.org/10.1002/per.1901

De Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology*, *91*(3), 538–552. https://doi.org/10.1037/0022-3514.91.3.538

De Pauw, S. S. W., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry & Human Development*, *41*(3), 313–329. https://doi.org/10.1007/s10578-009-0171-8

Degnan, K. A., Almas, A. N., Henderson, H. A., Hane, A. A., Walker, O. L., & Fox, N. A. (2014). Longitudinal trajectories of social reticence with unfamiliar peers across early childhood. *Developmental Psychology*, *50*(10), 2311–2323. https://doi.org/10.1037/a0037751

Depue, R. A., Luciana, M., Arbisi, P., Collins, P., & Leon, A. (1994). Dopamine and the structure of personality: Relation of agonist-induced dopamine activity to positive emotionality. *Journal of Personality and Social Psychology*, *67*, 485–498. https://doi.org/10.1037/0022-3514.67.3.485

Duarté-Vélez, Y., & Bernal, G. (2007). Suicide behavior among Latino and Latina adolescents: Conceptual and methodological issues. *Death Studies, 31*:5, 435-455, DOI: 10.1080/07481180701244579

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. https://doi.org/10.13023/edp.tool.01 [Also available at http://sites.education.uky.edu/apslab/resources/]

Dumais, A., Lesage, A. D., Alda, M., Rouleau, G., Dumont, M., Chawky, N., Roy, M., Mann, J. J., Benkelfat, C., & Turecki, G. (2005). Risk factors for suicide completion in major depression: A case-control study of impulsive and aggressive behaviors in men. *American Journal of Psychiatry*, *162*(11), 2116–2124. https://doi.org/10.1176/appi.ajp.162.11.2116

Durbin, C. E., & Hicks, B. M. (2014). Personality and psychopathology: A stagnant field in need of development. *European Journal of Personality*, *28*(4), 362–386. https://doi.org/10.1002/per.1962

Durbin, C. E., Klein, D. N., Hayden, E. P., Buckley, M. E., & Moerk, K. C. (2005). Temperamental emotionality in preschoolers and parental mood disorders. *Journal of Abnormal Psychology*, *114*(1), 28–37. https://doi.org/10.1037/0021-843X.114.1.28

Eaton, D.K., Foti, K.E., Brener, N.D., Crosby, A.E., Flores, G., & Kann, L. (2011). Associations between risk behaviors and suicidal ideation and suicide attempts: Do racial/ethnic variations in associations account for increased risk of suicidal behaviors among Hispanic/Latina 9th- to 12th-grade female students? *Archives of suicide research: official journal of the International Academy for Suicide Research, 15*(2), 113-26.

Edwards, M. J., & Holden, R. R. (2001). Coping, meaning in life, and suicidal manifestations: Examining gender differences. *Journal of Clinical Psychology*, *57*(12), 1517–1534. https://doi.org/10.1002/jclp.1114

Eisenberg, N., Ma, Y., Chang, L., Zhou, Q., West, S. G., & Aiken, L. (2007). Relations of effortful control, reactive undercontrol, and anger to Chinese children’s adjustment. *Development and Psychopathology*, *19*(02). https://doi.org/10.1017/S0954579407070198

Eisenberg, N., Valiente, C., Spinrad, T. L., Cumberland, A., Liew, J., Reiser, M., Zhou, Q., & Losoya, S. H. (2009). Longitudinal relations of children's effortful control, impulsivity, and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. *Developmental Psychology*, *45*(4), 988–1008. https://doi.org/10.1037/a0016213

Ellis, L. K., & Rothbart, M. K. (2001). Revision of the Early Adolescent Temperament Questionnaire. Poster presented at the Biennial Meeting of the Society for Research in Child Development in Minneapolis, Minnesota.

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, *132*(1), 33–72. https://doi.org/10.1037/0033-2909.132.1.33

Enns, M. W., Cox, B. J., & Inayatulla, M. (2003). Personality predictors of outcome for adolescents hospitalized for suicidal ideation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(6), 720–727. https://doi.org/10.1097/01.CHI.0000046847.56865.B0

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality*, *41*(4), 868–888. https://doi.org/10.1016/j.jrp.2006.11.002

Fergusson, D. M., Beautrais, A. L., & Horwood, L. J. (2003). Vulnerability and resiliency to suicidal behaviours in young people. *Psychological Medicine*, *33*(1), 61–73. https://doi.org/10.1017/S0033291702006748

Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2017). Replicability and other features of a high-quality science: Toward a balanced and empirical approach. *Journal of Personality and Social Psychology*, *113*(2), 244–253. https://doi.org/10.1037/pspi0000075

Fino, E., Iliceto, P., Sabatello, U., Petrucci, F., & Candilera, G. (2014). Self/other perception mediates between personality and suicidal ideation in young adults. *European Journal of Psychiatry, 28*, 102-113.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*, 370–378. https://doi.org/10.1177/1948550617693063

Fortuna, L. R., Álvarez, K., Ramos Ortiz, Z., Wang, Y., Mozo Alegría, X., Cook, B. L., & Alegría, M. (2016). Mental health, migration stressors and suicidal ideation among Latino immigrants in Spain and the United States. *European Psychiatry*, *36*, 15–22. https://doi.org/10.1016/j.eurpsy.2016.03.001

Frances, A., Fyer, M., & Clarkin, J. (1986). Personality and suicide. *Annals New York Academy of Sciences, 487*, 281-293.

Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2019). Hmisc: Harrell Miscellaneous. R package version 4.2-0. https://CRAN.R-project.org/package=Hmisc

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 187–232. https://doi.org/10.1037/bul0000084

Frick, P. J. (2004). Integrating research on temperament and childhood psychopathology: Its pitfalls and promise. *Journal of Clinical Child & Adolescent Psychology*, *33*(1), 2–7. https://doi.org/10.1207/S15374424JCCP3301_1

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

from age 10 to age 25 as related to trajectory of suicide attempt in a community cohort. *Suicide & life-threatening behavior, 41*(2), 180–192. doi:10.1111/j.1943-278X.2011.00017.x

García Coll, C. E., & Marks, A. K. E. (2012). *The immigrant paradox in children and adolescents: Is becoming American a developmental risk?* (pp. xiv-328). American Psychological Association.

Glenn, C. R., Lanzillo, E. C., Esposito, E. C., Santee, A. C., Nock, M. K., & Auerbach, R. P. (2017). Examining the course of suicidal and nonsuicidal self-injurious thoughts and behaviors in outpatient and inpatient adolescents. *Journal of abnormal child psychology, 45*(5), 971–983. doi:10.1007/s10802-016-0214-0

Goldston, D. B., Molock, S. D., Whitbeck, L. B., Murakami, J. L., Zayas, L. H., & Hall, G. C. (2008). Cultural considerations in adolescent suicide prevention and psychosocial treatment. *The American psychologist, 63*(1), 14–31. doi:10.1037/0003-066X.63.1.14

Göllner, R., Roberts, B. W., Damian, R. I., Lüdtke, O., Jonkmann, K., & Trautwein, U. (2017). Whose "storm and stress" is it? Parent and child reports of personality development in the transition to early adolescence. *Journal of Personality*, *85*(3), 376–387. https://doi.org/10.1111/jopy.12246

Gomez, J., Miranda, R., & Polanco, L. (2011). Acculturative stress, perceived discrimination, and vulnerability to suicide attempts among emerging adults. *Journal of Youth and Adolescence*, *40*(11), 1465–1476. https://doi.org/10.1007/s10964-011-9688-9

González Salinas, M.D.C., Carranza Carnicero, J.A. and Hidalgo Montesinos, M.D. (1999). Adaptation to the Spanish population of the "Toddler Behavior Assessment Questionnaire" to measure temperament in childhood. *Behavioral Sciences Methodology, 1*, 207-218.

González Salinas, M.D.C., Hidalgo Montesinos, M.D., Carranza Carnicero, J.A. and Ato Garcia, M. (2000). Preparation of a Spanish adaptation of the Infant Behavior Questionnaire for the measurement of temperament in childhood. *Psicothema, 12*, 513-519.

Goodwin, R. D., Weinberger, A. H., Kim, J. H., Wu, M., & Galea, S. (2020). Trends in anxiety among adults in the United States, 2008–2018: Rapid increases among young adults. *Journal of Psychiatric Research*, *130*, 441–446. https://doi.org/10.1016/j.jpsychires.2020.08.014

Gvion, Y., & Apter, A. (2011). Aggression, Impulsivity, and Suicide Behavior: A Review of the Literature. *Archives of Suicide Research*, *15*(2), 93–112. https://doi.org/10.1080/13811118.2011.565265

Hernández, M. G., Nguyen, J., Casanova, S., Suárez-Orozco, C., & Saetermoe, C. L. (2013). Doing no harm and getting it right: Guidelines for ethical research with immigrant communities. *New Directions for Child and Adolescent Development*, *2013*(141), 43–60. https://doi.org/10.1002/cad.20042

Hill, P. L., & Edmonds, G. W. (2017). Personality development in adolescence. In *Personality Development Across the Lifespan* (pp. 25–38). Elsevier. https://doi.org/10.1016/B978-0-12-804674-6.00003-X

Hoffmann, M., Pérez, J. C., García, C., Rojas, G., & Martínez, V. (2017). Chilean adaptation and validation of the Early Adolescent Temperament Questionnaire-Revised version. *Frontiers in Psychology*, *8*, 2131. https://doi.org/10.3389/fpsyg.2017.02131

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*(3), 332–346. https://doi.org/10.1177/1088868310361240

Hopwood, C. J., Wright, A. G. C., & Bleidorn, W. (2022). Person–environment transactions differentiate personality and psychopathology. *Nature Reviews Psychology*, *1*(1), 55–63. https://doi.org/10.1038/s44159-021-00004-0

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

John, O. P., Naumann, L. P. & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research, 3*, 114-158.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2019). semTools: Useful tools for structural equation modeling. R package version 0.5-2. Retrieved from https://CRAN.R-project.org/package=semTools

Kann, L., Kinchen, S., Shanklin, S. L., Flint, K. H., Hawkins, J., Harris, W. A., et al. (2014). Youth Risk Behavior Surveillance - United States, 2015. MMWR Surveillance Summary

63(SS04):1–168. Retrieved from: https://www.cdc.gov/mmwr/preview/mmwrhtml/ss6304a1.htm

Karam, E. G., Itani, L., Fayyad, J., Hantouche, E., Karam, A., Mneimneh, Z., Akiskal, H., & Rihmer, Z. (2015). Temperament and suicide: A national study. *Journal of Affective Disorders*, *184*, 123–128. https://doi.org/10.1016/j.jad.2015.05.047

Kasen, S., Cohen, P., & Chen, H. (2011). Developmental course of impulsivity and capability from age 10 to age 25 as related to trajectory of suicide attempt in a community cohort: Suicide attempt, impulsivity, and capability. *Suicide and Life-Threatening Behavior*, *41*(2), 180–192. https://doi.org/10.1111/j.1943-278X.2011.00017.x

Kawakami, N., Abdulghani, E. A., Alonso, J., Bromet, E. J., Bruffaerts, R., Caldas-de-Almeida, J. M., Chiu, W. T., de Girolamo, G., de Graaf, R., Fayyad, J., Ferry, F., Florescu, S., Gureje, O., Hu, C., Lakoma, M. D., LeBlanc, W., Lee, S., Levinson, D., Malhotra, S., … Kessler, R. C. (2012). Early-life mental disorders and adult household income in the world mental health surveys. *Biological Psychiatry*, *72*(3), 228–237. https://doi.org/10.1016/j.biopsych.2012.03.009

Kerby, D. S. (2003). CART analysis with unit-weighted regression to predict suicidal ideation from Big Five traits. *Personality and Individual Differences*, *35*(2), 249–261. https://doi.org/10.1016/S0191-8869(02)00174-5

Kim, S., Brody, G. H., & Murry, V. M. (2003). Factor structure of the Early Adolescent Temperament Questionnaire and measurement invariance across gender. *The Journal of Early Adolescence*, *23*(3), 268–294. https://doi.org/10.1177/0272431603254178

Klein, D. N., Kotov, R., & Bufferd, S. J. (2011). Personality and depression: Explanatory models and review of the evidence. *Annual Review of Clinical Psychology*, *7*(1), 269–295. https://doi.org/10.1146/annurev-clinpsy-032210-104540

Klimstra, T. A., Hale, W. W., Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. H. J. (2009). Maturation of personality in adolescence. *Journal of Personality and Social Psychology*, *96*(4), 898–912. https://doi.org/10.1037/a0014746

Koode, D. A., & Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica, 54*(5), 1243-1248.

Kotelnikova, Y., Olino, T. M., Klein, D. N., Kryski, K. R., & Hayden, E. P. (2016). Higher- and lower-order factor analyses of the Children's Behavior Questionnaire in early and middle childhood. *Psychological Assessment*, *28*(1), 92–108. https://doi.org/10.1037/pas0000153

Kotelnikova, Y., Olino, T. M., Klein, D. N., Mackrell, S. V. M., & Hayden, E. P. (2017). Higher and lower order factor analyses of the Temperament in Middle Childhood Questionnaire. *Assessment*, *24*(8), 1050–1061. https://doi.org/10.1177/1073191116639376

Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin, 21,* 995–998.

Kushner, S. C., Tackett, J. L., & Bagby, R. M. (2012). The structure of internalizing disorders in middle childhood and evidence for personality correlates. *Journal of Psychopathology and Behavioral Assessment*, *34*(1), 22–34. https://doi.org/10.1007/s10862-011-9263-4

Laceulle, O. M., Nederhof, E., Karreman, A., Ormel, J., & Aken, M. A. G. (2012). Stressful events and temperament change during early and middle adolescence: The TRAILS study. *European Journal of Personality*, *26*(3), 276–284. https://doi.org/10.1002/per.832

Laceulle, O. M., Ormel, J., Vollebergh, W. A. M., van Aken, M. A. G., & Nederhof, E. (2014). A test of the vulnerability model: Temperament and temperament change as predictors of future mental disorders - the TRAILS study. *Journal of Child Psychology and Psychiatry*, *55*(3), 227–236. https://doi.org/10.1111/jcpp.12141

Latham, M. D., Dudgeon, P., Yap, M. B. H., Simmons, J. G., Byrne, M. L., Schwartz, O. S., Ivie, E., Whittle, S., & Allen, N. B. (2020). Factor structure of the Early Adolescent Temperament Questionnaire–Revised. *Assessment*, *27*(7), 1547–1561. https://doi.org/10.1177/1073191119831789

Lawson, K. M. & Robins, R. W. (2021). Sibling constructs: What are they, why do they matter, and how should we handle them? *Personality and Social Psychology Review*. https://doi.org/10.1177/10888683211047101

Lawson, K. M., Atherton, O. E., & Robins, R. W. (2021). The structure of adolescent temperament and associations with psychological functioning: A replication and extension of Snyder et al. (2015). *Journal of Personality and Social Psychology.*

Lengua, L. J. (2006). Growth in temperament and parenting as predictors of adjustment during children's transition to adolescence. *Developmental Psychology*, *42*(5), 819–832. https://doi.org/10.1037/0012-1649.42.5.819

Levey, D. F., Polimanti, R., Cheng, Z., Zhou, H., Nuñez, Y. Z., Jain, S., He, F., Sun, X., Ursano, R. J., Kessler, R. C., Smoller, J. W., Stein, M. B., Kranzler, H. R., & Gelernter, J. (2019). Genetic associations with suicide attempt severity and genetic overlap with major depression. *Translational Psychiatry*, *9*(1), 22.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A*

Multidisciplinary Journal, *9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694.

Lolas, F., Gomez, A., & Suarez, L. (1991). EPQ-R and suicide attempt: The relevance of psychoticism. *Personality and Individual Differences*, *12*(9), 899–902. https://doi.org/10.1016/0191-8869(91)90178-E

Lonigan, C. J., Phillips, B. M., & Hooe, E. S. (2003). Relations of positive and negative affectivity to anxiety and depression in children: Evidence from a latent variable longitudinal study. *Journal of Consulting and Clinical Psychology*, *71*(3), 465–481. https://doi.org/10.1037/0022-006X.71.3.465

Loukas, A., & Roalson, L. A. (2006). Family environment, effortful control, and adjustment among European American and Latino early adolescents. *Journal of Early Adolescence, 26*(4), 432-455.

Maiya, S., Carlo, G., Davis, A. N., & Streit, C. (2021). Relations among acculturative stress, internalizing symptoms, and prosocial behaviors in Latinx college students. *Journal of Latinx Psychology*, *9*(2), 77–91. https://doi.org/10.1037/lat0000177

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child & Adolescent Psychiatry, 36,* 554–565. http://dx.doi.org/ 10.1097/00004583-199704000-00019

Marks, A. K., Ejesi, K., & García Coll, C. (2014). Understanding the U.S. immigrant paradox in childhood and adolescence. *Child Development Perspectives*, *8*(2), 59–64. https://doi.org/10.1111/cdep.12071

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

Masyn, K. E. (2003). Discrete-time survival mixture analysis for single and recurrent events using latent variables. [Unpublished doctoral dissertation]. University of California, Los Angeles.

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, *45*(8), 1027–1035. https://doi.org/10.1016/j.jpsychires.2011.03.006

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2006). The utility of alternative fit indices in tests of measurement invariance. *Academy of Management Proceedings*, *2006*(1), B1–B6. https://doi.org/10.5465/ambpp.2006.27182124

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*(7), 1539–1550. https://doi.org/10.1016/j.paid.2008.01.010

Mezulis, A., Simonson, J., McCauley, E., & Vander Stoep, A. (2011). The association between temperament and depressive symptoms in adolescence: Brooding and reflection as potential mediators. *Cognition & Emotion*, *25*(8), 1460–1470. https://doi.org/10.1080/02699931.2010.543642

Millner, A. J., Lee, M. D., & Nock, M. K. (2015). Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. *PLOS ONE*, *10*(10), e0141606. https://doi.org/10.1371/journal.pone.0141606

Minkoff, K., Bergman, E., Beck, A. T., & Beck, R. (1973). Hopelessness, depression, and attempted suicide. *American Journal of Psychiatry, 130*(4), 455-459.

Miranda, R., Ortin, A., Scott, M., & Shaffer, D. (2014). Characteristics of suicidal ideation that predict the transition to future suicide attempts in adolescents. *Journal of child psychology and psychiatry, and allied disciplines, 55*(11), 1288–1296.

Mojtabai, R., Olfson, M., & Han, B. (2016). National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*, *138*(6), e20161878–e20161878. https://doi.org/10.1542/peds.2016-1878

Morgan, G., Hodge, K., Wells, K., & Watkins, M. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A Comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, *3*(1), 2–20. https://doi.org/10.3390/jintelligence3010002

Muris, P., & Meesters, C. (2009). Reactive and regulative temperament in youths: Psychometric evaluation of the Early Adolescent Temperament Questionnaire-Revised. *Journal of Psychopathology and Behavioral Assessment*, *31*(1), 7–19. https://doi.org/10.1007/s10862-008-9089-x

Muris, P., & Ollendick, T. H. (2005). The role of temperament in the etiology of child psychopathology. *Clinical Child and Family Psychology Review*, *8*(4), 271–289. https://doi.org/10.1007/s10567-005-8809-y

Muris, P., Meesters, C., & Blijlevens, P. (2007). Self-reported reactive and regulative temperament in early adolescence: Relations to internalizing and externalizing problem behavior and "Big Three" personality factors. *Journal of Adolescence*, *30*(6), 1035–1049. https://doi.org/10.1016/j.adolescence.2007.03.003

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407-422.

Muthén, B., & Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics*, *30*(1), 27–58. https://doi.org/10.3102/10769986030001027

Muthén, L. K., & Muthén, B. O. (1998–2017). Mplus user's guide (8th ed.). Los Angeles, CA: Muthén & Muthén.

Naicker, K., Galambos, N. L., Zeng, Y., Senthilselvan, A., & Colman, I. (2013). Social, demographic, and health outcomes in the 10 years following adolescent depression. *Journal of Adolescent Health*, *52*(5), 533–538. https://doi.org/10.1016/j.jadohealth.2012.12.016

Nigg, J. T. (2006). Temperament and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *47*(3–4), 395–422. https://doi.org/10.1111/j.1469-7610.2006.01612.x

Nock, M.K., Borges, G., Bromet, E.J., Cha, C.B., Kessler, R.C., & Lee, S. (2008). Suicide and suicidal behavior. *Epidemiologic Reviews, 30*, 133–154.

Nock, M.K., Green, J.G., Hwang, I., McLaughlin, K.A., Sampson, N.A., Zaslavsky, A.M., & Kessler, R.C. (2013). Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: Results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry, 70*, 300–310.

Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science*, *10*(5), 173–176. https://doi.org/10.1111/1467-8721.00142

Nordström, P., Schalling, D., & Asberg, M. (1995). Temperamental vulnerability in attempted suicide. *Acta Psychiatrica Scandinavica*, *92*(2), 155–160.

Ojanen, T., Findley, D., & Fuller, S. (2012). Physical and relational aggression in early adolescence: Associations with narcissism, temperament, and social goals. *Aggressive Behavior, 38*, 99-107.

Oldehinkel, A. J., Hartman, C. A., De Winter, A. F., Veenstra, R., & Ormel, J. (2004). Temperament profiles associated with internalizing and externalizing problems in preadolescence. *Development and Psychopathology*, *16*(02). https://doi.org/10.1017/S0954579404044591

Ormel, J., Oldehinkel, A. J., Ferdinand, R. F., Hartman, C. A., De Winter, A. F., Veenstra, R., Vollebergh, W., Minderaa, R. B., Buitelaar, J. K., & Verhulst, F. C. (2005). Internalizing and externalizing problems in adolescence: General and dimension-specific effects of familial loadings and preadolescent temperament traits. *Psychological Medicine*, *35*(12), 1825–1835. https://doi.org/10.1017/S0033291705005829

Peña, J. B., Wyman, P. A., Brown, C. H., Matthieu, M. M., Olivares, T. E., Hartel, D., Zayas, L. H. (2008). Immigration generation status and its association with suicide attempts, substance use, and depressive symptoms among Latino adolescents in the USA. *Prevention Science, 9*, 299-310.

Pérez-Edgar, K., & Fox, N. A. (2005). Temperament and anxiety disorders. *Child and Adolescent Psychiatric Clinics of North America, 14*, 681-706.

Pickles, A., Aglan, A., Collishaw, S., Messer, J., Rutter, M., & Maughan, B. (2010). Predictors of suicidality across the life span: The Isle of Wight study. *Psychological Medicine*, *40*(9), 1453–1466. https://doi.org/10.1017/S0033291709991905

Piquet, M. L., & Wagner, B. M. (2003). Coping Responses of adolescent suicide attempters and their relation to suicidal ideation across a 2-year follow-up: A preliminary study. *Suicide and Life-Threatening Behavior*, *33*(3), 288–301. https://doi.org/10.1521/suli.33.3.288.23212

Pompili, M., Girardi, P., Ruberto, A., & Tatarelli, R. (2005). Suicide in borderline personality disorder: A meta-analysis. *Nordic Journal of Psychiatry*, *59*(5), 319–324. https://doi.org/10.1080/08039480500320025

Price, J. H., & Khubchandani, J. (2017). Latina adolescents' health risk behaviors and suicidal ideation and suicide attempts: Results from the National Youth Risk Behavior Survey 2001–2013. *Journal of Immigrant and Minority Health, 19*(3), 533–542.

Prinstein, M. J., Boergers, J., & Vernberg, E. M. (2001). Overt and relational aggression in adolescents: Social-psychological adjustment of aggressors and victims. *Journal of Clinical Child & Adolescent Psychology, 30*(4), 479–491. http://dx.doi.org/10.1207/S15374424JCCP3004_05

Przeworski, A., Newman, M. G., Pincus, A. L., Kasoff, M. B., Yamasaki, A. S., Castonguay, L. G., & Berlin, K. S. (2011). Interpersonal pathoplasticity in individuals with generalized anxiety disorder. *Journal of Abnormal Psychology*, *120*(2), 286–298. https://doi.org/10.1037/a0023334

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

Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasz & A. Angleitner *Advances in Research on Temperament* (p. 163-180).

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Robins, R. W., Caspi, A., & Moffitt, T. (2002). It's not just who you're with, it's who you are: Personality and relationship experiences across multiple relationships. *Journal of Personality*, *70*, 925-964.

Robins, R. W., Donnellan, M. B., Widaman, K. F., & Conger, R. D. (2010). Evaluating the link between self-esteem and temperament in Mexican origin early adolescents. *Journal of Adolescence*, *33*(3), 403–410. https://doi.org/10.1016/j.adolescence.2009.07.009

Rodriguez, A., Reise, S.P., & Haviland, M.G. (2015). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods.* http://dx.doi.org/10.1037/met0000045

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. http://www.jstatsoft.org/v48/i02/.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16,* 207–212. http://dx.doi .org/10.1111/j.1467-8721.2007.00505.x

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development.* New York, NY: Guilford Press.

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology, 103*(1), 55-66. https://doi.org/10.1037/0021-843X.103.1.55

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 99-166). Hoboken, NJ, US: John Wiley & Sons Inc.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology*, *78*(1), 122–135. https://doi.org/10.1037/0022-3514.78.1.122

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (2020). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21-39.

Rueter, M. A., Holm, K. E., McGeorge, C. R., & Conger, R. D. (2008). Adolescent suicidal ideation subgroups and their association with suicidal plans and attempts in young adulthood. *Suicide and Life-Threatening Behavior, 38*, 564–575.

Salzer, S., Pincus, A. L., Hoyer, J., Kreische, R., Leichsenring, F., & Leibing, E. (2008). Interpersonal subtypes within Generalized Anxiety Disorder. *Journal of Personality Assessment*, *90*(3), 292–299. https://doi.org/10.1080/00223890701885076

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037//1082-989X.7.2.147

Schwab-Stone, M., Shaffer, D., Dulcan, M., Jensen, P., Fisher, P., Bird, H., …Rubio-Stipec, & Rae, D. (1996). Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC 2.3). *Journal of the American Academy of Child & Adolescent Psychiatry, 35*, 878-888. http://dx.doi.org/10.1097/00004583-199607000-00013

Schwaba, T., Bleidorn, W., Hopwood, C.J., Manuck, S.B., & Wright, A.G.C. (2022). Refining the maturity principle of personality development: Facets, close others, and co-maturation. *Journal of Personality and Social Psychology.* https://doi.org/10.1037/pspp0000400

Shain, B. and AAP Committee on Adolescence. (2016). Suicide and suicide attempts in adolescents. *Pediatrics,138*, e20161420.

Sharaf, T., & Tsokos, C. P. (2014). Predicting survival time of localized melanoma patients using discrete survival time method. *Journal of Modern Applied Statistical Methods*, *13*(1), 140–156. https://doi.org/10.22237/jmasm/1398917280

Shiner, R. L. (2015). The development of temperament and personality traits in childhood and adolescence. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences.* (pp. 85–105). American Psychological Association. https://doi.org/10.1037/14343-004

Shiner, R. L., & DeYoung, C. G. (2013). The structure of temperament and personality traits: A developmental perspective. In P. D. Zelazo (Eds.), *The Oxford Handbook of Developmental Psychology* (pp. 113-141). Oxford University Press.

Shiner, R. L., Soto, C. J., & De Fruyt, F. (2021). Personality assessment of children and adolescents. *Annual Review of Developmental Psychology*, *3*(1), 113–137. https://doi.org/10.1146/annurev-devpsych-050620-114343

Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences: Personality development. *Journal of Child Psychology and Psychiatry*, *44*(1), 2–32. https://doi.org/10.1111/1469-7610.00101

Silva, C., & Van Orden, K. A. (2018). Suicide among Hispanics in the United States. *Current Opinion in Psychology*, 22, 44-49.

Simonds, J. (2006). *The role of reward sensitivity and response: Execution in childhood extraversion.* Unpublished doctoral dissertation, University of Oregon. https://research.bowdoin.edu/rothbart-temperament-questionnaires/files/2016/09/Simonds_Dissertation_2006.pdf

Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press.

Snyder, H. R., Gulley, L. D., Bijttebier, P., Hartman, C. A., Oldehinkel, A. J., Mezulis, A., Young, J. F., & Hankin, B. L. (2015). Adolescent emotionality and effortful control: Core latent constructs and links to psychopathology and functioning. *Journal of Personality and Social Psychology*, *109*(6), 1132–1149. https://doi.org/10.1037/pspp0000047

Sorenson, S. B., & Shen, H. (1996). Youth suicide trends in California: An examination of immigrant and ethnic group risk. *Suicide and Life-Threatening Behavior, 26*(2).

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science*, *24*(5), 358–362. https://doi.org/10.1177/0963721415589345

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*(2), 330–348. https://doi.org/10.1037/a0021717

Spirito, A., Esposito-Smythers, C. (2006). Attempted and completed suicide in adolescence. *Annual Review of Clinical Psychology, 2*, 237–266.

Stein, G.L., Castro-Schilo, L., Cavanaugh, A., Mejia, Y.M., Christophe, K., & Robins, R. W. (2019). When discrimination hurts: The longitudinal impact of increases in peer discrimination on anxiety and depressive symptoms in Mexican-origin youth. *Journal of Youth and Adolescence*, *48*, 864-875.

Suarez-Morales, L., & Lopez, B. (2009). The impact of acculturative stress and daily hassles on pre-adolescent psychological adjustment: Examining anxiety symptoms. *The Journal of Primary Prevention*, *30*(3–4), 335–349. https://doi.org/10.1007/s10935-009-0175-y

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., ... Wu, M. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. Journal of the American Academic of Child & Adolescent Psychiatry, 40, 168-179. http://dx.doi.org/10.1097/00004583-200102000-00011

Tackett, J. L. (2006). Evaluating models of the personality–psychopathology relationship in children and adolescents. *Clinical Psychology Review*, *26*(5), 584–599. https://doi.org/10.1016/j.cpr.2006.04.003

Tackett, J. L., & Durbin, C. E. (2017). Advances in research on youth personality: Introduction to the special issue. *Journal of Research in Personality, 67*, 1-2.

Tackett, J. L., Kushner, S. C., De Fruyt, F., & Mervielde, I. (2013). Delineating personality traits in childhood and adolescence: Associations across measures, temperament, and behavioral problems. *Assessment*, *20*(6), 738–751. https://doi.org/10.1177/1073191113509686

Tackett, J. L., Kushner, S. C., Herzhoff, K., Smack, A. J., & Reardon, K. W. (2014). Viewing relational aggression through multiple lenses: Temperament, personality, and personality pathology. *Development and Psychopathology*, *26*(3), 863–877. https://doi.org/10.1017/S0954579414000443

Tanabe, S., Terao, T., Shiotsuki, I., Kanehisa, M., Ishii, K., Shigemitsu, O., Fujiki, M., Hoaki, N. (2016). Anxious temperament as a risk factor of suicide attempt. *Comprehensive Psychiatry*, 68, 72– 77.

Tanji, F., Kakizaki, M., Sugawara, Y., Watanabe, I., Nakaya, N., Minami, Y., Fukao, A., & Tsuji, I. (2015). Personality and suicide risk: The impact of economic crisis in Japan. *Psychological Medicine*, *45*(3), 559–573. https://doi.org/10.1017/S0033291714001688

Taylor, Z. E., Widaman, K. F., & Robins, R. W. (2018). Longitudinal relations of economic hardship and effortful control to active coping in Latino youth. *Journal of Research on Adolescence*, *28*, 396–411. https://doi.org/10.1111/jora.12338

Thompson, M.P., & Swartout, K.M. (2018). Epidemiology of suicide attempts among youth transitioning to adulthood. *Journal of Youth and Adolescence, 47*, 807-817. *Threatening Behavior, 31*, 122-131.

Torres, L. (2010). Predicting levels of Latino depression: Acculturation, acculturative stress, and coping. *Cultural Diversity and Ethnic Minority Psychology*, *16*(2), 256–263. https://doi.org/10.1037/a0017357

Tucker, R. P., O'Keefe, V. M., Cole, A. B., Rhoades-Kerswill, S., Hollingsworth, D. W., Helle, A. C., DeShong, H. L., Mullins-Sweatt, S. N., & Wingate, L. R. (2014). Mindfulness tempers the impact of personality on suicidal ideation. *Personality and Individual Differences*, *68*, 229–233. https://doi.org/10.1016/j.paid.2014.05.001

Valiente, C., Eisenberg, N., Spinrad, T. L., Haugen, R., Thompson, M. S., & Kupfer, A. (2013). Effortful control and impulsivity as concurrent and longitudinal predictors of academic achievement. *The Journal of Early Adolescence*, *33*(7), 946–972. https://doi.org/10.1177/0272431613477239

Van Beveren, M.-L., McIntosh, K., Vandevivere, E., Wante, L., Vandeweghe, L., Van Durme, K., Vandewalle, J., Verbeken, S., & Braet, C. (2016). Associations between temperament, emotion regulation, and depression in youth: The role of positive temperament. *Journal of Child and Family Studies*, *25*(6), 1954–1968. https://doi.org/10.1007/s10826-016-0368-y

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01064

van den Akker, A. L., Deković, M., & Prinzie, P. (2010). Transitioning to adolescence: How changes in child personality and overreactive parenting predict adolescent adjustment problems. *Development and Psychopathology*, *22*(1), 151–163. https://doi.org/10.1017/S0954579409990320

van den Akker, A. L., Deković, M., Asscher, J., & Prinzie, P. (2014). Mean-level personality development across childhood and adolescence: A temporary defiance of the maturity principle and bidirectional associations with parenting. *Journal of Personality and Social Psychology*, *107*(4), 736–750. https://doi.org/10.1037/a0037248

Velting, D. M. (1999). Suicidal ideation and the five-factor model of personality. *Personality and Individual Differences, 27*, 943-952.

Verona, E., Patrick, C. J., & Joiner, T. E. (2001). Psychopathy, antisocial personality, and suicide risk. *Journal of Abnormal Psychology*, *110*(3), 462–470.

Verstraeten, K., Vasey, M. W., Raes, F., & Bijttebier, P. (2009). Temperament and risk for depressive symptoms in adolescence: Mediation by rumination and moderation by effortful control. *Journal of Abnormal Child Psychology*, *37*(3), 349–361. https://doi.org/10.1007/s10802-008-9293-x

Viñas, F., González, M., Gras, E., Jane, C., & Casas, F. (2015). Psychometric properties of the EATQ-R among a sample of Catalan-speaking Spanish adolescents. *Universitas Psychologica, 14*, 747-758.

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97,* 346-353.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negativeaffect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070.

Watson, D., Clark, L. A., Weber, K., & Assenheimer, J. S., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of Abnormal Psychology, 104*(1), 15-25.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology, 104(1)*, 4-14.

Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 1-19.

Werner, E. E., & Smith, R. S. (1992). *Overcoming the odds: High risk children from birth to adulthood.* Cornell University Press.

Wetter, E. K., & Hankin, B. L. (2009). Mediational pathways through which positive and negative emotionality contribute to anhedonic symptoms of depression: A prospective study of adolescents. *Journal of Abnormal Child Psychology*, *37*(4), 507–520. https://doi.org/10.1007/s10802-009-9299-z

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. New York.

Wilson, S., & Olino, T. M. (2021). A developmental perspective on personality and psychopathology across the life span. *Journal of Personality*, *89*(5), 915–932. https://doi.org/10.1111/jopy.12623

World Health Organization (WHO) (2014). *Preventing suicide: A global imperative*. Geneva, Switzerland: World Health Organization.

Xie, H., McHugo, G., Drake, R., & Sengupta, A. (2002). Using discrete-time survival analysis to examine patterns of remission from substance use disorder among persons with severe mental illness. *Mental Health Services Research*, 10.

Yen, S., Shea, T., Pagano, M., Sanislow, C. A., Grilo, C. M., McGlashan, T. H., Skodol, A. E., Bender, D. S., Zanarini, M. C., Gunderson, J. G., & Morey, L. C. (2003). Axis I and Axis II disorders as predictors of prospective suicide attempts: Findings from the Collaborative Longitudinal Personality Disorders Study. *Journal of Abnormal Psychology*, *112*(3), 375–381. https://doi.org/10.1037/0021-843X.112.3.375

Young, G. R., Karnilowicz, H. R., Mauss, I. B., Hastings, P. D., Guyer, A. E., & Robins, R. W. (2022). Prospective associations between emotion regulation and depressive symptoms among Mexican-origin adolescents. *Emotion, 22*(1), 129–141. https://doi.org/10.1037/emo0001060

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*(02), 601–617.
https://doi.org/10.1017/S0954579418000159

Endnotes

[1] The article describing the initial development of the EATQ (Capaldi & Rothbart, 1992) has been cited almost 600 times, and the publication describing the development of the revised EATQ-R (Ellis & Rothbart, 2001) has been cited over 700 times.

[2] More information about the EATQ-R is available at the following website: https://research.bowdoin.edu/rothbart-temperament-questionnaires/instrument-descriptions/the-early-adolescent-temperament-questionnaire/

[3] Rothbart and colleagues developed a suite of age-appropriate temperament measures, including measures designed for infancy, early childhood, childhood, middle childhood, adolescence (i.e., the EATQ-R), and adulthood.

[4] Notably, many of these concerns about bifactor models were not widely discussed until after Snyder et al. conducted their research.

[5] The present study does not have data from two PE subscales used in Snyder et al. (2015). Therefore, we fit a model where Surgency and Affiliation form two separate factors, rather than one general factor. This is consistent with findings from Snyder et al. (2015), which showed that Surgency and Affiliation share little common variance, and also follows the approach recommended by the first author of the original article (Dr. Hannah Snyder, personal communication, May 10, 2019).

[6] Nine published papers have used data from the California Families Project (CFP) to examine temperament measured via the EATQ-R (Atherton, Lawson, Ferrer, & Robins, 2020; Atherton, Lawson, & Robins, 2020; Atherton, Schofield, Sitka, Conger, & Robins, 2016; Atherton, Tackett, Ferrer, & Robins, 2017; Atherton, Zheng, Bleidorn, Robins, in press; Clark, Donnellan, Conger, & Robins, 2015; Clark, Donnellan, & Robins, in press; Robins, Donnellan, Widaman, & Conger, 2010; Taylor, Widaman, Robins, 2018). Some of these studies have computed correlations between EATQ-R domains and the adolescent functioning measures from the present study. Specifically, Atherton et al. (2017) examined how Effortful Control, Negative Emotionality, and Positive Emotionality were associated with relational aggression and used data overlapping with the present study at two out of four waves. Additionally, Atherton, Lawson, Ferrer, & Robins (2020) examined how Effortful Control was associated with ADHD symptoms with data overlapping with the present study at two out of four waves. Importantly, these papers differ significantly from the present study in terms of research questions, scoring of the EATQ-R, and type of analyses. Moreover, no previous CFP publications have examined the association of any EATQ-R temperament domain with any of the other measures of adolescent functioning included in Snyder et al. (2015). For a full list of California Families Project publications, see: https://osf.io/ky7cw/.

[7] Participants completed the EATQ-R in either English (90%) or Spanish (10%). Researchers have translated a number of Rothbart's temperament measures into Spanish (e.g., González Salinas et al., 1999; González Salinas et al., 2000), including the EATQ-R (see

https://research.bowdoin.edu/rothbart-temperament-questionnaires/instrument-descriptions/the-early-adolescent-temperament-questionnaire/). There is work examining the validity of the EATQ-R for the Chilean Spanish version (Hoffmann et al., 2017) and the Spanish version in a sample of Catalan-speaking Spanish adolescents (Viñas et al., 2015).

[8] Our project attempted to replicate and extend the analyses from Snyder et al. (2015). Given our primary goal of replication, we did not preregister our initial analyses, and instead used the exact same methods from Snyder et al. (2015) to replicate their work. During the publication process, the reviewers and editor recommended that we conduct additional analyses. These analyses are preregistered on the project OSF page, which also includes R scripts to run all analyses and materials: https://osf.io/5rhjb/. We made one deviation from our supplemental preregistration. Specifically, when we conducted new model comparisons, we considered the interpretability of the factor loadings (e.g., whether all items loaded in the expected direction), in addition to model fit indices, to adjudicate between models. Additionally, we are not legally or ethically allowed to publicly post data for this project because the participants in the study have not given informed consent to have their personal data publicly shared, and we do not have IRB approval to post data publicly. Researchers interested in replicating findings can contact the corresponding author to gain access to individual-level data. Further, we are unable to share full materials for the NIMH DISC-IV depression, anxiety, and ADHD scales as they are copyrighted psychiatric assessments.

[9] For NE, we ran analyses both with and without the Depressed Mood and Aggression subscales. To be consistent with Snyder et al. (2015, p. 1138, Footnote 6), we report results from the analyses including Depressed Mood and Aggression in the main text, and we report results excluding Depressed Mood and Aggression in the supplemental materials (see Tables S12 – S13).

[10] We also ran additional exploratory analyses examining whether model fit was improved when we used an estimator that is appropriate for *ordinal* data, the diagonally weighted least squares with mean and variance adjustments (WLSMV) estimator. In contrast to WLSMV, ML estimation relies on *interval*-level indicators, but the Likert scales used to rate the EATQ-R items arguably fail to attain interval-level measurement, which could contribute to the suboptimal model fit we found for many models using ML. Consistent with this reasoning, we found substantially better model fit for EC and NE (but not PE) when we used WLSMV estimation instead of ML estimation (Table S1.4).

[11] Items that had correlated residual variances for the original and modified models can be found in Table S1.2. For the EC scale, none of the pairs of item residual covariances overlapped between Snyder et al. (2015) and the present study. For the NE scale, one pair of item correlated residuals overlapped (i.e., residual variances between item 60 and 64). For the PE scale, the same single pair of items had correlated residuals in both Snyder et al. (2015) and the present study, resulting in identical model fit for both the original model and the modified correlated residuals model.

[12] For EC, the hierarchical model did not converge. Additionally, in the correlated factors model, the correlation between the Attention and Inhibitory Control subscales had to be constrained to 1.

[13] We also ran these model comparisons using WLSMV estimation and found a similar pattern of findings, but with substantially better model fit for all models (Table S5). In other words, our conclusions about which model fit best were the same regardless of whether we used ML or WLSMV estimation.

[14] Indeed, Rothbart did not include Aggression as part of the EATQ-R NE domain (Ellis & Rothbart, 2001).

[15] Ten published papers have used data from the California Families Project (CFP) to examine temperament measured via the EATQ-R (Atherton, Lawson, Ferrer, & Robins, 2020; Atherton, Lawson, & Robins, 2020; Atherton, Schofield, Sitka, Conger, & Robins, 2016; Atherton, Tackett, Ferrer, & Robins, 2017; Atherton, Zheng, Bleidorn, Robins, 2019; Clark, Donnellan, Conger, & Robins, 2015; Clark, Donnellan, & Robins, 2018; Lawson et al., in press; Robins, Donnellan, Widaman, & Conger, 2010; Taylor, Widaman, Robins, 2018). No previous CFP publications have examined suicidal ideation, plans, or attempts. For a complete list of California Families Project publications, see: https://osf.io/ky7cw/.

[16] All materials, scripts, and output files for this project are available on the Open Science Framework (OSF): https://osf.io/j7vfb/. We are not legally or ethically allowed to publicly post data for this project because the participants in the study have not given informed consent to have their personal data publicly shared, and we do not have IRB approval to post data publicly. Researchers interested in replicating findings can contact the corresponding author to gain access to individual-level data.

[17] Beginning at age 18, skip logic was used to assess suicidal ideation and behaviors. Specifically, participants were first asked the YRBS item about ideation and were only presented with the subsequent YRBS planning item if they endorsed ideation. Similarly, participants were only presented with the YRBS attempt item if they endorsed planning. Therefore, the data might *underestimate* planning and attempts at ages 18 through 21 because youth sometimes report planning or attempting suicide without endorsing ideation.

[18] Rothbart's EATQ-R measure also includes a Perceptual Sensitivity scale (assessing awareness of low-intensity stimulation in the environment) and a Pleasure Sensitivity scale (assessing pleasure related to activities or stimuli involving low intensity), which are conceptually related to Affiliation (Snyder et al., 2015). When the present study was launched in 2006, the Perceptual Sensitivity and Pleasure Sensitivity subscales were not considered relevant to the study aims and were therefore not included at any assessment.

[19] We believe that the data are missing at random (MAR) because the propensity for missing suicidal ideation, plans, and/or attempts data can be explained by other variables in the dataset – namely temperament (adolescents with missing temperament data at all waves were not included in the survival analyses).

[20] Given the slight dip in sample size at age 18 and the drop in ideations, plans, and attempts at this age, we examined whether the participants who expressed suicidal ideation, plans, and

attempts at age 17 were more likely to have missing data at age 18. We found that, of the 65 participants who expressed suicidal ideation at age 17, only 1 participant was missing suicide data at age 18. We found the same results (i.e., 1 participant missing suicide data at age 18) for the 27 participants who expressed making a suicide plan at age 17 and the 21 participants who endorsed making a suicide attempt at age 17. Thus, this drop does not appear to be due to selective attrition.

[21] To gauge robustness of the results, we also bootstrapped the confidence intervals of all significant results to see if the results remained largely unchanged. Using maximum likelihood estimation and bootstrapping 10,000 iterations, we found no meaningful differences in the point estimates or upper/lower limits of the confidence intervals (differences were all less than or equal to .03).

[22] For comparison, we also present odds ratios for the observed variables in Table S2.5. Results are the same between observed and latent variables with respect to significance, but the odds ratios are substantially larger for observed variables, which is due to the fact that the observed and latent variables are on different scales (see Tables 2.1 and S2.1 to compare).

[23] To examine the potential confounding impact of gender, we ran models where both gender and temperament domains/facets were included as predictors in the models. We found that 35 out of the 39 (90%) results remained the same. For the results that changed, two significant associations became nonsignificant and two nonsignificant associations became significant. In particular, the relation between suicide attempts and Activation Control became significant (OR = 0.60, $p < .001$), which also happened with suicide attempts and Inhibitory Control (OR = 0.67, $p = .001$). On the other hand, the association between suicide attempts and the broad NE domain became nonsignificant (OR = 1.52, $p = .009$), which also happened with suicide attempts and the Frustration facet (OR = 1.44, $p = .021$).

[24] Adolescent self-reported Big Five traits (i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness) were measured using the 44-item Big Five Inventory (BFI; John et al., 1991; John et al., 2008) at ages 14 and 16. Given that we do not have Big Five data at age 12, we began these survival analyses at age 14 (versus age 12, as was the case for the temperament analyses).

[25] In our preregistration, we stated that we would also examine pubertal timing as a moderator, but we decided to omit these analyses and focus on our core research questions. We did not run any analyses examining pubertal timing and temperament and anxiety/depression.

[26] These predictions were preregistered on the Open Science Framework (https://osf.io/qa27t/) before any analyses were conducted. However, we discovered a mistake in the preregistration after it was uploaded. In particular, we preregistered that we already knew the mean-level trajectories of the EC domain and facets based on Atherton et al. (2020), but Atherton et al. study examined change from age 10 to 19 whereas the present study examines change from age 10 to 16. Given that the best-fitting trajectory can differ based on the waves of data included, we edited the hypotheses to include the age 10 to 16 EC trajectory identified by Damian et al. (2020)

and to remove hypotheses about the EC facets, since Damian et al. did not examine facet-level trajectories.

[27] This project involves secondary data analysis of the California Families Project (CFP). Previous published work has used CFP data to examine the development of Effortful Control measured via the EATQ-R (Atherton, Lawson, & Robins, 2020; Atherton, Lawson, Ferrer, & Robins, 2020; Damian et al., 2020). Another published paper (Lawson et al., 2021) examined the relation between Effortful Control, Positive Emotionality, and Negative Emotionality at age 12 and depression symptoms measured via the *DISC-IV* at ages 12 and 14. For a complete list of CFP publications, see: https://osf.io/ky7cw/.

[28] The EATQ-R PEM domain also includes a Perceptual Sensitivity scale (assessing awareness of low-intensity stimulation in the environment) and a Pleasure Sensitivity scale (assessing pleasure related to activities or stimuli involving low intensity). However, these scales were not deemed relevant to the aims of the CFP study when the data were collected and, thus, were not included at any assessment.

[29] For four of the temperament facets (Inhibitory Control, Depressed Mood, Surgency, and Affiliation), model fit of the partial strong models was worse than the weak models (e.g., change in CFI > .01). However, absolute model fit assessed via CFI and RMSEA was still adequate for these partial strong models. Therefore, to facilitate latent growth curve models, we retained partial strong invariance for these four facets.

[30] Because the main analyses for Surgency included three waves of data (ages 12, 14, and 16), we only compared the no-growth and linear growth models for Surgency.

[31] An alpha level of .006 was calculated by dividing .05 by nine, which is calculated by multiplying the three unique temperament domains (i.e., EC, NEM, PEM) by the three unique components of anxiety/depression (general distress, anxious arousal, anhedonic depression).

[32] These gender moderation analyses deviate from our preregistration, where we said that we would run multiple group models and compare model fit of constrained vs. unconstrained models. Given that the analyses of temperament and anxiety/depression involved saved factor scores from multiple models, we decided that including the main effect and interaction of gender and temperament in multiple regression models was a more appropriate analysis.

[33] Although we did not find any significant differences in the EC means for the girls and boys, the covariance between the level and slope of Inhibitory control was negative for girls (*Covariance*= -.008, p = .031), but positive (and nonsignificant) for boys (*Covariance* = .001, p = .525), and the slope variance was significant for girls (*SlopeVariance* = .002, p = .040) but had to be fixed to 0 for boys. These differences contributed to the worse model fit for the constrained Inhibitory Control model.

[34] We also find linear increases in Surgency when we look at the trajectory from age 10 (where there are only 6 items) to age 16 (see Table S6 for details).