Adolescent emotionality and effortful control: Core latent constructs and links to psychopathology and functioning

Hannah R. Snyder [a*], Lauren D. Gulley [a], Patricia Bijttebier [b], Catharina A. Hartman [c,] Albertine J. Oldehinkel [c], Amy Mezulis [d], Jami F. Young [e], & Benjamin L. Hankin [a]

[a] Department of Psychology, University of Denver, Denver, CO, USA

[b] Department of Psychology, KU Leuven, Leuven, Belgium

[c] Department of Psychiatry, University of Groningen and University Medical Center Groningen, Groningen, Netherlands

[d] Department of Clinical Psychology, Seattle Pacific University, Seattle, WA, USA

[e] Graduate School of Applied and Professional Psychology, Rutgers University, New Brunswick, NJ, USA

[*] Correspondence regarding this article should be addressed to Hannah R. Snyder, Dept. of Psychology, University of Denver, Frontier Hall, 2155 S. Race St., Denver, CO 80208, USA. Email: hsnyders@du.edu.

Collection of data from Samples 1 and 2 and preparation of this manuscript were supported by grants from the National Institute of Mental Health: 5R01MH077195 and 5R01MH077178 (Hankin and Young) and F32MH098481 (Snyder). Collection of data from Sample 3 was supported by grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grants 60-60600-97-118; ZonMw Culture and Health

grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013), the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities (University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group), and Accare Center for Child and Adolescent Psychiatry. Collection of data from Samples 4 and 5 was supported by grants from the Fund of Scientific Research-Flanders: G.0357.08 and G.0923.12 (Bijttebier). Collection of data from Sample 6 was supported by a grant from the National Institute of Mental Health: R15MH098294-01A1 (Mezulis).

Abstract

Temperament is associated with important outcomes in adolescence, including academic and interpersonal functioning and psychopathology. Rothbart's temperament model is among the most well-studied and supported approaches to adolescent temperament, and contains three main components: positive emotionality (PE), negative emotionality (NE), and effortful control (EC). However, the latent factor structure of Rothbart's temperament measure for adolescents, the Early Adolescent Temperament Questionnaire Revised (EATQ-R, Ellis & Rothbart, 2001) has not been definitively established. To address this problem and investigate links between adolescent temperament and functioning, we used confirmatory factor analysis to examine the latent constructs of the EATQ-R in a large combined sample. For EC and NE, bifactor models consisting of a common factor plus specific factors for some sub-facets of each component fit best, providing a more nuanced understanding of these temperament dimensions. The nature of the PE construct in the EATQ-R is less clear. Models replicated in a hold-out dataset. The common components of high NE and low EC where broadly associated with increased psychopathology symptoms, and poor interpersonal and school functioning, while specific components of NE were further associated with corresponding specific components of psychopathology. Further questioning the construct validity of PE as measured by the EATQ-R, PE factors did not correlate with construct validity measures in a way consistent with theories of PE. Bringing consistency to the way the EATQ-R is modeled and using purer latent variables has the potential to advance the field in understanding links between dimensions of temperament and important outcomes of adolescent development.

Keywords: *temperament, adolescent, EATQ, factor analysis, psychopathology, interpersonal*

functioning, school functioning

Adolescent emotionality and effortful control: Core latent constructs and links to psychopathology and functioning

Temperament, broadly defined, refers to individual differences in behavioral response styles or dispositional traits that are present early in life. These individual differences are assumed to have a constitutional basis, meaning that they are a fairly stable part of the biological makeup of an organism but can be influenced over time by heredity, maturation, and experience (e.g., Rothbart, 2007). A preponderance of studies have shown that temperament is associated with a variety of outcomes in childhood and adolescence, including academic achievement (Valiente et al., 2013), interpersonal functioning (Eisenberg, Vaughan, & Hofer, 2009), cognitive processing (e.g., Lonigan, Vasey, Phillips, & Hazen, 2004), and emotion regulation (e.g., Yap et al., 2011). Furthermore, maladaptive forms of temperament are associated with psychopathology, including externalizing and internalizing problems (see Nigg, 2006 for review).

Over the past few decades, several theoretical frameworks have been used to conceptualize temperament (e.g., Buss & Plomin, 1975; Chess & Thomas, 1977; Rothbart, 1981). Of these original accounts, Rothbart's temperament model has become among the most well-studied and supported approaches to conceptualizing individual differences in adolescent temperament (e.g., Derryberry & Rothbart, 1997; Putnam, Ellis, & Rothbart, 2001). Rothbart defines temperament as individual differences in three main superordinate factors: positive emotionality (PE), negative emotionality (NE), and self-regulation (i.e., effortful control, EC). Two of these, positive emotionality PE and NE, involve affective reactivity, which refers to excitability, responsivity, or arousability of the behavioral and physiological systems of an

organism (Rothbart & Rueda, 2005). PE (e.g., smiling/laughter, activeness, assertiveness) directs approach behavior towards reward and overlaps with other well-established reward-related constructs, such as extraversion and Gray's Behavioral Activation System (BAS) (Derryberry & Rothbart, 1997; Muris & Ollendick, 2005). Individuals who are high on PE are receptive to reward, sociable, and actively engaged with their environment. NE (e.g., sadness, anger, frustration), on the other hand, mobilizes avoidance behavior away from non-reward or punishment and is closely related to constructs such as neuroticism and Gray's Behavioral Inhibition System (BIS) (Derryberry & Rothbart, 1997; Muris & Ollendick, 2005). Individuals who are high on NE demonstrate vigilance for negative cues and restricted engagement with the environment.

Effortful control (EC) represents the last domain of Rothbart's model, and involves the recruitment of attentional and behavioral processes to modulate affective reactivity (Rothbart & Rueda, 2005; Rothbart, Ellis, Rueda, & Posner, 2003). Broadly, these processes facilitate the ability to employ flexible, strategic, and ultimately, effective coping strategies to modulate high levels of emotional reactivity (e.g., Lengua & Long, 2002). Processes of EC include the ability to maintain or shift attentional focus, inhibit maladaptive behavioral responses, or activate appropriate responses in light of changing task demands (e.g., Eisenberger, Lieberman, & Satpute, 2005).

Associations between Temperament and Adolescent Functioning

Research on adolescent temperament has identified key associations with important domains of adolescent functioning. Broadly speaking, there is ample evidence showing associations between high levels of NE and low levels of EC, on the one hand, and maladaptive adolescent functioning and psychopathology, on the other hand. Findings also show associations

among particular components of PE and these outcomes. For example, low levels of EC are associated with poor academic performance (Valiente et al., 2013) and difficulty adapting both emotionally and behaviorally to the social demands of a classroom environment (Al-Hendawi, 2013). Low levels of EC and high levels of anger and frustration, two lower order constructs of NE, are also related to problematic peer interactions, including aggression (e.g., hitting) and the experience of peer victimization. This is likely due to difficulties regulating negative emotions and behaviors in the context of stressful interpersonal interactions (Coplan and Bullock, 2012; Eisenberg et al., 2009). Shyness, another lower order construct of NE, is associated with fewer and lower-quality friendships, as these youth are less likely to engage with peers and instead withdraw themselves from social interactions (e.g. Coplan & Bullock, 2012).

In addition to academic and interpersonal functioning, temperament represents an ideal construct for understanding adolescent psychopathology because it is related conceptually, as well as empirically, to hierarchical models of psychopathology, including externalizing and internalizing problems (see Griffith et al., 2010; Markon, Krueger, & Watson, 2005). Externalizing problems, such as ADHD and conduct problems, are typically characterized by low levels of EC, high sensitivity to reward, a component of PE, and high levels of anger and frustration (Muris, Meesters, Blijlevens, 2007). In terms of internalizing disorders, anxiety is associated with low levels of EC and high levels of fear, another lower order construct of NE (Nigg, 2006). Depression, an internalizing disorder often comorbid with anxiety, is linked to low levels of EC and high levels of NE, including sadness and fear, but is distinguished from anxiety by low levels of PE (Nigg, 2006). Evidence also shows associations between low levels of EC and high levels NE and other forms of psychopathology, including non-suicidal self-injury and substance abuse (e.g. Baetens, Claes, Willem, Muehlenkamp, & Bijtebier, 2011).

Research has recently focused on elucidating the mediating mechanisms linking adolescent temperament to psychopathology. Findings suggest that high levels of NE and low levels of EC are related to maladaptive cognitive processes and deficits in emotion regulation, and subsequently, adolescent psychopathology. Rumination, for instance, has been found to mediate the link between high levels of NE and depression, especially for individuals with low levels of EC (Verstaeten, Vasey, Raes, & Jijttebier, 2008). High levels of NE and low levels of EC are also related to attention bias to threatening emotional information, a well-established correlate of anxiety (Lonigan et al., 2004). High levels of NE are also linked to other types of maladaptive emotion regulation strategies, including both suppression of negative affect and dysregulated expression of negative affect (Yap et al., 2011). Taken together, research on temperament and adolescent functioning highlights the idea that understanding the construct of temperament has important implications for understanding adolescent functioning in academic and social contexts, as well as adolescent psychopathology. Next, we turn to the measurement of adolescent temperament.

Measurement of Temperament: EATQ-R

Recently the issue of replicability in science, and especially in psychology, has re-emerged and been hotly debated (e.g., Pashler & Wagenmakers, 2012). In the area of individual differences in traits, one key barrier to evaluating whether important findings replicate is lack of consensus and consistency in measuring core latent constructs. In other words, when the key trait constructs are measured inconsistently across studies, it is difficult to compare the results and build a systematic, replicable knowledge base. In the individual differences literature of temperament traits, this problem is surprisingly common, even when researchers use the same, frequently used measures, for example, because different studies combine different sets of items

or subscales. In this paper, we specifically focus on a frequently used measure of temperament traits in adolescents, the Early Adolescent Temperament Questionnaire Revised (EATQ-R, Ellis & Rothbart, 2001).

The EATQ-R has been widely adopted and used in numerous studies of adolescent temperament. For example, there are 240 citations to the original Ellis & Rothbart (2001) citation in Google Scholar as of March 2015. However, despite its widespread use, there is a lack of consensus among researchers regarding the core latent constructs measured by the EATQ-R. Specifically, its latent factor structure has not been definitively established, has not been used consistently across different studies, nor has it been consistently analyzed in line with the latent structural model postulated by Rothbart and colleagues (e.g., Derryberry & Rothbart, 1997; Putnam et al., 2001). This lack of a definitive latent structure has limited the ability to compare and interpret results about core temperament dimensions and associations across studies (Muris & Meesters, 2009), and thus has impeded the key goal of establishing the replicability of effects. Therefore, the main goal of the present study is to more definitively determine the factor structure of the EATQ-R, and then to test the resulting models with regards to important aspects of adolescent functioning.

Rothbart and colleagues developed the EATQ-R to assess the main facets postulated in their model of temperament in adolescents, building on their earlier scales for children[1]. The

[1] The original EATQ (Capaldi & Rothbart, 1992) consisted of 12 subscales, covering negative emotionality, positive emotionality, reactivity and self-regulation. However, factor analyses did not fully support this model; instead this early psychometric work yielded a variety of factor structures that did not clearly correspond to these dimensions of temperament (Capaldi & Rothbart, 1992; Kim, Brody, & Murry, 2003). Thus, Rothbart and colleagues revised and

EATQ-R subscales have been combined in different ways, as discussed below, but have most often been considered to represent three of the main temperament dimensions in Rothbart's model: NE, PE and EC. Specifically, the creators of the EATQ-R currently recommend combining the subscales into three main composite scales: (1) EC, consisting of the Attention, Activation Control, and Inhibitory Control subscales, (2) NE, consisting of the Aggression, Fear, Frustration and Shyness subscales (Depressed Mood is not included), and (3) PE, consisting of the Surgency, Pleasure Sensitivity, Perceptual Sensitivity and Affiliation subscales (Personal Communication, Lesa Ellis, August 1, 2007). However, this recommended grouping of subscales has not been published, and there have been no published confirmatory factor analyses.

---

expanded the EATQ (EATQ-R). Their goal was to better assess the core aspects of temperament in their model, especially aspects of temperament related to self-regulation (Ellis, 2001; Ellis & Rothbart, 2001; Putnam et al., 2001). The revised self-report scale includes 65 items to assess 11 subscales: Attention, Inhibitory Control, Activation Control, Fear, Shyness, Frustration, Surgency, Pleasure Sensitivity, and Perceptual Sensitivity, Affiliation, Aggression and Depressed Mood (see Measures and Table S1). Aggression and Depressed Mood have sometimes been presented by the scale developers as part of the negative emotionality temperament construct (Ellis, 2001), but at other times have been presented and used as separate measures of social-emotional functioning (Ellis & Rothbart, 2001; Putnam et al., 2001). A parent report version of the EATQ-R was also developed, which does not include the Pleasure Sensitivity and Perceptual Sensitivity scales (which were judged to be less observable to parents), and contains some additional items and different wording of items in other subscales (Ellis, 2001). Thus, self-report and parent versions are not directly comparable. In the current paper we thus focus on the more complete adolescent self-report version.

There have been several exploratory factor analyses of all or part of the EATQ-R (Ellis & Rothbart, 2001; Muris et al., 2007; Muris & Meesters, 2009; Putnam et al., 2001). However, these studies have produced inconsistent results, ranging from four (Ellis & Rothbart, 2001; Putnam et al., 2001) to nine (Muris & Meesters, 2009) components, which often do not readily correspond to the originally hypothesized temperament dimensions of EC, NE and PE. Thus, exploratory factor analyses have yielded mixed results which have not produced a clear, replicable factor structure of the EATQ-R, and the resulting factors have not always aligned clearly with the latent temperament dimensions they were designed to assess. In addition, and perhaps partly as a consequence of the lack of an established factor structure, the EATQ-R has not been used consistently across studies and in line with the latent structure postulated by Rothbart. Different research groups have excluded and included different subscales when assessing each core temperament dimension.[2]

[2] While effortful control has been fairly consistently assessed with all three subscales (Attention, Inhibitory Control and Activation Control; but see e.g., Oldehinkel, Hartman, De Winter, Veenstra, and Ormel (2004) for exclusion of Inhibitory Control), the composition of the negative emotionality and positive emotionality composite scales has been more inconsistent and controversial. For Negative Emotionality, many studies have excluded either or both Aggression and Depressed Mood from analyses (e.g., Mezulis, Simonson, McCauley, & Vander Stoep, 2011; Oldehinkel, Hartman, Ferdinand, Verhulst, & Ormel, 2007), while others have included them (Checa, Rodriguez-Bailon, & Rueda, 2008). Moreover, many studies include Shyness, Fear and Frustration in NE (e.g., Mezulis et al., 2011), while others have included only Frustration (e.g., Baetens et al., 2011). Assessment of positive emotionality has been even more mixed, with some authors including only Surgency (e.g., Mezulis et al., 2011), while others combine items

Relations Between the EATQ-R and Adolescent Functioning

Despite these measurement issues, the EATQ-R has been shown to predict many aspects of adolescent mental health and functioning. Higher EATQ-R effortful control has been shown to predict multiple positive outcomes, including lower levels of internalizing and externalizing symptoms and less impact of negative emotionality on symptoms (Muris, Meesters, & Blijlevens, 2007; Oldehinkel, Hartman, Ferdinand, Verhulst, & Ormel, 2007; Vasey et al., 2013), lower levels of interpersonal conflict (Swanson, Valiente, & Lemery-Chalfant, 2012; Yap et al., 2011), and higher school achievement (Checa & Rueda, 2011; Checa, Rodríguez-Bailón, & Rueda, 2008; Swanson et al., 2012). EATQ-R negative emotionality also predicts many negative outcomes, including adolescent depression (Loukas & Murphy, 2007; Mezulis & Rudolph, 2012; Mezulis, Simonson, McCauley, & Vander Stoep, 2011), externalizing symptoms and conduct problems (Loukas & Murphy, 2007; Muris et al., 2007), and interpersonal conflict (Yap et al., 2011). Last, positive emotionality, specifically surgency, has been associated with both positive and negative outcomes, including lower levels of internalizing symptoms (Oldehinkel, Hartman, De Winter, Veenstra, & Ormel, 2004) and higher levels of externalizing symptoms (Muris et al., 2007; Oldehinkel et al., 2004). Thus, the EATQ-R, especially at the super-factor level assessing EC, PE and NE, has been shown to predict important aspects of adolescent functioning, including psychopathology, interpersonal functioning, and academic achievement. However, these effects have not always replicated, potentially because of inconsistencies across studies in the way the EATQ-R was analyzed (e.g., which subscales are included).

The Current Study

---

from the Affiliation, Pleasure Sensitivity, and Perceptual Sensitivity subscales (e.g., Baetens et al., 2011; de Boo & Kolk, 2007).

In sum, while the EATQ-R has been widely used to assess adolescent temperament, its factor structure has not been established, and it has been used inconsistently. Overall, these limitations make it difficult to compare results across studies and reliably, systematically advance knowledge on temperamental traits at both the super-factor and specific facet level. Failure to find a clear and replicable factor structure may be due in part to the use of exploratory (EFA) rather than confirmatory factor analysis (CFA) methods (e.g., van Prooijen & van der Kloot, 2001), which is a more appropriate method when the goal is to identify latent constructs and there is a theoretical basis for specifying models *a priori* (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999). The current study therefore used CFA to test the factor structure of the EATQ-R adolescent self-report scale, and test links between the resulting latent temperament dimensions and important aspects of adolescent functioning.

In addition to taking a CFA approach, the study has several additional methodological strengths. We used a very large sample (n= 2026) of adolescents collected across six independent studies conducted at different sites. The large sample allows for more precise estimates and testing of more complex models. The use of data from geographically and demographically diverse sites enhances the robustness and generalizability of the findings. Second, given the large sample size, we were able to split the data into one set for model development and initial testing and a second hold-out set for replication of the resultant models. Demonstrating that the final models generalize well to the hold-out data set provides needed replication and ensures the models are not over-fitted to idiosyncratic features (noise) in the data set used for model development.

We first tested the factor structure of the EATQ-R by testing models of the three dimensions of temperament as currently hypothesized by Rothbart and colleagues: EC, NE, and

PE. Next, we tested relations among these dimensions in a model of the full scale. Finally, we assessed relations between the final EATQ-R temperament models and aspects of adolescent functioning hypothesized to be related to temperament, including social functioning (antisocial behavior towards peers and victimization by peers), school functioning (grades and school disciplinary action) and psychopathology (depression, anxiety and ADHD symptoms). Based on the literature discussed above, we predict that (1) higher EC should be associated with lower levels of psychopathology and better interpersonal and school functioning, (2) higher NE should be associated with higher levels of psychopathology broadly and more interpersonal problems, and that specific aspects of NE should further show specificity with corresponding specific aspects of psychopathology (e.g., EATQ-R Fear with harm avoidance). Predictions for PE are less clear given the relative paucity of research and mixed findings with this temperament dimension. But, we reasoned that if the PE scale does capture positive emotionality as hypothesized in Rothbart's model, it should be associated with lower levels of psychopathology, perhaps especially depression (e.g., Anderson & Hope, 2008).

Method

Participants

EATQ-R self-report data were available from six different samples across five sites, and were combined to achieve a large sample size for the current study that spans the full age range in which the EATQ-R is commonly used (from late childhood through late adolescence) and multiple geographic regions. The combined sample had a total of *n*= 2026 participants (56% female), with a mean age of 13.02 years (SD=2.57, range 8-19). Community samples of adolescents were recruited from (1) public schools in the Denver, CO metro area (*n*= 294; e.g., Barrocas, Hankin, Young, & Abela, 2012; Hankin, Jenness, Abela, & Smolen, 2011), (2) public

schools in the New Brunswick, NJ metro area (*n*= 242; e.g., Barrocas, Hankin, Young, & Abela, 2012), (3) five municipalities in the north of The Netherlands, including both urban and rural areas (*n*=340[3]; e.g., Huisman et al., 2008; Ormel et al., 2012), (4) Belgian secondary schools (*n* =307; Willem, Bijttebier, Claes, Vanhalst, & Raes, 2013), (5) Belgian elementary and secondary schools (*n*=588; Verstraeten, Vasey, Claes, & Bijttebier, 2010; Verstraeten, Vasey, Raes, & Bijttebier, 2010), and (6) public and private middle schools in the Seattle, WA metro area (*n*=220). Additional measures of adolescent functioning (see Measures) were available from samples 1 and 2 (*n* = 562). For all samples, participants provided informed consent (parents for adolescents under 18 and adolescents 18 and older) and assent (adolescents under 18), and were treated in compliance with procedures approved by their appropriate local human subjects review boards.

Measures

EATQ-R. Participants were administered the full EATQ-R self-report scale in all studies except for sample 6 (see Participants), in which participants were administered a subset of the subscales: Frustration; Shyness; Fear; Activation Control; Attention; Inhibitory Control; and Surgency. Samples 1, 2, and 6 completed the English version of the EATQ-R, while samples 3, 4, and 5 completed the Dutch version. The comparability of the Dutch translation (Hartman, 2000) was verified through a backward and forward translation procedure and piloting, in consultation with the EATQ-R developers (Rothbart, Evens and Ellis), and the Dutch version has been used extensively in previous research (e.g., Baetens et al., 2011; de Boo & Kolk, 2007;

[3] A random subsample of 340 participants was selected from the total Tracking Adolescents' Individual Lives Survey (TRAILS) sample (*n*=2230) to match the mean number in the other datasets, so that each dataset has approximately equal weight in the analyses.

Oldehinkel et al., 2004, Oldehinkel, Hartman, Ferdinand, Verhulst, & Ormel, 2007; Oldehinkel, Hartman, Nederhof, Riese, & Ormel, 2011). For all samples, the rate of missing data for all items administered was low (0.4% total). See Tables S1 and S4 for descriptive statistics. The EATQ-R subscales are as follows (http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/instrument-descriptions/early-adolescent-temperament.html):

EC subscales. Activation Control consists of 5 items assessing ability to begin and complete tasks when there is a strong tendency to avoid it. Attention consists of 6 items assessing ability to focus and sustain attention as well as to shift attention when desired. Inhibitory Control consists of 5 items assessing ability to suppress or stop inappropriate behaviors, wait and plan before acting.

NE Subscales. Aggression consists of 6 items assessing hostile reactivity and aggressive physical and verbal actions. Depressed mood consists of 6 items assessing lowered mood, and loss of enjoyment and interest in activities. Fear consists of 6 items assessing anticipation of distress, including worry and fear. Frustration consists of 7 items assessing negative affect related to interruption of ongoing tasks or goal blocking. Shyness consists of 4 items assessing behavioral inhibition to novelty and challenge, especially in social situations.

PE Subscales. Perceptual Sensitivity consists of 4 items assessing awareness of slight, low-intensity stimulation in the environment. Pleasure Sensitivity consists of 5 items assessing pleasure related to activities or stimuli involving low intensity. Affiliation consists of 5 items assessing the desire for warmth and closeness with others. Surgency consists of 6 items assessing pleasure derived from activities involving high intensity or novelty.

Adolescent Functioning Measures. In samples 1 and 2 (see Participants) participants or their parents additionally completed the following measures.

Children's Depression Inventory (CDI). The CDI (Kovacs, 1985) is a 27-item self-report measure of depressive symptoms. The CDI has adequate internal consistency and 1-month test–retest reliabilities and the scale correlates with clinician-rated depression (e.g., r = .55; Kovacs, 1992). In the current sample the CDI had good internal consistency (α = .88). The rate of missing data was 2.1%.

Manifest Anxiety Scale for Children (MASC). The MASC (March, Parker, Sullivan, Stallings, & Conners, 1997) is a widely used self-report measure of anxious symptoms in children and adolescents. The MASC contains 39 items that assesses the subscales (1) physical symptoms of anxiety, (2) harm avoidance, (3) social anxiety, and (4) separation anxiety/panic. The MASC has high internal consistency (Muris, Merckelbach, Ollendick, King, & Bogie, 2002) and test–retest reliability (March et al., 1997). In the current sample, the MASC subscales all had adequate internal consistency (α > .75 for all subscales). The rate of missing data was 2.8%.

MTA Swanson, Nolan, and Pelham scale (MTA SNAP-IV). Parents completed the NIMH Collaborative Multisite Multimodal Treatment Study of Children With Attention-Deficit/Hyperactivity Disorder (MTA) version of the SNAP-IV (Swanson et al., 2001). The SNAP-IV questionnaire includes the 18 DSM-IV criteria for ADHD, with 9 items assessing inattention, 5 hyperactivity, and 4 impulsivity. It is frequently used in research and clinical settings to diagnose ADHD subtypes. The measure is reliable (α = .94) and valid (Bussing et al., 2008). In the current sample, the SNAP subscales all had good internal consistency (α > .80 for all subscales). The rate of missing data was 1.9%.

Revised Peer Experiences Questionnaire (RPEQ). The RPEQ is a self-report measure of antisocial behavior towards peers and victimization by peers (De Los Reyes & Prinstein, 2004; Prinstein, Boergers, & Vernberg, 2001). Participants in the current study completed a shortened

version of the scale, containing seven items each about antisocial behavior towards peers (excluding, damaging the reputation of, or behaving aggressively toward peers), and victimization by peers (being the victim of the previously listed behaviors by peers). Other versions of the RPEQ have shown good internal consistency and validity (De Los Reyes & Prinstein, 2004). In the current sample, the antisocial behavior toward peers (α= .72) and victimization by peers (α=. 79) scales both had adequate internal consistency. The rate of missing data was 3.3%.

School behavior and grades. Parents reported on their child's typical letter grades, from "mostly A's" (1) to "mostly F's" (5). The rate of missing data was 3.9%. Parents also reported the number of times their child had been sent to the office for misbehavior during the year, from none (1) to more than five times (6). The rate of missing data was 1.1%.

Data Analytic Plan

EATQ-R Factor Structure Analyses. We divided the data at random into two sets (*n*=1013 in each set), one for model development and one as a hold-out set for replication of the final models. Confirmatory factor analyses (CFA) were tested with Mplus (Muthén & Muthén, 2012), using full information maximum likelihood (FIML) estimation to address missing data. For all models, factor variance was set to 1, to allow all item loadings to be estimated (rather than setting an item loading to 1); item loadings are thus standardized with respect to latent variable variance (i.e., STD Standardized). For all models, good fit was defined as RMSEA < .05, CFI >.95, and acceptable fit was defined as RMSEA < .08, CFI >.90 (Hu & Bentler, 1999).

Model development and testing with the first data set was conducted with the following steps. First, individual factor models were run for each subscale. Each model was checked for adequate item loadings; since even weak factor loadings are significant given the large sample

size, .30 was chosen as a cut-off for acceptability (e.g., Kline, 2010), below which items were removed. For subscale models that did not have good fit, modification indices were examined and correlations between item residual variances were added in order of largest to smallest modification index values until good model fit was achieved (e.g., Mueller & Hancock, 2008). These modifications were then included in all further models. To prevent over-fitting models, no new modifications were added in subsequent models. Second, for each super-scale (EC, NE, PE), the relevant subscale models from step 1 were modeled together, with correlations between all subscale factors. Chi-square difference tests were used to compare model fit for each correlated factor model to a one-factor model with all super-scale items loading on a single factor.

Finally, we tested bifactor models for each super-scale, in which all items loaded onto a common factor representing the shared variance across items in that super-scale, as well as loading on their specific subscale factor that represent the unique variance associated with each subscale not accounted for by the common factor (e.g., Chen, 2006; Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Friedman et al., 2008). Factor correlations are set to 0 because what is shared between factors is already captured by the common factor (Chen, 2006; Muthén & Muthén, 2012). This bifactor parameterization has two advantages over hierarchical models in which first-level factors load onto a higher-level factor. First, from a practical standpoint, convergence problems are very common for hierarchical models, whereas this is not frequently a problem for bifactor models. Second, from a conceptual perspective, bifactor models allow examination of how other variables are related to both the common and specific aspects of a construct (Chen, 2006; Chen et al., 2012; Friedman et al., 2008), about which there are frequently distinct hypotheses (e.g., general NE vs. what is unique to depressed mood).

Full bifactor models were tested first and then modified based on the significance of

factor variances and pattern of loadings.[4] If variance was not significant for a specific factor, this was taken as evidence that the specific factor was not needed to account for variance on the items in that subscale (i.e., they are fully accounted for by the common factor), and the specific factor was eliminated. In contrast, if items from a subscale loaded strongly on their specific factor but had any non-significant and/or negative loadings on the common factor, this suggested that that subscale was best considered a separate factor, and loadings on the common factor were eliminated. Model fit for these final models was then compared to that of the correlated factor and one-factor models. In all cases, the bifactor models fit the data best. Thus, only the final bifactor models are reported here. Results from the individual subscale models and correlated subscale models are reported in Supplemental Materials.

Correlations with Adolescent Functioning. After development of final EATQ-R models, correlations were tested between the factors in each EATQ-R dimension model (EC, NE and PE) and each of the adolescent functioning measures (CDI, MASC, SNAP, RPEQ antisocial behavior towards peers, REPQ victimization towards peers, school behavior, grades) in samples 1 and 2, for which these measures were available. As school behavior and grades were assessed with a single question each, they were analyzed as manifest variables. All other measures were analyzed as latent variables, based on their established subscale structure. The MASC and SNAP have correlated subscales that have been supported by previous factor-analyses analysis (March et al., 1997; Pillow, Pelham, Hoza, Molina & Stulz, 1998). Thus, they were analyzed using bifactor models, with common and subscale-specific factors. The MASC model included a

[4] To test the significance of factor variances, models were re-parameterized with the first item loading constrained to 1 instead of factor variance constrained to 1. Model fit is identical regardless of which parameterization is used.

common factor and specific factors for Physical Symptoms, Separation/Panic and Harm avoidance subscales; the Social Anxiety subscale was fully accounted for by the common factor (i.e., there was not significant variance associated with the specific factor, so it was eliminated.) The SNAP model included a common factor and Inattention, Hyperactivity and Impulsivity specific factors. The CDI, RPEQ antisocial behavior towards peers, and REPQ victimization towards peers scales were each analyzed as single factors.[5]

Finally, we compared the findings regarding links to adolescent functioning using our final EATQ-R models to those obtained using the traditional method of computing temperament dimension scores from the EATQ-R. Specifically, we examined correlations between each of our adolescent functioning measures and the three main composite measures as currently recommended by the creators of the EATQ-R (Personal Communication, Lesa Ellis, August 1, 2007): (1) EC, consisting of the Attention, Activation Control, and Inhibitory Control subscales, (2) NE, consisting of the Aggression, Fear, Frustration and Shyness subscales (Depressed Mood is not included), and (3) PE, consisting of the Surgency, Pleasure Sensitivity, Perceptual

[5] Exploratory factor analyses have reported multiple factors for the CDI, but the number of factors and the items loading on each have varied widely across studies (e.g., Garcia, Aluja, & Del Bario, 2008), and in the current sample the five subscales proposed in the CDI manual (Kovacs, 1992) were not supported: extremely high correlations between subscale factors indicated that the scale was better treated as unitary, which is also consistent with common practice in analyzing the CDI as a single score without subscale scores. An exploratory factor analysis of the full version of the RPEG found support for multiple factors representing different types of aggression. However, with only seven items in the short version used in the current sample, examination of subscales was not possible.

Sensitivity and Affiliation subscales.

Results

Model development was conducted in Dataset 1 (n =1013), and model replicability was tested in the hold-out data set (n =1013).

EATQ-R Effortful Control

EC models included the Attention, Inhibitory Control and Activation Control subscales. The initial version of the bifactor model, with a Common EC factor and specific factors for each EC subscale, demonstrated that there was no significant variance associated with the Inhibitory Control-Specific or Attention-Specific factors, but there was significant variance for the Activation Control-Specific factor and Common EC factor. We therefore modified the model to eliminate the Inhibitory Control-Specific and Attention-Specific factors (Figure 1). Model fit was acceptable by RMSEA and nearly acceptable by CFI (Table 1). Model fit was significantly better than the one factor model ($\Delta\chi^2$ (5) = 161.17 p <.001) and equivalent to the correlated subscale model ($\Delta\chi^2$ (2) = 2.13 p=.34) while being more parsimonious. Thus, this model was used in all further analyses.

Negative Emotionality

NE models included the five NE subscales: Aggression, Depressed Mood, Fear, Frustration and Shyness.[6] The Common NE factor and all specific factors had significant variance and thus were retained. One item (37) had a weak negative loading on the Depressed Mood-Specific factor (it loaded very strongly on Common NE), and was therefore eliminated

[6] Given controversy about the inclusion of Aggression and Depressed mood in NE, we also report models excluding these subscales in Supplementary Materials.

from the Depressed Mood-Specific factor. Model fit was good by RMSEA and acceptable by CFI (Table 1). Model fit was significantly better than both the one factor model ($\Delta\chi^2$ (26) = 1309.59, p <.001) and the correlated subscale model ($\Delta\chi^2$ (16) = 108.76 p <.001). Thus, this model was used in all further analyses.

Positive Emotionality

PE models included the four PE subscales: Affiliation, Pleasure Sensitivity, Perceptual Sensitivity, and Surgency. Surgency items did not load adequately on the Common PE factor, with two negative loadings and the remaining loadings < .3. Thus, the model was modified to not load Surgency items on the common factor, leaving Common PE, Surgency (as a separate factor), Affiliation-Specific, Perceptual Sensitivity-Specific and Pleasure Sensitivity-Specific factors (Figure 3). There was significant variance for all factors, so they were retained. One item (24) had a non-significant loading on the Perceptual Sensitivity-Specific factor, so it was eliminated from that factor (it loaded adequately on the Common PE factor). The Surgency factor was only very weakly correlated with Common PE or any of the specific factors. Model fit was good by RMSEA and acceptable by CFI (Table 1). Model fit was significantly better than both the one factor model ($\Delta\chi^2$ (18) = 1306.06, p <.001) and the correlated subscale model ($\Delta\chi^2$ (12) = 120.45 p <.001). Thus, this model was used in all further analyses.

Full Combined Model

Next, we combined the final bifactor EC, NE, and PE/Surgency models in order to examine latent correlations among factors across the three temperament dimensions. Initially, this model was not positive definite. Examination of factor loadings revealed that in the full model there was no longer evidence of an Affiliation-specific factor (negative or very weak positive loadings on all but one item). Thus, the Affiliation-Specific factor was eliminated (i.e.,

Affiliation items were loaded only onto the Common PE factor). This allowed the model to run successfully. The model had good fit by RMSEA but not by CFI (Table 1).

Correlations between all factors in the full model are presented in Table 2, and item loadings in Table S3 (Supplementary Materials). Alpha was set to $p < .0005$ using Bonferroni correction for the number of correlations to set the family-wise error rate to .05. Common EC was negatively correlated with most NE factors (Common NE $r = -.48$, Aggression-Specific $r = -.42$, Depressed Mood-Specific $r = -.33$, and Frustration-Specific $r = -.23$), while the Activation Control-Specific factor was *positively* correlated with the Fear-Specific ($r = .36$) and Common NE ($r = .24$) factors. Common EC did not correlate with any of the PE factors ($r < +/-.15$), but the Activation Control-Specific factor correlated positively with the Pleasure Sensitivity-Specific factor ($r = .27$) and negatively with Surgency ($r = -.28$). The Fear-Specific factor was strongly negatively correlated with the Surgency factor ($r = -.61$) and more weakly with Common PE ($r = -.30$), but positively correlated with the Pleasure Sensitivity-Specific factor ($r = .40$). The Depression-Specific factor correlated strongly negatively with Common PE ($r = -.63$) but positively with the Perceptual Sensitivity-Specific factor ($r = .34$). The Shyness-specific factor was negatively correlated with Surgency ($r = -.25$). Unexpectedly, Common NE and Common PE were *positively* correlated ($r = .55$), an issue we return to when discussing construct validity.

Replication of Final Models in Hold-Out Sample

To test replicability, we ran each of the final bifactor models in the hold-out data set (n =1013). For all models, model fit was similar in the model development and hold-out data set, although fits were slightly worse in the hold-out dataset (Table 3). Comparing factor loadings between datasets, there was little bias (i.e., factor loadings were not systematically higher in one set, difference= -0.03-0.00 across models), and relatively small differences in factor loadings on

average (absolute value of differences= 0.06-0.09 across models). By way of comparison, average factor loadings across models were .44-.51, with average SEs= .04-.06, so these differences between datasets were modest on average relative to the size and precision of the loading estimates in each dataset. For the full model, correlations between factors were also similar across datasets on average (absolute value of differences= 0.07), with little bias (difference= -0.01; Table 3). Again by way of comparison, the average absolute value of the correlations was .21, average SE= .06, so these differences between datasets were again modest on average relative to the size and precision of the correlation estimates in each dataset. Values were similar when excluding near zero correlations (< .2) in both models (average diff.= .03, average abs. diff.= .08, average absolute *r*= .30, average SE= .06).

Correlation with Measures of Adolescent Functioning

Finally, to assess construct validity and the relation between temperament and functioning, we correlated each of the final EATQ-R models (EC, NE and PE) with models for each of the adolescent functioning measures in Samples 1 and 2 (*n* =562): depression (CDI), anxiety (MASC), ADHD (SNAP), interpersonal functioning (RPEQ antisocial behavior towards peers and victimization by peers) and school functioning (grades and school discipline problems). Alpha was set to $p < .0003$ using Bonferroni correction for the number of correlations to set the family-wise error rate to .05. Full correlation results are presented in Table 4.

Correlations with EATQ-R EC. Higher Common EC was associated with fewer symptoms of depression (CDI, *r* = -.58), anxiety (Common MASC *r* = -.38, MASC physical symptoms-specific *r* = -.28) and ADHD (Common SNAP, *r* = -.25), less antisocial behavior towards peers (RPEG Antisocial, *r* = -.45), less victimization by peers (RPEG Victim, *r* = -.35), better grades (*r* = .36), fewer school discipline problems (*r* = -.18) and more harm avoidance

(MASC Harm Avoidance-specific, *r* = .38). In contrast, the Activation-specific factor was only associated with more harm avoidance (*r* = .39).

Correlations with EATQ-R NE. Higher Common NE was associated with more symptoms of depression (CDI, *r* = .57), anxiety (Common MASC *r* = .75, MASC physical symptoms-specific *r* = .17), more antisocial behavior towards peers (RPEG Antisocial, *r* =.36), and more victimization by peers (RPEG Victim, *r* =.34). Higher Aggression-specific was associated with more antisocial behavior towards peers (RPEG Antisocial, *r* =.46), and more victimization by peers (RPEG Victim, *r* =.37), lower grades (*r* =-.35), and more school discipline problems (*r* =.26). Higher Depressed Mood-specific was associated with more depression symptoms (CDI, *r* =.50) and more physical symptoms (MASC physical symptoms-specific, *r* =.39). Higher Fear-specific was associated with more anxiety symptoms (MASC Separation/Panic-specific *r*=1.0, MASC Harm Avoidance-specific, *r* =.42) and lower grades (*r* =.-.25). The Frustration-specific and Shyness-specific factors were not significantly associated with any measures (*r* <.2), potentially because measures that might be expected to correlate specifically with these factors (e.g., loneliness with shyness) were not included in the current study.

Correlations with EATQ-R PE. Common PE was associated with higher levels of MASC Harm Avoidance-specific (*r* =.32). Surgency was associated with lower levels MASC Separation/Panic-specific (*r* =-.35). Affiliation-specific was associated with higher CDI depressive symptoms (*r* =.31), higher Common MASC (*r* =.50), but lower MASC Separation/Panic-specific (*r* =-.31)., more antisocial behavior towards peers (RPEG Antisocial, *r* = .43), and more victimization by peers (RPEG Victim, *r* =.30). Perceptual-Sensitivity-specific was associated with higher CDI depressive symptoms (*r* =.23)., higher MASC Physical-Symptoms-specific (*r* =.28), and more victimization by peers (RPEG Victim, *r* = .32).

Comparison with Traditional Methods of Analyzing the EATQ-R. In contrast to analyses using the final latent variable models of each EATQ-R dimension, analyses using EATQ-R manifest super-scale (EC, NE, and PE) and subscale measures differed in two main ways. First, they often yielded patterns of correlations with adolescent functioning variables that were much less specific than those found with the latent variable models (Table S5). These effects probably arose because each manifest subscale score is a mixture of common (e.g., common NE) and specific (e.g., frustration-specific) variance, such that common variance can drive correlations with other measures and lead to false conclusions that a specific aspect of temperament is related to those variables. For example, all manifest NE subscale scores were correlated with CDI depression and MASC anxiety, whereas the latent variable models demonstrated that these correlations were actually driven by a combination of common NE and aspects of NE specific to each form of psychopathology. Likewise, while all manifest EC subscale scores, including activation control, were correlated with depression, anxiety, interpersonal functioning and grades, the latent variable models demonstrated that these effects were specifically related to common EC, and unrelated to the activation control specific factor. Second, contamination by common variance sometimes appeared to mask specific effects. For example, the relationship of affiliation with increased depression, anxiety and interpersonal problems was much weaker for the manifest Affiliation subscale score than the Affiliation-specific factor. Both of these problems were further exacerbated when manifest variables were used for the adolescent functioning measures as well (Table S7).

Discussion

While the EATQ-R has been widely and productively used to assess adolescent temperament, its factor structure has not been definitively established, making it difficult to

establish construct validity and replicate key findings related to adolescent temperament. Indeed, lack of consensus and consistency in measuring core latent constructs is a common problem across many areas of psychology, leading to difficulties in comparing results and building a systematic, replicable knowledge base. Confirmatory factor analytic methods can be a valuable approach for establishing robust empirically and theoretically-justified models needed for replication and for better understanding links between key constructs of interest.

The current study therefore conducted the first confirmatory factor analysis of the EATQ-R, using data collected from six separate studies, to test alternative hypothesized structural models of the three key dimensions of Rothbart and colleagues' temperament model, effortful control (EC), negative emotionality (NE) and positive emotionality (PE), and the underlying facets of these super-factors. Importantly, these models replicated in a hold-out dataset, suggesting that the results are robust and likely to generalize. Furthermore, these models revealed links between dimensions of temperament and important aspects of adolescent functioning, including psychopathology, interpersonal functioning, and school functioning, which are hypothesized by the literature but not always apparent using previous ways of analyzing the EATQ-R. These associations demonstrate the utility of these newly developed EATQ-R models for understanding links between adolescent temperament and functioning. Below we discuss the main insights from the current study into the constructs of EC, NE and PE, the relations among them, and their relation to measures of adolescent functioning. We include suggested directions for future research throughout.

EATQ-R Models

Our results indicate that temperament, as assessed via the EATQ-R, cannot be reduced simply to the three dimensions of EC, NE and PE. Rather, there is both unity and diversity within

each of these dimensions. Specifically, the best fitting models for these dimensions of temperament were ones in which there was both a common latent factor capturing what is shared across subscales in that construct as well as specific latent factors capturing what is unique to items in particular subscales. Importantly, while this bifactor modeling approach has not previously been applied to the EATQ or other measures of temperament, it has been found to best account for the structure of adult personality traits (e.g., Chen et al., 2012; Costa & McCrae, 1995), adolescent personality disorder traits (e.g., Roose, Bijttebier, Decoene, Claes, & Frick, 2010), and dimensions of psychopathology in both adults and adolescents (e.g., Caspi et al., 2014; Lahey et al., 2012; Noordhof, Krueger, Ormel, Oldehinkel, & Hartman, 2014; Tackett et al., 2013). Moreover, as we discuss below, these bifactor models enable investigation of links between other measures and both common and specific facets of each temperament dimension, revealing a more nuanced picture of how temperament affects adolescent functioning.

Thus, these results suggest that a more complex approach to analyzing and interpreting the EATQ-R is needed, as opposed to using one single, summed super-scale for each dimension of temperament as is currently common practice with the EATQ-R. Specifically, we suggest that whenever the sample size is sufficient, the EATQ-R should be analyzed using latent variable models rather than a manifest variable approach.[7] Specifically, only latent variable models

[7] If sample size is too small to permit CFA analysis, the results of the current study suggest that the use of individual subscale scores may be justified, as subscale factors generally had good fit. However, researchers may wish to consider dropping items 3, 19, and 41, which did not load well on their subscales, or on the final bifactor models (from which they were dropped). Our results further suggest that, with the exception of the Attention and Inhibitory Control subscales, any combination of subscales at the manifest level should be done with caution and the

enable (1) separation of common and specific factors for each temperament dimension, which in turn allow more specific and nuanced links to other variables of interest to be examined, and (2) elimination of error variance, improving power to detect relations with other variables of interest. While switching to a latent variable approach may inevitably pose some challenges (e.g., time required for analysis and the need for relatively larger sample sizes), it also opens many exciting possibilities for achieving new insights into adolescent temperament dimensions and their relation to important life outcomes, both in new studies and through re-analysis of existing datasets.

Effortful Control. EC is defined in Rothbart's temperament model as a self-regulatory component that supports the ability to appropriately control behavior and attention (e.g., Putnam et al., 2001; Rueda, Posner, & Rothbart, 2005). This model includes three aspects of EC: (1) attention (capacity to focus and shift attention appropriately), (2) inhibitory control (capacity to suppress inappropriate responses and plan future action), and (3) activation control (ability to perform an action when there is a strong tendency to avoid it). The findings from this study indicate that most variance in EC as measured by the EATQ-R is accounted for by what is common across these three aspects of EC (Common EC),[8] but there was also a factor specific to

---

understanding that the resulting measures will be an imprecise admixture of both common and specific variances for each temperament dimension. Further, our results suggest that the EATQ-R should not be used to assess PE, given the lack of demonstrated construct validity.

[8] One reason that Common EC may fully account for Inhibitory Control and Attention in the EATQ-R, despite hypothesized differences between these two components, is that the Attention subscale items almost exclusively ask about sustained maintenance of attention to complete goals (with the exception of two items, one of which did not load adequately). These attention items

Activation Control. This structure has not been previously captured by EFA analyses of the EATQ-R, or of Rothbart's temperament measures for children (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) or adults (ATQ; Evens & Rothbart, 2007), which have combined all aspects of EC into a single factor, which in the case of the CBQ also contained seemingly unrelated aspects of temperament.

This structure, with both common and unique aspects of EC, parallels that of the closely related construct of executive function (EF), where there are both specific EF abilities and a common EF ability, which spans these components (Friedman et al., 2008; Miyake & Friedman, 2012). While future research is needed to determine how EC and EF constructs are related, the EATQ-R Common EC factor may represent a similar construct to this common EF factor, which is posited to be the ability to actively maintain goals and use them to guide behavior (Miyake & Friedman, 2012). Likewise, most EC items in the EATQ-R are related to the ability to maintain goals and use them to control behavior, such as maintaining focus on tasks, completing tasks, and following rules. In contrast, given the focus of Activation Control items on diligently completing schoolwork, and its correlations with other measures discussed below, we

---

are likely to be strongly related to Inhibition items, which are related to the maintenance of goals to direct behavior appropriately. Adding items related to shifting attention to potential future revisions of the EATQ-R might allow an attentional shifting component of EC to be differentiated from inhibition and Common EC, much as there is a shifting-specific component of EF that is separable from Common EF ability (Friedman et al., 2008; Miyake & Friedman, 2012). Alternatively, it may be that, while the distinction can be made in fine-grained neuropsychological EF tasks, inhibition and attention turn out to be fully intertwined when assessed by questionnaires as molar aspects of behavior in daily life.

hypothesize that the Activation Control-Specific factor may represent the motivation to be thorough, self-disciplined, and do tasks well, similar to the adult personality trait of conscientiousness (e.g., Costa & McCrae, 1995), and may include the over-control and fear of failure (e.g., harm avoidance) sometimes also seen with high conscientiousness (e.g., Boyce, Wood, Brown, 2010). We return to this issue when discussing correlations among temperament dimensions and with adolescent functioning measures below.

Negative Emotionality. Negative emotionality is a broad construct subsuming emotions including anxiety, sadness, frustration, anger, and discomfort (e.g., Putnam et al., 2001). NE has consistently emerged in Rothbart's model as a dimension of temperament from infancy to adulthood (e.g., Putnam et al., 2001) and also features prominently in many other models of temperament and personality (e.g., for review see Tackett et al., 2013) The findings from the current study indicate that there is both a common NE factor and emotion-specific factors for each NE subscale (Aggression, Depressed Mood, Fear, Frustration, and Shyness).[9] This finding is consistent with the views of Rothbart and colleagues (e.g., Evans & Rothbart, 2007), although the existence of common and specific NE factors had not been directly tested with appropriate latent variable analyses in the EATQ or other temperament measures (e.g., CBQ, ATQ) previously. Importantly, psychopathology, which is closely linked to negative emotionality (e.g., for review see Lahey, 2009; Tackett et al., 2013), has also been shown to consist of both common and specific factors. Specifically, bifactor models of psychopathology in both

[9] This was true regardless of whether Aggression and Depressed Mood were included. See Supplemental Material for models excluding these subscales and discussion of the pros and cons of including them depending on study goals

adolescents and adults find that there is a common factor that spans all aspects of common psychopathologies and is related to broad negative emotionality, in addition to factors for more specific aspects of psychopathology (Caspi et al., 2014; Lahey et al., 2012; Tackett et al., 2013; c.f. Tellegen, Watson, & Clark, 1999). Thus, the current results are highly consistent with evidence from other measures for both common and specific aspects of NE.

Positive Emotionality. Positive emotionality is a broad construct that involves directing approach behavior towards reward, including positive anticipation, sociability, active engagement with the environment, and positive affect (e.g., happiness), and it overlaps with constructs such as extraversion and behavioral activation (e.g., Derryberry & Rothbart, 1997; Muris & Ollendick, 2005). The current recommendation of Rothbart and colleagues to combine the Surgency, Affiliation, Perceptual Sensitivity and Pleasure Sensitivity scales into a PE composite scale (Personal Communication, Lesa Ellis, August 1, 2007) were not supported by the current study.[10] The Surgency subscale (which is closely related to novelty or sensation seeking), failed to load on a common factor with the other PE subscales, and instead it formed a separate and largely uncorrelated factor. Thus if surgency is the construct of interest, only the Surgency subscale should be used. If, on the other hand, PE more broadly construed is the construct of interest, the common factor formed by Affiliation, Perceptual Sensitivity and

[10] The original exploratory factor analysis of the EATQ-R (Ellis & Rothbart, 2001) found separate factors for Surgency and the three other subscales, which formed a factor that was termed *affiliativeness*. The results of the current study using CFA are consistent with this original exploratory factor analysis. While this factor was previously termed *affiliativeness*, affiliation is only one component of this factor. Thus, there is lack of clarity about what exactly this construct is measuring via these EATQ-R items.

Pleasure Sensitivity does not measure this latent construct of PE as usually defined by most temperament and personality theorists.

First, in terms of face validity, some items, especially the Perceptual Sensitivity subscale, are not clearly conceptually related to PE as generally conceptualized, while other core components of PE, such as the experience of positive emotions (e.g., happiness) are not directly queried. Second, as discussed below, the Common PE factor did not correlate with other measures as would be expected if it represented PE (e.g., the Common PE factor positively correlated with NE, whereas other measures of PE and NE are generally uncorrelated or weakly negatively correlated (e.g., Crawford & Henry, 2004; Lonigan, Hooe, David, & Kistner, 1999). We speculate, based on the items in the component subscales, that the Common PE factor may instead represent *sensitivity and reactivity to the environment*, which has been proposed as a core temperament trait in other temperament models (e.g., Aron, Aron, & Jagiellowicz, 2012). Individuals who are high in sensory processing sensitivity are better at perceiving subtle sensory and social cues, and, as a consequence, are more emotionally reactive to the environment, responding to negative, or over-stimulating, environments with increased NE and to positive environments with increased PE (Aron et al., 2012).

In sum, the PE subscales appear to be measuring two distinct constructs—surgency/sensation-seeking and general sensitivity. This division is more consistent with temperament dimensions in Rothbart and colleagues measures at other ages. Specifically, the Children's Behavior Questionnaire (Rothbart et al., 2001) as an extraversion/surgency factor rather than a general PE factor. In addition, the Adult Temperament Scale (Evans & Rothbart, 2007) includes a dimension termed *orienting sensitivity,* which includes items similar to the Perceptual Sensitivity and Pleasure Sensitivity subscales of the EATQ-R, as well as an

extraversion/surgency dimension. Neither of these closely matches the latent trait construct of broad PE as generally conceptualized by trait theorists. Thus, future research may wish to consider including alternative measures of PE (e.g., Positive Affect subscale from the Adult Temperament Questionnaire (Evans & Rothbart, 2007) and/or the PANAS-C (Laurent et al., 1999)).

Correlations among temperament dimensions and measures of adolescent functioning

The majority of correlations among temperament factors, and between temperament factors and measures of adolescent functioning, were consistent with previous theory, although a few associations were unexpected. Critically, the EATQ-R latent variable models revealed specific links between facets of temperament and aspects of adolescent functioning which, while hypothesized in the literature, are often not apparent when using traditional manifest measures of temperament. Specifically, analyses using EATQ-R manifest super-scale (EC, NE, and PE) and subscale measures frequently yielded a much less specific pattern of correlations with adolescent functioning measures because each manifest subscale score is a mixture of common (e.g., common NE) and specific (e.g., frustration-specific) variance, such that common variance can drive correlations with other measures and lead to false conclusions that a specific aspect of temperament is related to those variables. In other cases contamination by common variance appears to mask specific effects when using manifest temperament variables. In both cases, our latent variable models are able to provide a much more nuanced picture of how specific aspects of adolescent temperament are related and affect important adolescent outcomes, and suggest topics for future research.

Effortful Control. As expected based on previous research associating poor EC with negative affect and psychopathology (Muris et al., 2007; Oldehinkel et al., 2007; Vasey et al.,

2013), Common EC was negatively correlated with NE temperament factors and with external measures of psychopathology symptoms, including symptoms of depression, anxiety, and ADHD, consistent with previous research (e.g., Muris et al., 2007; Oldehinkel et al., 2007; Vasey et al., 2013). It has been proposed that high EC may enable individuals to over-ride maladaptive responses, including impulsive and aggressive behaviors, attentional biases towards negative information, and repetitive negative thinking patterns (rumination, worry), thus reducing negative affect and risk for psychopathology (e.g., Ellis, Rothbart, & Posner, 2004; Lonigan & Vasey, 2008; Vasey et al., 2013). Unlike other aspects of anxiety, higher levels of harm avoidance were associated with better EC. While often considered an aspect of anxiety, harm avoidance is also related to lower levels of risk taking (e.g., Wills, Sandy, & Shinar, 1999), which in turn is related to effortful control (e.g., Magar, Phillips, & Hosie, 2008). Additionally, consistent with prior research showing that individuals with good effortful control have better social and academic outcomes (e.g., Checa & Rueda, 2011; Checa et al., 2008; Swanson, Valiente, & Lemery-Chalfant, 2012; Yap et al., 2011), Common EC was also associated with better interpersonal functioning (less antisocial behavior towards peers and victimization by peers) and better school functioning (higher grades and fewer school discipline problems).

Importantly, these positive effects of EC were specific to the common EC, and did not extend to the specific aspect of EC related to activation control. Indeed, the Activation Control-Specific factor was *positively* correlated with some aspects of NE temperament and lower surgency, as well as higher levels of harm avoidance. Taken together, these relations suggest that individuals higher in activation control may be risk-averse and potentially experience over-control and fear of failure. These findings are novel, given that EC has never been decomposed into common and specific factors before. However, they are compatible with evidence that high

levels of conscientiousness can be associated with more negative emotion following achievement failures (Boyce et al., 2010), higher levels of guilt and shame (Rothbart, Ahadi, & Hershey, 1994), perfectionism (e.g., Stoeber, Otto, & Dalbert, 2009) and less risk taking (e.g., Carver, 2005; Gullone & Moore, 2000). In addition, worry is associated with motivation to undertake anticipatory preparation and planning (e.g., Watkins, 2008), and thus may lead to completing tasks on time. Investigating the potential costs, as well as benefits, of specific aspects of EC is thus an important area for future research.

Negative Emotionality. As expected, the NE temperament dimension was associated with psychopathology symptoms. Importantly, the common and specific NE factors differentially predicted different psychopathology symptoms. Common NE was strongly associated with both higher levels of depression and anxiety symptoms (common anxiety and physical symptoms), consistent with theories and evidence that depression and anxiety share broad negative emotionality as a common component (e.g., Anderson & Hope, 2008; Khan, Kristen, Gardner, Prescott, & Kendler, 2005; Ormel et al., 2013; Tellegen et al., 1999). The Depressed mood-specific and Fear-specific temperament factors showed good specificity, with the Fear-specific factor specifically predicting anxiety symptoms (and indeed being isomorphic with the separation/panic factor of the MASC), and the Depressed mood-specific factor predicting depression symptoms, as well as physical symptoms (which occur in depression as well as anxiety, e.g., fatigue and restlessness/agitation are symptoms of both major depression and generalized anxiety disorder, American Psychiatric Association, 2013). In addition, both Common NE and the Aggression-specific temperament factor predicted interpersonal functioning (more antisocial behavior towards peers and victimization from peers), while the Aggression-specific factor further predicted school functioning (more school discipline problems

and lower grades). Critically, these specific links would not be apparent if NE was analyzed as a single factor, demonstrating the importance of examining links between both common and specific aspects of NE with outcome variables.

Positive Emotionality. Correlations between PE factors and other measures yielded both expected and unexpected relations. As expected (Ellis & Rothbart, 2001), the Fear-Specific NE factor and MASC separation/panic factors were negatively correlated with the Surgency factor (i.e., most Surgency items refer to lack of fear of high-risk activities). Also as might be expected, Common PE was negatively correlated with Fear-Specific and Depressed Mood-Specific temperament factors, but Common NE and Common PE were *positively* correlated. Additionally, other correlations with PE factors are inconsistent with the view that they represent aspects of PE as usually defined. For example, the Affiliation-specific factor was associated with higher levels of depression and general anxiety, and worse interpersonal functioning. These findings are consistent with evidence that a high need for affiliation and corumination with close others are associated with psychopathology risk in adolescents (e.g., Cyranowski, Frank, Young, & Shear, 2000; Hankin, Stone, & Wright, 2010), but do not suggest that affiliation is necessarily related to PE.

We posit that the Common PE factor may be tapping a latent core construct that is related to sensitivity to the environment, and this in turn is related to greater general emotional reactivity, both towards negative and positive stimuli (Aron et al., 2012; but see Evans & Rothbart, 2008). Thus, if PE, as measured by the EATQ-R, is really assessing sensitivity as we speculate, then both positive and negative correlations between aspects of sensitivity NE might be expected, and could potentially explain the seemingly contradictory, and originally unexpected, correlations in the current study. For example, it is possible that given a generally

positive environment, sensitivity is associated with greater reward sensitivity (Aron et al., 2012). Reward sensitivity may protect against depressed mood (captured by the Depressed Mood-Specific factor; e.g., Forbes & Dahl, 2005) and potentially worry (captured by the Fear-Specific factor; Fairchild, 2011), but at the same time, may be associated with greater irritability, frustration and anger (captured by the Common NE factor) when rewards are not readily attained and instead reward seeking is thwarted (e.g., Carver, 2004; Hundt et al., 2013). However, we acknowledge that these propositions remain speculative, and future research is needed to evaluate the current results and expand the inquiry with other measures of sensitivity to investigate the construct validity of these new proposals on the underlying facets of PE, as currently measured in the EATQ-R.

Conclusions

In the 13 years since it was developed, the EATQ-R has been widely used, and has proved valuable in predicting many aspects of adolescent functioning and mental health. However, lack of an established factor structure has led to widespread inconsistency in the way it has been used (which subscales included and how they are combined) from study to study, which has made comparing results difficult and impeded progress in understanding adolescent temperament and how it is related to psychopathology and functioning. To address this problem, the current study developed and replicated latent variable models of temperament using confirmatory factor analysis in a large combined sample of adolescents. We identified best-fitting models for EC and NE that are readily interpretable, consistent with the broader trait literature on temperament and personality, and provide a more nuanced understanding of the structure of these temperament dimensions and their relation to one another and important adolescent outcomes. These models revealed specific, theoretically predicted, and meaningful

patterns of links with these outcome measures which would have been obscured if these temperament dimensions had been analyzed as in previous studies, either as unidimensional super-scales or individual subscales. Specifically, the common component of effortful control was broadly related to positive outcomes, including lower levels of psychopathology symptoms (depression, anxiety and ADHD), and better interpersonal and school functioning, while dimensions of negative emotionality temperament were related to psychopathology both through common risk from general NE and specific risk relating specific aspects of NE temperament to related specific aspects of psychopathology. These models can easily be applied both to future research and to gain new insights through re-analysis of existing EATQ-R data. The nature of the PE construct in the EATQ-R is less clear, and future research may benefit from inclusion of additional scales or items more clearly assessing PE as usually defined. Bringing consistency to the way the EATQ-R is modeled across studies and using purer latent variables has the potential to advance the field in understanding links between dimensions of temperament and important outcomes of adolescent development.

References

Al-Hendawi, M. (2013). Temperament, school adjustment, and academic achievement: Existing research and future directions. *Educational Review, 65,* 177-205. doi: 10.1080/00131911.2011.64837

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders, fifth edition. Arlington, VA: American Psychiatric Association.

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

Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory processing sensitivity: A review in the light of the evolution of biological responsivity. *Personality and Social Psychology Review*, *16*, 262–282. doi:10.1177/1088868311434213

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*, 527–530. doi:10.1016/j.paid.2010.11.015

Barrocas, A. L., Hankin, B. L., Young, J. F., & Abela, J. R. Z. (2012). Rates of nonsuicidal self-injury in youth: Age, sex, and behavioral methods in a community sample. *Pediatrics*, *130*, 39–45. doi:10.1542/peds.2011-2094

Boyce, C. J., Wood, A. M., & Brown, G. D. A. (2010). The dark side of conscientiousness: Conscientious people experience greater drops in life satisfaction following unemployment. *Journal of Research in Personality*, *44*, 535–539. doi:10.1016/j.jrp.2010.05.001

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development.* Oxford:

Wiley-Interscience.

Bussing, R., Fernandez, M., Harwood, M., Wei Hou, Garvan, C. W., Eyberg, S. M., & Swanson, J. M. (2008). Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: Psychometric properties and normative ratings from a school district sample. *Assessment*, *15*, 317–328. doi:10.1177/1073191107313888

Capaldi, D. M., & Rothbart, M. K. (1992). Development and validation of an early adolescent temperament measure. *The Journal of Early Adolescence*, *12*, 153–173. doi:10.1177/0272431692012002002

Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion*, *4*, 3–22. doi:10.1037/1528-3542.4.1.3

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*, 312–333. doi:10.1207/s15327957pspr0904_2

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

Checa, P., & Rueda, M. R. (2011). Behavioral and brain measures of executive attention and school competence in late childhood. *Developmental Neuropsychology*, *36*, 1018–1032. doi:10.1080/87565641.2011.591857

Checa, P., Rodríguez-Bailón, R., & Rueda, M. R. (2008). Neurocognitive and temperamental systems of self-regulation and early adolescents' social and academic outcomes. *Mind, Brain, and Education*, *2*, 177–187. doi:10.1111/j.1751-228X.2008.00052.x

Chen, F. F. (2006). A comparison of bifactor and second-order models of quality of life,

Multivariate Behavioral Research, 41, 189–225. doi:10.1207/s15327906mbr4102_5

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*, 219–251. doi:10.1111/j.1467-6494.2011.00739.x

Chess, S., & Thomas, A. (1977). Temperamental individuality from childhood to adolescence. *Journal of the American Academy of Child Psychiatry, 16*, 218–226. doi:10.1016/s0002-7138(09)60038-8

Coplan, R. J., & Bullock, A. (2012). Temperament and peer relationships. In M. Zentner & R. L. Shiner (Eds.) *Handbook of Temperament* (pp. 442-461). New York: Guilford Press,

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64,* 21–50. doi:10.1207/s15327752jpa6401_2

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245–265. doi:10.1348/0144665031752934

Cyranowski, J. M., Frank, E., Young, E., & Shear, M. K. (2000). Adolescent onset of the gender difference in lifetime rates of major depression: A theoretical model. *Archives of General Psychiatry, 57*, 21-27. doi:10.1001/archpsyc.57.1.21

de Boo, G. M., & Kolk, A. M. (2007). Ethnic and gender differences in temperament, and the relationship between temperament and depressive and aggressive mood. *Personality and Individual Differences, 43*, 1756–1766. doi:10.1016/j.paid.2007.05.012

De Los Reyes, A., & Prinstein, M. J. (2004). Applying depression-distortion hypotheses to the

assessment of peer victimization in adolescents. *Journal of Clinical Child & Adolescent Psychology, 33,* 325–335. doi:10.1207/s15374424jccp3302_14

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Developmental Psychopathology, 9,* 633–652.

Eisenberg, N., Vaughan, J., & Hofer, C. (2009). Temperament, self-regulation and peer social competence. In *Handbook of Peer Interactions, Relationships, and Groups* (pp. 473–489). Guilford Press.

Eisenberger, N. I., Lieberman, M. D., & Satpute, A. B. (2005). Personality from a controlled processing perspective: An fMRI study of neuroticism, extraversion, and self-consciousness. *Cognitive Affective & Behavioral Neuroscience, 5,* 169–181. doi:10.3758/CABN.5.2.169

Ellis, L. K. (2001). *Individual differences and adolescent psychosocial development*. Unpublished doctoral dissertation.

Ellis, L. K., & Rothbart, M. K. (2001). Revision of the early adolescent temperament questionnaire. *Poster Presented at the 2001 Biennial Meeting of the Society for Research in Child Development, Minneapolis, MN*.

Ellis, L. K., Rothbart, M. K., & Posner, M. I. (2004). Individual differences in executive attention predict self-regulation and adolescent psychosocial behaviors. *Annals of the New York Academy of Sciences, 1021,* 337–340. doi:10.1196/annals.1308.041

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

Evans, D. E., & Rothbart, M. K. (2008). Temperamental sensitivity: Two constructs or one? *Personality and Individual Differences, 44,* 108–118. doi:10.1016/j.paid.2007.07.016

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use

of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272

Fairchild, G. (2011). The developmental psychopathology of motivation in adolescence. *Developmental Cognitive Neuroscience*, *1*, 414–429. doi:10.1016/j.dcn.2011.07.009

Forbes, E. E., & Dahl, R. E. (2005). Neural systems of positive affect: Relevance to understanding child and adolescent depression? *Development and Psychopathology*, *17*, 827–850. doi:10.1017/S095457940505039X

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. doi:10.1037/0096-3445.137.2.201

García, L. F., Aluja, A., & Del Barrio, V. (2008). Testing the hierarchical structure of the Children's Depression Inventory: A multigroup analysis. *Assessment*, *15*, 153–164. doi:10.1177/1073191107310310

Griffith, J. W., Zinbarg, R. E., Craske, M. G., Mineka, S., Rose, R. D., Waters, A. M., & Sutton, J. M. (2010). Neuroticism as a common dimension in the internalizing disorders. *Psychological Medicine*, *40*, 1125–1136. doi:10.1017/S0033291709991449

Gullone, E., & Moore, S. (2000). Adolescent risk-taking and the five-factor model of personality. *Journal of Adolescence*, *23*, 393–407. doi:10.1006/jado.2000.0327

Hankin, B. L., Jenness, J., Abela, J. R. Z., & Smolen, A. (2011). Interaction of 5-HTTLPR and idiographic stressors predicts prospective depressive symptoms specifically among youth in a multiwave design. *Journal of Clinical Child & Adolescent Psychology*, *40*, 572–585. doi:10.1080/15374416.2011.581613

Hankin, B. L., Stone, L., & Wright, P. A. (2010). Corumination, interpersonal stress generation,

and internalizing symptoms: accumulating effects and transactional influences in a multiwave study of adolescents. *Development and Psychopathology*, *22*, 217–235. doi:10.1017/S0954579409990368

Hartman, C. A. (2000). Dutch translation of the Early Adolescent Temperament Questionnaire. Unpublished manuscript, University Medical Center Groningen, Department of Psychiatry.

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

Huisman, M., Oldehinkel, A. J., de Winter, A., Minderaa, R. B., de Bildt, A., . . . Ormel, J. (2008). Cohort profile: The Dutch "TRacking Adolescents" Individual Lives' Survey'; TRAILS. *International Journal of Epidemiology*, *37*, 1227–1235. doi:10.1093/ije/dym273

Hundt, N. E., Brown, L. H., Kimbrel, N. A., Walsh, M. A., Nelson-Gray, R., & Kwapil, T. R. (2013). Reinforcement sensitivity theory predicts positive and negative affect in daily life. *Personality and Individual Differences*, *54*, 350–354. doi:10.1016/j.paid.2012.09.021

Khan, A. A., Jacobson, K. C., Gardner, C. O., Prescott, C. A., & Kendler, K. S. (2005). Personality and comorbidity of common psychiatric disorders. *British Journal of Psychiatry*, *186*, 190–196. doi:10.1192/bjp.186.3.190

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

Kline, R. (2010). *Principles and Practice of Structural Equation Modeling, Third Edition*. Guilford Press.

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

21, 995–998.

Kovacs, M. (1992). Children's Depression Inventory (CDI) manual. Toronto (ON): Multi-Health Systems.

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, *64*, 241–256. doi:10.1037/a0015309

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, *121*, 971–977. doi:10.1037/a0028355

Laurent, J., Catanzaro, S. J., Joiner, T. E., Rudolph, K. D., Potter, K. I., . . . Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment, 11*, 326–338. doi:10.1037/1040-3590.11.3.326

Lengua, L. J., & Long, A. C. (2002). The role of emotionality and self-regulation in the appraisal–coping process: tests of direct and moderating effects. *Journal of Applied Developmental Psychology*, *23*, 471–493. doi:10.1016/S0193-3973(02)00129-6

Lonigan, C. J., & Vasey, M. W. (2008). Negative affectivity, effortful control, and attention to threat-relevant stimuli. *Journal of Abnormal Child Psychology*, *37*, 387–399. doi:10.1007/s10802-008-9284-y

Lonigan, C. J., Hooe, E. S., David, C. F., & Kistner, J. A. (1999). Positive and negative affectivity in children: Confirmatory factor analysis of a two-factor model and its relation to symptoms of anxiety and depression. *Journal of Consulting and Clinical Psychology*, *67*, 374–386.

Lonigan, C. J., Vasey, M. W., Phillips, B. M., & Hazen, R. A. (2004). Temperament, anxiety, and the processing of threat-relevant stimuli. *Journal of Clinical Child & Adolescent*

Psychology, *33*, 8–20. doi:10.1207/S15374424JCCP3301_2

Loukas, A., & Murphy, J. L. (2007). Middle school student perceptions of school climate: Examining protective functions on subsequent adjustment problems. *Journal of School Psychology*, *45*, 293–309. doi:10.1016/j.jsp.2006.10.001

Magar, E., Phillips, L. H., & Hosie, J. A. (2008). Self-regulation and risk-taking. *Personality and Individual Differences*, *45*, 153–159. doi: 10.1016/j.paid.2008.03.014

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 and Adolescent Psychiatry*, *36*, 554–565. doi:10.1097/00004583-199704000-00019

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, *88*, 139–157. doi:10.1037/0022-3514.88.1.139

Mezulis, A. H., & Rudolph, M. E. (2012). Pathways linking temperament and depressive symptoms: A short-term prospective diary study among adolescents. *Cognition & Emotion*, *26*, 950–960. doi:10.1080/02699931.2012.665027

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*, 1460–1470. doi:10.1080/02699931.2010.543642

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. doi:10.1177/0963721411429458

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling (pp.

488–508). Thousand Oaks, CA: SAGE.

Muris, P., Merckelbach, H., Ollendick, T., King, N., & Bogie, N. (2002). Three traditional and three new childhood anxiety questionnaires: Their reliability and validity in a normal adolescent sample. *Behaviour Research and Therapy*, *40*, 753–772.

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*, 7–19. doi:10.1007/s10862-008-9089-x

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*, 1035–1049. doi:10.1016/j.adolescence.2007.03.003

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

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7 ed.). Los Angeles, CA: Muthén & Muthén.

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

Noordhof, A., Krueger, R. F., Ormel, J., Oldehinkel, A. J., & Hartman, C. A. (2014). Integrating autism-related symptoms into the dimensional internalizing and externalizing model of psychopathology. The TRAILS Study. *Journal of Abnormal Child Psychology,* Advance online publication. doi:10.1007/s10802-014-9923-4

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*, 421–440. doi:10.10170S0954579404044591

Oldehinkel, A. J., Hartman, C. A., Ferdinand, R. F., Verhulst, F. C., & Ormel, J. (2007). Effortful control as modifier of the association between negative emotionality and adolescents' mental health problems. *Development and Psychopathology*, *19*, 523–539. doi:10.1017/S0954579407070253

Oldehinkel, A. J., Hartman, C. A., Nederhof, E., Riese, H., & Ormel, J. (2011). Effortful control as predictor of adolescents“ psychological and physiological responses to a social stress test: The Tracking Adolescents” Individual Lives Survey. *Development and Psychopathology*, *23*, 679–688. doi:10.1017/S0954579411000095

Ormel, J., Jeronimus, B. F., Kotov, R., Riese, H., Bos, E. H.,. . . Oldehinkel, A. J. (2013). Neuroticism and common mental disorders: Meaning and utility of a complex relationship. *Clinical Psychology Review*, *33*, 686–697. doi:10.1016/j.cpr.2013.04.003

Ormel, J., Oldehinkel, A. J., Sijtsema, J., van Oort, F., Raven, D., Veenstra, R., et al. (2012). The TRacking Adolescents' Individual Lives Survey (TRAILS): design, current status, and selected findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 1020–1036. doi:10.1016/j.jaac.2012.08.004

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspectives on Psychological Science, 7*, 528–530. doi:10.1177/1745691612465253

Pillow, D. R., Pelham, W. E., Hoza, B., Molina, B. S., & Stultz, C. H. (1998). Confirmatory

factor analyses examining attention deficit hyperactivity disorder symptoms and other childhood disruptive behaviors. *Journal of Abnormal Child Psychology, 26,* 293–309.

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 Psychology, 30,* 479–491. doi:10.1207/S15374424JCCP3004_05

Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In *Advances in Research on Temperament* (pp. 165–182). Lengerich, Germany, Pabst Science.

Roose, A., Bijttebier, P., Decoene, S., Claes, L., & Frick, P. J. (2010). Assessing the affective features of psychopathy in adolescence: A further validation of the Inventory of Callous and Unemotional Traits. *Assessment, 17,* 44–57. doi:10.1177/1073191109344153

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52,* 569–578. doi:10.2307/1129176

Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16,* 207–212.

Rothbart, M. K., & Rueda, M. R. (2005). The development of effortful control. *Developing Individuality in the Human Brain: A Tribute to Michael I. Posner*, 167–188.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40,* 21-39. doi:10.2307/23087906

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72,* 1394–1408.

Rothbart, M. K., Ellis, L. K., Rueda, M. R., & Posner, M. I. (2003). Developing mechanisms of

temperamental effortful control. *Journal of Personality, 71,* 1113–1143.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28,* 573–594. doi:10.1207/s15326942dn2802_2

Stoeber, J., Otto, K., & Dalbert, C. (2009). Perfectionism and the Big Five: Conscientiousness predicts longitudinal increases in self-oriented perfectionism. *Personality and Individual Differences, 47,* 363–368. doi:10.1016/j.paid.2009.04.004

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., . . . 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 Academy of Child and Adolescent Psychiatry, 40,* 168–179. doi:10.1097/00004583-200102000-00011

Swanson, J., Valiente, C., & Lemery-Chalfant, K. (2012). Predicting academic achievement from cumulative home risk: The mediating roles of effortful control, academic relationships, and school avoidance. *Merrill-Palmer Quarterly, 58,* 375–408. doi:10.1353/mpq.2012.0014

Tackett, J. L., Lahey, B. B., van Hulle, C., Waldman, I., Krueger, R. F., & Rathouz, P. J. (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology, 122,* 1142–1153. doi:10.1037/a0034151

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10,* 297-303. doi:10.1111/1467-9280.00157

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,* 946-972. doi:10.1177/0272431613477239

van Prooijen, J. W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61,* 777–792. doi:10.1177/00131640121971518

Vasey, M. W., Harbaugh, C. N., Lonigan, C. J., Phillips, B. M., Hankin, B. L., Willem, L., & Bijttebier, P. (2013). Dimensions of temperament and depressive symptoms: Replicating a three-way interaction. *Journal of Research in Personality, 47,* 908–921. doi:10.1016/j.jrp.2013.09.001

Verstraeten, K., Vasey, M. W., Claes, L., & Bijttebier, P. (2010). The assessment of effortful control in childhood: Questionnaires and the Test of Everyday Attention for Children compared. *Personality and Individual Differences, 48,* 59–65. doi:10.1016/j.paid.2009.08.016

Verstraeten, K., Vasey, M. W., Raes, F., & Bijttebier, P. (2008). Temperament and risk for depressive symptoms in adolescence: Mediation by rumination and moderation by effortful control, *37*(3), 349–361. doi:10.1007/s10802-008-9293-x

Verstraeten, K., Vasey, M. W., Raes, F., & Bijttebier, P. (2010). Brooding and reflection as components of rumination in late childhood. *Personality and Individual Differences, 48,* 367–372. doi:10.1016/j.paid.2009.11.001

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134,* 163–206. doi:10.1037/0033-2909.134.2.163

Willem, L., Bijttebier, P., Claes, L., Vanhalst, J., & Raes, F. (2013). The cross-temporal associations between rumination subtypes and substance use in adolescence: Exploring the

moderating role of gender. *Journal of Psychopathology and Behavioral Assessment*, *36*, 143–154. doi:10.1007/s10862-013-9373-2

Wills, T. A., Sandy, J. M., & Shinar, O. (1999). Cloninger's constructs related to substance use level and problems in late adolescence: A mediational model based on self-control and coping motives. *Experimental and Clinical Psychopharmacology*, *7*, 122–134. doi:10.1037//1064-1297.7.2.122

Yap, M. B. H., Allen, N. B., O'Shea, M., di Parsia, P., Simmons, J. G., & Sheeber, L. (2011). Early adolescents' temperament, emotion regulation during mother–child interactions, and depressive symptomatology. *Development and Psychopathology*, *23*, 267–282. doi:10.1017/S0954579410000

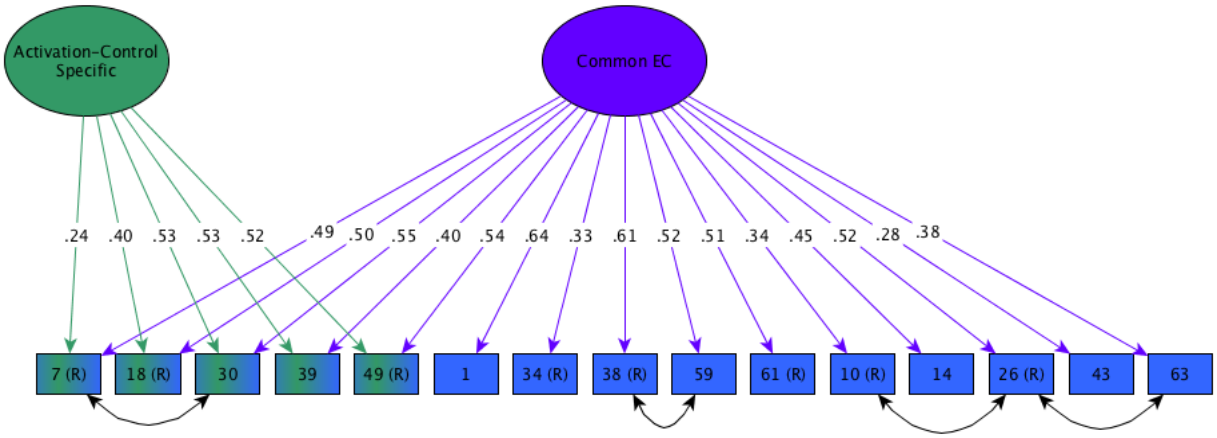

Figure 1. Bifactor model for Effortful Control. The Common EC factor captures what is shared across all EC items, while the Activation Control-Specific factor captures what is unique to the Activation Control items. Numbers in boxes are EATQ-R item numbers; (R) indicates that the item is reverse coded. Numbers on straight arrows are factor loadings. Curved arrows between item boxes indicate model modifications allowing the residual variance of the items to correlate.

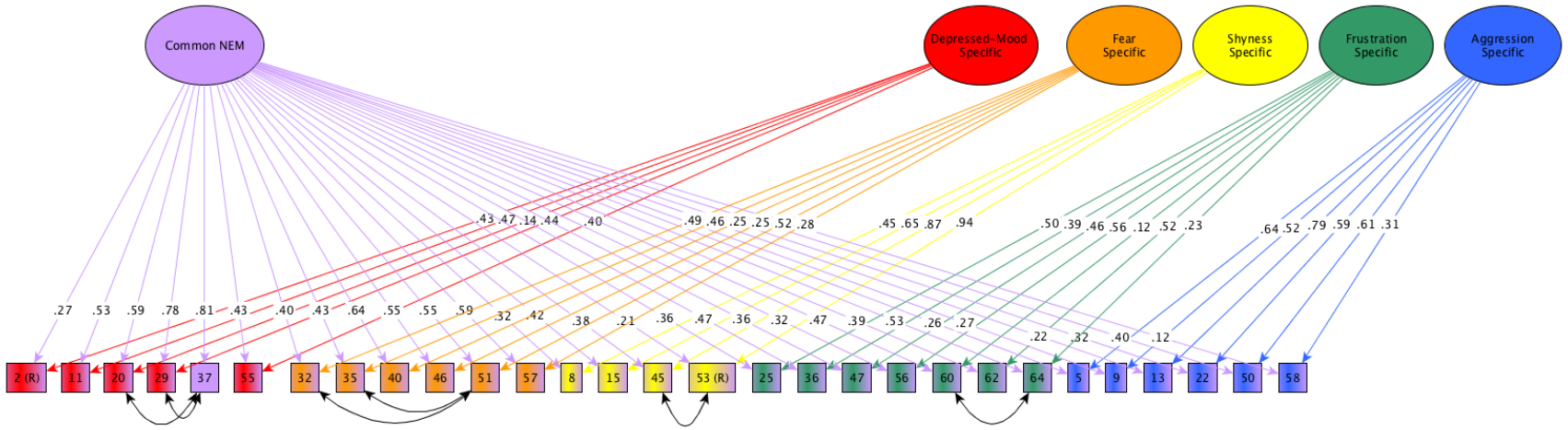

Figure 2. Bifactor model for Negative Emotionality. The Common NE factor captures what is shared across all NE items, while the each specific factor captures what is unique to the items in that subscale. Numbers in boxes are EATQ-R item numbers; (R) indicates that the item is reverse coded. Numbers on straight arrows are factor loadings. Curved arrows between item boxes indicate model modifications allowing the residual variance of the items to correlate.

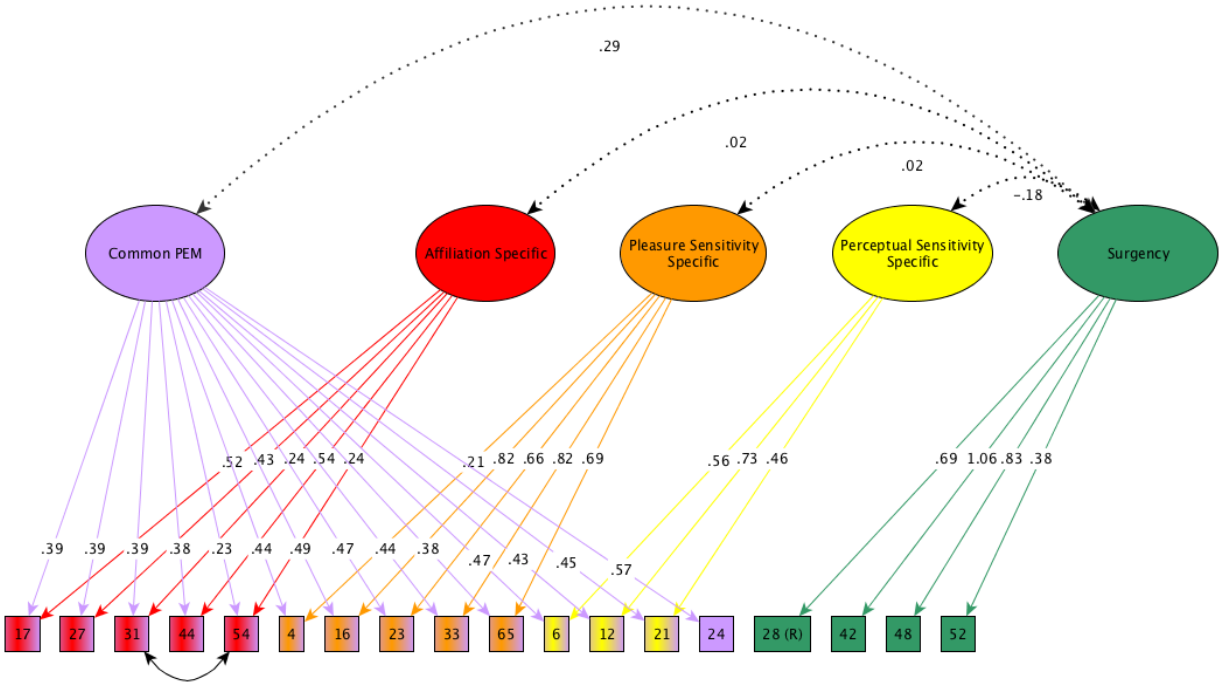

Figure 3. Nested bifactor model for Positive Emotionality/Surgency. The Common PE factor captures what is shared across all PE items, while the each specific factor captures what is unique to the items in that subscale. Surgency items did not load adequately on Common PE, and are therefore modeled as a separate factor. Numbers in boxes are EATQ-R item numbers; (R) indicates that the item is reverse coded. Numbers on straight arrows are factor loadings. Numbers on curved arrows between factors are latent correlations. Dashed lines indicate weak correlations <.30, and solid lines represent stronger correlations. Curved arrows between item boxes indicate model modifications allowing the residual variance of the items to correlate.

Table 1

EATQ-R Model Fit Statistics

Scale	Model	χ^2 (df)	χ^2 /df	CFI	RMSEA
Effortful Control (EC)	3 Factor Correlated	393.19 (83)**	4.74	.88	.061
	1 Factor	552.23 (86)**	6.42	.81	.073
	Common EC + Activation-Control Specific	391.06 (81)**	4.83	.88	.061
Negative Emotionality (NE)	5 Factor Correlated	1082.92 (359)**	3.02	.89	.045
	1 Factor	2283.75 (369)*	6.19	.72	.072
	Common NE + Subscale-Specific	974.16 (343)**	2.84	.91	.043
Positive Emotionality (PE)	4 Factor Correlated	478.60 (128)**	3.74	.89	.052
	1 Factor	1664.21 (134)**	12.42	.52	.104
	Common PE + Affiliation, Pleasure Sensitivity and Perceptual Sensitivity -Specific + Surgency	358.15 (116)**	3.09	.92	.045
Full Model	Combined EC, NE and PE models (no Affiliation-specific, see Results)	4812.26 (1731)**	2.78	.80	.042

** $p < .001$

Table 2

EATQ-R Full Model Estimated Factor Correlations

Factor Correlations	Common EC		Activation-Specific		Surgency		Common PE		Pleasure sensitivity-specific		Perceptual sensitivity-specific	
Dataset	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
Common NE	**-.48***	**-.36***	**.24***	**.30***	-.16	**-.25***	**.55***	**.57***	.09	.09	.07	.10
Aggression-specific	**-.42***	**-.43***	.09	-.17	.08	**.17***	-.16	-.15	-.08	**-.30***	.05	.16
Depressed mood specific	**-.33***	**-.40***	-.11	.09	-.13	**-.22***	**-.63***	**-.56***	.06	.03	**.34***	**.31***
Fear-specific	-.10	-.17	**.36***	.29	**-.61***	**-.48***	**-.30***	-.29	**.40***	.19	.04	.03
Frustration-specific	**-.23***	**-.41***	-.23	-.13	.07	.17	.10	.03	-.07	-.20	.12	.11
Shyness-specific	.00	-.10	-.04	-.08	**-.25***	**-.21***	-.15	**-.21***	.01	.07	-.04	.07
Common EC					.13	.11	.10	.14	.01	.09	-.14	**-.20***
Activation-specific					**-.28***	**-.26***	-.04	-.13	**.27***	**.35***	.15	.17
Common PE	.10	.14	-.04	-.13	**.23***	.14						
Pleasure sensitivity-specific	.01	.09	**.27***	**.35***	-.13	-.11						
Perceptual sensitivity-specific	-.14	**-.20***	.15	.17	.09	.08						

Note. Dataset 1= model development dataset, Dataset 2= Hold-out dataset. *n*=1013 for each dataset. Grey boxes in factor correlations table indicate self-correlations or factors constrained not to correlate (e.g., Specific factors within each scale do not correlate with each other or their Common factor, because their shared variance is already captured by their Common factor).

* $p < .0005$ (Bonferroni corrected *p* value for family-wise error rate of .05)

Table 3

Final Bifactor Model Replication in Hold-Out Sample:

Model Fit Statistics and Differences from Model Development Sample Models

Model	χ^2 (df)	χ^2 /df	CFI	RMSEA	Loading Diff.	Abs. Loading Diff.
Effortful Control	414.10 (81)**	5.11	.84	.066	-0.03	0.06
Negative Emotionality	1064.67 (343)**	3.10	.89	.047	-0.01	0.08
Positive Emotionality	380.32 (116)**	3.28	.91	.049	0.00	0.09
Full Model	4693.71 (1731)**	2.71	.79	.043	-0.01	0.07

Note. Loading Diff. = Average item factor loadings for hold-out dataset – model development dataset. Abs. Loading Diff. = Absolute value of the Loading Diff.

** $p < .001$

Table 4
Estimated factor correlations between EATQ-R model factors and measures of adolescent functioning

		Effortful Control		Negative Emotionality						Positive Emotionality				
		C. EC	Act.-spec.	C. NE	Agg.-spec.	Dep.- spec.	Fear-spec.	Frust.-spec.	Shy-spec.	C. PE	Surgency	Affil.-spec.	Pleasure sens.-spec.	Percep. sens.-spec.
Depression	CDI	**-.58***	.06	**.57***	.12	**.50***	-.14	-.07	-.01	-.25	-.15	**.31***	.00	**.23***
Anxiety	C. MASC	**-.38***	-.17	**.75***	-.09	.00	.05	-.18	.16	-.11	-.14	**.50***	.10	.17
	MASC Physical Symptom-Spec.	**-.28***	.04	**.17***	.14	**.39***	.16	.09	-.08	.06	.07	.10	-.02	**.28***
	MASC Separation/ Panic-Spec.	.07	-.03	.02[a]	-.01[a]	.12[a]	**1.0**[a]	.13[a]	.05[a]	.03	**-.35***	**-.31***	.09	-.16
	MASC Harm Avoidance-Spec.	**.38***	**.39***	-.06	-.20	-.26	**.42***	-.07	.03	**.32***	.10	-.19	.06	.08
ADHD	C. SNAP	**-.25***	.23	-.08	.16	.21	-.01	.19	-.12	-.04	.12	-.09	-.09	.13
	SNAP Inattention-Spec.	-.21	-.07	-.02	.05	.19	.10	.14	.08	-.12	-.02	.07	.10	.05
	SNAP Hyperactivity -Spec.	.15	.05	-.10	-.12	.14	.08	.07	.10	-.12	.01	-.04	.15	.00
	SNAP Impulsivity-Spec.	.14	-.11	-.02	.03	-.03	.19	.10	-.05	-.06	-.06	-.01	-.07	.07
Interpersonal Funct.	RPEG Antisocial	**-.45***	-.04	**.36***	**.46***	.02	-.13	-.16	-.09	-.12	-.01	**.43***	-.10	.18
	RPEG Victim	**-.35***	.07	**.34***	**.27***	.19	-.05	-.12	-.09	-.04	.05	**.30***	-.02	**.32***
School Funct.	Grades	**.36***	-.06	-.11	**-.35***	-.10	**-.25***	-.04	-.01	.07	-.05	.09	.16	-.03
	School Discipline Problems [b]	**-.18***	-.03	-.11	**.26***	.20	.04	.20	-.10	-.08	.11	.04	-.21	.12

Note. *n* = 562. Spec. = Specific C= Common, Act. = Activation, Agg. = aggression, Dep. = Depressed Mood, Frust. = Frustration, Affil. = Affiliation Pleasure = Pleasure Sens. Sensitivity, Percep. Sens. = Perceptual Sensitivity, Func. = Functioning

[a] The EATQ-R Fear-specific factor is perfectly correlated with the MASC Separation/Panic-Specific factor, causing the covariance matrix to be not positive definite when the MASC Separation/Panic-Specific factor is included in the model. The model was therefore re-run without this factor, and correlations between EATQ NE factors and other MASC factors are reported for the model without a MASC Separation/Panic-Specific factor. Correlations were very similar with and without this factor included.

[b] Because school discipline problems were positively skewed, the MLR estimator in mplus, which is robust to non-normality, was used. Results were very similar using the standard FIML estimator.

* $p < .0003$ (Bonferroni corrected *p* value for family-wise error rate of .05).