Stony Brook University

The Structural and Relative Stability of Temperament in Young Children Based on a Laboratory-Observational Measure

A Dissertation Presented

by

Margaret Wegner Dyson

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Clinical Psychology

Stony Brook University

August 2013

Stony Brook University

The Graduate School

Margaret Wegner Dyson

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Daniel N. Klein, Ph.D. – Dissertation Advisor
Psychology Professor, Clinical Psychology

Joanne Davila, Ph.D - Chairperson of Defense
Psychology Professor, Clinical Psychology

Anne Moyer, Ph.D
Psychology Professor, Psychology

Sheldon Weintraub, Ph.D.
Professor, Stony Brook University Medical Center-Psychiatry

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

The Structural and Relative Stability of Temperament in Young Children Based on a Laboratory-Observational Measure

by

Margaret Dyson

Doctor of Philosophy

in

Clinical Psychology

Stony Brook University

2013

Despite the continued debate regarding the structure of temperament in young children, it is often assumed that temperament traits demonstrate temporal stability over time. This research has relied predominately on parent-report measures. The present study used an alternative approach, a laboratory-observational measure (Laboratory Temperament Assessment Battery [Lab-TAB]), to examine the structural and relative stability of temperament traits in a community sample of young children (N = 440). Using structural equation modeling (SEM), we derived a similar five-factor structure consisting of the dimensions of Positive Affect/Interest, Sociability, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint at both age 3 and 6 years, suggesting good structural stability. In addition, all five latent factors exhibited significant relative stability from age 3 to 6, as well as two significant temporal associations between different latent factors. This represents one of the first studies to use SEM procedures to examine the structural and relative stability in young children based on a laboratory-observational measure.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my amazing, brilliant, and supportive advisor, Dr. Daniel Klein. Dan, thank you for all your guidance, advice, and encouragement over the years.

I would also like to thank my dissertation committee members for their feedback and support throughout this process: Dr. Joanne Davila, Dr. Anne Moyer, and Dr. Sheldon Weintraub.

I must also acknowledge all of my wonderful colleagues and friends of the Klein lab who were essential to my success in graduate school: Lea Dougherty, Ph.D., Thomas Olino, Ph.D., Rebecca Laptook, Ph.D., Dana Torpey, Ph.D., Sara Bufferd, Ph.D., Jiyon Kim, Ph.D., Autumn Kujawa, M.A., Sarah Black, M.A., and Allison Pennock, M.A, Laura Klein, M.A., and Sue Rose, M.A.

I must also thank my two favorite "Rs," my best friend and fellow graduate student, Rachel Hershenberg, and my fiancé, Rich Grousset . Rach, I am so grateful that we were able to take this journey together. I could not have done it without your continued support, humor, cheerleading, and attempts at capitalization. Rich, I feel so incredibly lucky to have met you and am excited to begin our new life together. Thank you for always believing in me and comforting me during this process. Finally, I would like to thank my wonderful parents and sister. Mom, Dad, and Cate, thank you for your endless support and everything you have done for me over the years to enable me to accomplish my goals. I love you all!

The Structural and Relative Stability of Temperament in Young Children Based on a Laboratory-Observational Measure

Individual differences in reactivity and regulation in young children have traditionally been studied within a temperament framework. Several research traditions (e.g., Thomas and Chess, Buss and Plomin, Goldsmith and Campos, and Rothbart and colleagues) have developed models of the structure of temperament traits (see Zentner & Bates 2008 for detailed review of different theoretical models). There has also been a growing interest in the applicability of the Five Factor Model (FFM; Digman, 1990; McCrae & Costa, 1987, 1997) of adult personality to young children. Most of these theoretical traditions concur that the structure of temperament traits in young children is multidimensional; however, there is little agreement with regard to the number and nature of these primary trait dimensions (De Pauw, Mervielde, & Leeuwan, 2009; De Pauw & Mervielde, 2010). Despite the extensive debate regarding the structure and nature of traits, an implicit assumption of these models is that temperament traits demonstrate temporal stability (Caspi & Roberts, 2005; Goldsmith et al., 1987; Pedlow, Sanson, Prior, & Oberklaid, 1993).

Types of Stability in Young Children

Developmental researchers propose at least four types of stability in longitudinal research (e.g., Caspi & Shiner, 2005; De Fruyt, Bartels, Leeuwen, Clereq, Decuyper, and Mervielde, 2006; Putnam, Rothbart, & Garstein, 2008): (a) *ipsative*, the degree to which the relative ordering of traits within an individual are preserved across time (i.e., if an individual stays the same over time), (b) *mean-level*, the changes in the average trait level in the population, (c) *relative* or *rank order*, the consistency of individual differences within a sample of individuals and the degree to which the relative ordering of individuals on a trait is maintained over time,

and (d) *structural*, the degree of continuity in the inter-correlations among traits across time. The present study will focus on the latter two types of stability[1].

Factors That May Impact Stability in Young Children

This study will focus on examining the structural and relative stability of temperament in early childhood, rather than teasing apart the factors that influence stability; nonetheless, it is important to acknowledge what some of these factors may be. As discussed below, the method of assessing temperament (e.g., parent-report versus laboratory-observational) may be one such factor. Another important factor is measurement error, which may lead to inaccurate estimates of the stability of temperament traits; however, this can be addressed and controlled using structural equation modeling (SEM). Maturation or normative development has also been observed to impact the stability of temperament over time, especially in early childhood when changes are widespread and rapid (Caspi & Roberts, 2001; Rothbart & Bates, 2006). As discussed below, lower relative stability estimates have been obtained for traits assessed in the early stages of life (e.g., infancy, preschool period; Caspi, Roberts, & Shiner, 2005; Caspi & Shiner, 2006; Durbin et al., 2007; Rothbart & Bates, 2006). Maturation may also influence the structural stability of temperament, as there may be increasing differentiation of traits over the course of childhood (e.g., the number of traits increases from infancy to preschool; Caspi, Roberts, & Shiner 2005; Eisenberg, 2000). Thus, lower structural and relative stability may be expected in early childhood, compared to older ages, due to maturational or normative developmental changes.

[1] Unlike the three other types of stability, the ipsative approach focuses on change that occurs at the individual level and usually requires more than the two assessment time points available for this study. Also, we are precluded from using the mean-level approach because we used age-appropriate measures/tasks, which necessarily differed between assessment time points (to be discussed further in the methods section); hence differences in stability are confounded with differences in tasks.

Finally, genetic and environmental factors may influence the stability of temperament (Kandler et al., 2010; Krueger & Johnson, 2008). According to the cumulative continuity principle, genetic factors contribute to stability by influencing the set-points to which individuals will revert back (Caspi, Roberts, & Shiner, 2005). However, there is also evidence that environmental factors can sometimes alter these set-points and contribute to enduring changes in levels of traits (Fraley & Roberts, 2005; Kandler et al., 2010; Krueger & Johnson, 2008).

Methods of Assessing Temperament in Young Children

As noted previously, the method of measurement used to assess temperament in young children may impact relative and structural stability estimates (Durbin et al., 2007; Majdandžić & van den Boom, 2007). Parent-report questionnaires are the most common method of evaluating temperament in young children and have provided consistent support for the stability of childhood temperament (e.g., Pedlow et al., 1993; Lemery et al., 1999; Rothbart et al., 2000). In addition to being relatively inexpensive, convenient to administer, and time-efficient, parent-report measures tap a parent's extensive experience with a child's emotional and behavioral responses across a variety of settings and situations (Mangelsdorf, Schoppe, & Buur, 2000; Rothbart & Bates, 2006). However, parent-report measures of child temperament also have several limitations (Kagan, 1998; Kagan, & Fox, 2006; Mangelsdorf et al., 2000; Rothbart & Bates, 2006; Stifter, Willoughby, & Towe-Goodman, 2008). Parent-report measures are vulnerable to multiple response and perceptual biases, and thus represent a mixture of objective and subjective factors (Mangelsdorf et al., 2000; Rothbart & Bates, 2006; Zentner & Bates, 2008). For instance, parents may respond based on their own personality, emotional state, or psychopathology (Durbin et al., 2007; Jouriles & Thompson, 1993; Mangelsdorf et al., 2000; Stifner et al., 2008). Or, parents may be susceptible to social desirability and only portray their

child in a positive light. Moreover, stability estimates based on parent-report measures may reflect the stability of parent expectations or schemas of temperament rather than the stability of the child's behavior (Gagne, Hulle, Askan, Essex, & Goldsmith, 2011; Mangelsdorf et al., 2000; Saudino, 2003). Thus, stability estimates of temperament traits may be imprecise or inflated when derived from parent-report measures (Durbin et al., 2007, Mangelsdorf et al., 2000; Saudino, 2003).

Laboratory-observational measures of child temperament have a number of advantages relative to parent-report measures. These measures allow the researcher to utilize standardized procedures to elicit specific behaviors and emotions (Majdandžić & van den Boom; Zeman, Klimes-Dougan, Cassano, & Adrian, 2007). Also, observational measures use objective criteria to code behavior and emotion, circumventing the issue of parental interpretation (Durbin et al., 2007). Specifically, coders are trained to reliably detect specific emotions and behaviors and are less likely to be influenced by emotional state and psychopathology, implicit theories of the structure of temperament, the desire to be consistent in one's responses, beliefs about the consistency of child behavior, or social desirability (Stifner et al., 2008). Moreover, laboratory-observational measures provide the opportunity to place children in situations that evoke low frequency emotions and behaviors, such as fearfulness or inhibition (Durbin et al., 2007). Nonetheless, laboratory-observational measures also have notable limitations. In addition to being expensive, laborious, and time-intensive, laboratory-observational measures are also susceptible to "state" effects (i.e., transient influences that are specific to a particular time or environment) and may have questionable ecological validity (i.e., the laboratory represents a novel and atypical context; Goldsmith & Gagne, in press).

Studies have consistently reported low associations between parent-report and laboratory-observational measures of temperament (e.g., Durbin, Hayden, Klein, Olino, 2007; Goldsmith, Reiser-Danner, & Briggs, 1991; Majdandzic, & van den Boom, 2007; Saudino & Cherny, 2001; Stifter et al., 2008). However, high concordance between observational and parent-report measures should not be anticipated, as each has its own advantages and disadvantages (Rothbart & Goldsmith, 1985). Moreover, observational and parent-report measures may tap different aspects of temperament (Gagne et al., 2011). For instance, in a sample of toddlers, parent-reports' and laboratory assessments of activity level reflected different environmental and genetic influences (Saudino, 2009). Therefore, observational measures may provide different information about the structure of temperament than parent-report measures. As such, it may not be possible to generalize about the structural and relative stability of temperament from one method to the other.

Cross-Sectional Studies of Structural Stability in Young Children Using Parents' Reports and Exploratory Factor Analysis.

Few studies have investigated the structural stability of temperament in young children. Several cross-sectional studies have conducted exploratory factor analysis (EFA) on groups of different-aged children. Rothbart, Ahadi, Hershey, and Fisher (2001) conducted EFA on samples of 3, 4-5, and 6-7 year-old children to examine the structure of child temperament based on parent ratings on the Children's Behavior Questionnaire (CBQ). The three-factor solution of Negative Affectivity, Extraversion, and Effortful Control was found to be highly similar across all age groups. Using another parent-report measure, the Inventory of Child Individual Differences (ICID), Halverson and colleagues (2003) recovered the FFM traits (i.e., Extraversion, Neuroticism, Conscientiousness, Agreeableness, and Openness) in cross-sectional

samples of 3-5, 6-9, 11-13, 16-18, and 20-23 year-olds using exploratory factor analysis. Thus, although the two studies differed in the number of factors extracted, both suggest that the structure of temperament is similar across childhood, and possibly into adulthood.

Studies of Relative Stability in Young Children Using Parents'/Observers' Reports

Several longitudinal studies have investigated the relative stability of temperament traits in early childhood, with estimates generally falling in the moderate range. For example, in their comprehensive meta-analysis examining the rank-order stability of traits from infancy to late adulthood, Roberts and Delvecchio (2000) combined stability coefficients for temperament trait dimensions (i.e., approach, adaptability, task persistence, negative emotionality, activity level, rhythmicity, and threshold) and adult personality traits (i.e., Extraversion, Neuroticism, Conscientiousness, Agreeableness, Openness, Femininity/Masculinity, Type A). For the time interval from 3 to 5.9 years (the present study's period of interest), they estimated a stability correlation (i.e., cross-time correlation) of .52. Nonetheless, the majority of those studies investigating the stability of temperament traits have only utilized simple bivariate cross-time correlations to evaluate the stability of temperament traits, which are susceptible to attenuation due to measurement error (Roberts, Caspi, & Moffit, 2001).

Longitudinal Studies of Structural and Relative Stability in Young Children Using Parents' Reports and Confirmatory Factor Analysis/Structural Equation Modeling

The EFA structural studies discussed above have several limitations. First, in cross-sectional designs, different children are examined at each age, hence developmental and sample effects are confounded. Longitudinal designs, in which the same children are examined at several points over time, provide a more robust and sensitive approach to examining structural stability. Second, EFA has limitations in comparison to Confirmatory Factor Analysis (CFA). Unlike

EFA, CFA provides omnibus tests of model fit and permits formal comparisons between alternative models both within and across time. In addition, CFA is generally implemented in a structural equation modeling (SEM) context, which has the additional advantages of accounting for measurement error in estimating relative stability, and allowing for the examination of both structural and relative stability within the same analytic framework (Byrne, 2005; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hurley, Scandura, Schriesheim, Brannick, Seers, Vandenberg et al., 1997).

Only a few studies have used longitudinal designs and/or CFA/SEM to examine the structural and relative stability of temperament in young children. Based on maternal-reports of children ranging from infancy to age 8, the Australian Temperament Project used CFA to test the factor structures of temperament traits at five separate assessment periods (4-8, 18-22, 32-36, 44-52, 57-78, and 88-99 months). The authors identified two factors (Approach/Sociability, Rhythmicity) that emerged consistently from infancy to age 8, and four factors (Irritability, Inflexibility, Cooperation-Manageability, and Persistence) that emerged across most of the time intervals (Pedlow et al., 1993). Thus, these models exhibited fairly good structural stability over time, especially after infancy. The relative stability estimates based on the SEM analyses ranged from moderate to large for the six factors: Irritability (.60-.78), Inflexibility (.80-.81), Approach (.44-.82), Cooperation-Manageability (.53-.78), Persistence (.55-.78), and Rhythmicity (.57-.83) (Pedlow et al., 1993).

Also using maternal reports, Lemery, Goldsmith, Klinnert, and Mrazek (1999) examined the core temperament traits of Positive Emotionality, Activity Level, Fear, and Distress-Anger at 3, 6, 12, 18, 24, 36, and 48 months. Utilizing structural equation modeling, these temperament traits generally demonstrated moderate stability across time. Further, their findings suggest that

stability increases from infancy to the toddlerhood–preschool period, and that within the toddlerhood-preschool (2-4 years of age) period, there was a pattern of high stability (estimates in the .70s).

Stability of Temperament in Young Children Using Laboratory-Observational Measures

A number of studies have utilized laboratory-observational measures to examine the stability of single temperament traits, such as behavioral inhibition, which is believed to exhibit a moderate degree of stability from infancy to childhood (with higher stability at older ages) (e.g., Fox, Henderson, Rubin, Calkins, & Schmidt, 2001; Kagan, Snidman, Kahn, & Townsley, 2007; Pfiefer, Goldsmith, Davidson, & Rickman, 2002; Scarpa, Raine, Venables, & Mednick, 1995). However, only a handful studies have used laboratory-observational measures to examine the relative stability of multiple temperament traits within the same study. Most of these studies have focused on infants and toddlers and used short intervals between assessments, which should maximize stability.

Based on a sample of 9-month old twins, Goldsmith and Campos (1990) found that laboratory indices of pleasure (.55) and fear (.38) exhibited moderate to high stability over a two-week period. For the period between 12-13 and 18-20 months of age, Belsky and colleagues found high stability for laboratory-assessed positive emotionality (.79) and negative emotionality (.67) (Belsky, Hsieh, & Crnic, 1996). For the ages of 3, 6, 9 and 12 months, Carnicero, Perez-Lopez, Gonzalez-Salinas and Martinez-Fuentes (2000) estimated the stability of the laboratory-assessed traits of sociability, activity level, attentiveness, and emotional tone (positive vs. negative affect). Across the different time intervals, stability estimates ranged from low to high (.00-.52). However, during the interval between 9 to 12 months (the oldest age), moderate to high stability was found for all of the traits (.42-.52), with the exception of activity level (.14).

Finally, across the ages of 3, 6, 10, and 13.5 months, Rothbart, Derryberry, and Hershey (2000) found low to moderate relative stability for the traits of fear (.20-.34), frustration (.12-.37), distress (.05-.32), and smiling/laughter (.21-.35). In sum, these studies generally report relative stability estimates (i.e., cross-time correlations) in the moderate range for periods of up to a year.

Fewer studies have examined the relative stability of temperament traits based on a laboratory-observational measure in young children (i.e., preschool to early childhood). Across the ages of 3, 5, and 7 years, Durbin and colleagues (2007) found moderate to high stability for the laboratory-assessed traits of Positive Emotionality, which includes the lower-order traits of positive affect (PA) (.59-.70), anticipatory PA (.41), sociability (.52-.62), interest/engagement (.37-.48), and activity level (.62), and Negative Emotionality, which includes the lower-order traits of negative affect (.57-.59), sadness (.52), anger (.30-.40). The lower-order trait of fear yielded lower stability estimates (.21-.23).

Finally, only one study used SEM to account for measurement error in examining the relative stability of temperament in young children. Using a sample of 94 four year-old children assessed at two separate time points seven months apart, Majdandžić and van den Boom (2007) developed separate structural models for each temperament dimension (i.e., positive emotionality/exuberance, interest, anger, sadness) based on a combination of laboratory observations and parent-report questionnaires. A structural model could not be fit for the fear dimension due to low intercorrelations between the fear episodes. SEM estimates of relative stability for the dimensions of exuberance/positive emotionality (.76), interest (1.00), and anger (.55) were in the high range. However, sadness (.34) did not demonstrate significant stability across time. The stability estimates derived from SEM were higher than those based on correlational analyses. Unfortunately, the sample size was small for conducting SEM. As a

result, the authors were forced to test separate models for each trait and were unable to simultaneously examine structural, as well as relative, stability.

Rationale for the Present Study

As this literature review indicates, no studies have examined the structural stability of temperament assessed using laboratory-observational measures. Moreover, there are only a limited number of studies of the relative stability of laboratory-assessed temperament traits, and most have been focused on infants and used short time intervals between assessments. Considering the differences in approach and the low correlations between parent-report and laboratory-observational measures, it is possible that a laboratory-observational measure will produce a different picture of the structural and relative stability of temperament in young children compared with parent-report measures. In a recent study, we used a two-stage factor analytic approach (EFA on one sample followed by CFA on a second sample) to examine the structure of temperament in three year-old children using a laboratory-observational measure (Laboratory Assessment Temperament Battery [Lab-TAB; Goldsmith, Reilly, & Lemery, Longley, & Prescott, 1995]) (Dyson, Olino, Durbin, Goldsmith, & Klein, 2011). Our best-fitting model was comprised of five-higher order dimensions, Sociability, Positive Affect/Interest, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint, which overlapped somewhat with the five-factor traits of adult personality as well as several models derived from parent-report measures of temperament. However, our model also had several notable deviations from prior models. First, our model included two factors, PA/Interest and Sociability, which both generally fall under the broad dimension of Extraversion in the literature based on self- and parent-reports. This suggests that either the common facets of extraversion may not coalesce into one higher order factor until later in development, or that our laboratory tasks tap somewhat different

information than parent-report. Second, in our model, the Neuroticism/Negative Emotionality factor found in most prior models was split into two separate factors, Dysphoria and Fear/Inhibition. Again, this suggests that the core components of Neuroticism/Negative Emotionality may not be consolidated in early childhood (i.e., the lower-order traits of anger and sadness were highly associated, whereas fear was distinct from other negative emotions), or that this reflects the differences between parent-report and laboratory-observational measures. Third, we did not find Agreeableness or Openness factors. This finding was not surprising given that our Lab-TAB measure was not designed to directly assess either of these constructs. However, our Sociability factor overlapped with Agreeableness, as both are defined by prosocial, affiliative, and dominant behavior.

It is also important to note that this study was cross-sectional, which precluded us from investigating the structural and relative stability of these temperament dimensions. However, we recently conducted a follow-up of our sample at age 6. The current study uses these data to address two specific aims. The first aim involved deriving a structural model of temperament for the age 6 sample based on a laboratory-observational measure ([Lab-TAB], Goldsmith et al., 1995) and then examining the structural stability between our age 3 and 6 models. We hypothesized that our age 3 model would be a good fit to our sample at age 6, demonstrating structural stability over time. Our second aim was to estimate the relative stability of the temperament dimensions from age 3 to 6. Based on the research reviewed above assessing relative stability in young children (e.g., Durbin et al., 2007; Lemery et al., 1999; Majdandžić and van den Boom, 2007; Pedlow et al., 1993; Roberts & DelVecchio, 2000), we hypothesized that we would find moderate relative stability between our laboratory-assessed trait dimensions at ages 3 and 6.

Method

Participants and Demographics

Age 3. The sample consisted of 550 (54.0% male and 46.0% female) three-year old children from Long Island, NY who participated in a longitudinal study of temperamental emotionality. The mean age of the children was 42.2 months (*SD* = 3.1). Participants were recruited through commercial mailing lists and were initially contacted by the Stony Brook University Center for Survey Research. Families with a child between three-to-four years of age who lived with at least one English-speaking biological parent and did not have any significant medical conditions or developmental disabilities were eligible for participation. Of eligible families, 69.1% entered the study and completed the laboratory temperament assessment. Following a detailed description of the study, written informed consent was obtained from all the families. The families were financially compensated for their participation. The sample was primarily White/European-American (87.1%) and middle class, as measured by the Hollingshead's Four Factor Index of Social Status (Hollingshead, 1975; *M* = 54.2; *SD* = 11). At the age 3 assessment, the mean ages of the mothers and fathers were 36.0 (*SD* = 4.4) and 38.3 years (*SD* = 5.4), respectively. The majority (94.2%) of the children came from two-parent homes, and 51.4% of the mothers worked outside of the home part- or full-time. 55.0 % of the mothers and 47.0 % of the fathers had a college degree or higher.

Age 6. Of the original sample, 452 (54.0% male and 46.0 % female) children participated in the age 6 laboratory assessment. The mean age of the children was 73.1 months (*SD* = 4.98). The sample was primarily White/European-American (89.2%). Regarding schooling, 45.6% of the children attended full-day kindergarten, 39.8% of the children attended first grade, 5.3% of the children attended half-day kindergarten, 4.0% of the children attended second grade, and

4.4% of the children attended some other form of schooling or were homeschooled. At the age 6 assessment, the mean ages of the mothers and fathers were 39.3 (*SD* = 4.4) and 41.76 (*SD* = 6.0), respectively. The majority (89.4%) of the children came from two-parent homes, and 57.5% of the mothers worked outside of the home part- or full-time. Those children who participated in the age 6 laboratory assessment did not differ from those who did not participate on age 3 demographic variables (i.e., sex, age, ethnicity) or Lab-Tab variables of interest.

Age 3 and 6 combined. A total of 440 participants had complete data for the age 3 and 6 laboratory assessments. Thus, analyses were based on these participants.

Age 3 Assessment Procedures

Laboratory Assessment. The laboratory assessment lasted approximately two hours and included a standardized set of 11 laboratory episodes adopted from the Laboratory Temperament Assessment Battery (Lab TAB; Goldsmith et al., 1995) and one episode (Exploring New Objects) designed for this study. The Lab-TAB provides standardized episodes with emotional incentives or targets (e.g., Stranger has the emotional incentive of behavioral inhibition; Transparent Box has the emotional incentive of blocked goals) but allows flexibility in scoring these episodes (i.e., can employ both global and micro-level coding procedures). The 12 episodes that we used were selected to elicit a broad range of emotions and behaviors. Most of the episodes in the Lab-TAB were drawn from previous studies examining a number of research questions related to child social and personality development and thus have a history of successful usage in developmental psychology (Goldsmith et al., 1995). There was a short play break between episodes that allowed the children to return to a neutral affective state. All of the episodes were videotaped through a one-way mirror and later coded. A parent remained in the

room for all episodes except for Stranger Approach and Box Empty. Below is a description of each episode:

Risk Room. The child was left alone to explore a set of novel and ambiguous stimuli, including a large black box with eyes and teeth, a cloth tunnel, a Halloween mask, balance beam, and small staircase. After five minutes, the experimenter returned to the room and asked the child to engage in play with each object. This Lab-TAB episode was derived a series of studies by Kagan and colleagues (e.g., Kagan, 1998; Kagan, Reznick, & Snidman, 1986).

Tower of Patience. The child and experimenter alternated turns building a tower together with large blocks. During each turn, the experimenter increased delays in placing the block on the tower, making the child wait. This Lab-TAB episode, like some of the other episodes tapping impulsive behavior, was derived from prior research by Kochanska and colleagues (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996).

Arc of Toys. The child was allowed to play freely by him/herself in a room with toys for a few minutes, after which the experimenter returned and asked the child to clean up the toys.

Stranger. The child was briefly left alone in the empty assessment room while the experimenter went to look for other toys. In the experimenter's absence, a male research assistant entered the room and spoke to the child in a neutral tone while gradually walking closer to the child. At the end of the episode, the experimenter entered the room and introduced the male stranger to the child as her friend.

Car Go. The child and experimenter raced remotely controlled cars.

Transparent Box. The child selected a toy, which was then locked in a transparent box. The child was then left alone in the room with a set of incorrect keys to use to open the box.

After a few minutes, the experimenter returned, gave the child the correct key, and encouraged the child to use the new key to open the box and play with the toy.

Exploring New Objects. The child was left alone to explore a set of novel and ambiguous stimuli, including pretend mice in a cage, sticky water-filled gel balls, a mechanical bird, a mechanical spider, and a pretend skull covered under a blanket. After five minutes, the experimenter returned and asked the child to play with each object.

Pop-up Snakes. The experimenter showed the child what appeared to be a can of potato chips, which actually contained coiled spring "snakes." The experimenter then encouraged the child to surprise the child's parent with the can of snakes.

Impossibly Perfect Green Circles. The child was instructed to repeatedly draw a circle on a large piece of paper. After each drawing, the circle was mildly criticized

Popping Bubbles. The child and experimenter played with a bubble-shooting toy.

Snack Delay. The child was instructed to wait for the experimenter to ring a bell before eating a snack. The experimenter systematically delayed ringing the bell.

Box Empty. The child was given a box to unwrap, but rather than containing a present, the box was empty. After the child discovered that the box was empty, the experimenter returned with several small toys for the child to keep.

Laboratory Coding Procedures. We selected constructs based on the literature on the structure of temperament/personality in youth (e.g., Caspi & Shiner, 2006; De Pauw et. al., 2009; De Pauw & Mervielde, 2010), and attempted to include all constructs that could be coded using laboratory observations. Coding schemes were selected from existing coding systems (e.g., Carlson, 2005; Durbin et al., 2007; Goldsmith et al., 1995; Kagan et al., 1984). Different coding methods were employed for the affective, behavioral, behavioral inhibition (BI), and inhibitory

control variables. For almost all variables, we combined ratings across episodes to create cross-situational indices and reduce the impact of episode-specific influences. The episodes were coded by undergraduate research assistants, study staff, and graduate students who completed extensive training and were unaware of other study variables. Coders were assigned to specific episodes and had to reach at least 80% agreement on all specific codes within the episode with a "master" rater before coding independently. To examine interrater reliability, videotapes of 35 children were independently coded by a second rater (only 8 children were used to assess interrater reliability of Inhibitory Control because it uses simple count variables). In order to calculate the interclass correlation coefficient (ICC), a two-way random, absolute agreement interrater ICC was used (Shrout & Fleiss, 1979). We also examined the internal consistency of the scales using coefficient alpha based on the entire sample (N = 559).

Affective traits. Each instance (i.e., time stamp recorded) of facial, bodily, and vocal positive affect, anger, sadness, and fear were rated on a 3-point scale (low intensity, moderate intensity, high intensity) during all 12 episodes. Within each episode, these intensity ratings were summed within each channel (facial, bodily, vocal) (e.g., sum of low, moderate, and high facial affect in the Risk Room episode), which produced 36 scores (facial affect for 12 episodes, bodily affect for 12 episodes, vocal affect for 12 episodes) for each of the four affective traits. The intensity ratings were then averaged within each channel across all 12 episodes (e.g., we computed the mean of the sum of low, moderate, and high facial PA in Risk Room + the mean of the sum of low, moderate, and high facial PA in Tower; etc.), which resulted in scores for each of the three channels for each of the four affective traits. Each of these 12 variables was then standardized (e.g., we standardized the mean of the sum of the facial PA intensity scores across all episodes, standardized the mean of the sum of the bodily PA intensity scores across all

episodes, and standardized the mean of the sum of vocal PA intensity scores across all episodes). This resulted in a standardized score for each channel for each affective trait. Finally, the standardized scores for the three channels were then averaged for each affect (e.g., PA = (standardized facial PA + standardized bodily PA + standardized vocal PA)/3). Coefficient alpha for the positive affect (PA), anger, sadness, and fear scales were .87, .68, .81, and .63, respectively. Interrater ICCs (N=35) for PA, sadness, anger, and fear were .92, .79, .73, and .64, respectively.

Other behavioral traits. Global ratings of the behavioral trait variables were derived using all of the relevant behaviors during that episode. The following variables were rated on a single 4-point Likert scale (0 = low, 1 = moderate, 2 = moderate to high, and 3 = high): Interest (α = .68, ICC = .84) was based on how engaged the child appeared in play. Anticipatory PA (α = .70, ICC = .63) was based on PA that occurred in anticipation of a reward, reinforcer, or positive event. Initiative (α = .74, ICC = .70) was based on the degree of passivity or assertiveness the child displayed in their interactions with others. Activity (α = .73, ICC = .75) was based on movement during each episode as well as the amount of vigor exhibited in the manipulation of the stimuli. Sociability (α = .83, ICC = .83) was based on the child's attempts to engage and interact with the experimenter and the parent. Compliance (α = .77, ICC = .85) was based on the severity of "rule-breaking", the persistence of the noncompliance, and the degree to which these behaviors were judged to reflect an intentional unwillingness to comply with the experimenter's or parent's suggestions, requests, or demands. Impulsivity (α = .70, ICC = .75) was based on the child's tendency to act or respond without reflection or hesitation.

The following variables were rated on the degree to which the child exhibited the behavior during the episode (0 = none, 1 = slightly, 2 = somewhat, 3 = quite a bit, 4 = very

much): Assertiveness (α = .69, ICC = .59) was based on the degree to which the child made requests or demands, offered suggestions, or drew attention to him/herself. Domineering/Pushy (α = .70, ICC = .87) was based on the degree to which the child made demands, was actively noncompliant, and argued with the experimenter or mother. Hostility (α = .60, ICC = .84) was based on the degree to which the child directed physical or verbal aggression or angry comments at the experimenter or mother. Clingy (α = .70, ICC = .51) was based on the degree of clingy behavior, proximity-seeking, and reassurance-seeking directed at the experimenter or parent, and needing the experimenter or parent to participate in order to play. Lastly, social anxiousness (α = .50, ICC = .62) was based on the degree of nervous smiling, sad response to criticism, and submissive behavior.

Dominance vs. submissiveness and warmth vs. hostility were rated on a 11-point Likert scale (-5 [extremely negative] to 5 [extremely positive]) because these traits are bivalent. Dominance (α = .76, ICC = .59) was based on the degree of social potency demonstrated by the child in the interaction. High scores reflected high levels of dominant behavior, whereas negative scores indicated submissiveness and passivity. Warmth (α = .79, ICC =.77) was based on the degree of warmth or affiliation the child displayed in the interaction. Thus, high scores were indicative of high levels of warmth and affiliation, whereas negative scores were indicative of interpersonal hostility.

Behavioral inhibition (BI). BI refers to reactions of fearfulness, wariness, and low approach to unfamiliar people, objects, and contexts (Kagan, Reznick, Clarke, Snidman, & Garcia-Coll, 1984). The Risk Room, Stranger Approach, and Exploring New Objects episodes were coded using Goldsmith et al.'s (1995) system, which, consistent with most of the literature on BI, involves making highly specific ratings of behavioral responses at discrete time intervals

(20-30 second epochs). For the present study, the BI variable did not contain any affective ratings from the Risk Room, Stranger Approach, and Exploring New Object episodes because the affective ratings were used to create the fear variable described above. The BI composite variable (α =.74, interrater ICC = .90) was constructed by combining the average standardized ratings of the following variables from the Risk Room (RR), Stranger, and Exploring New Objects (ENO) episodes: total number of objects touched (RR and ENO only), latency to touch objects (RR and ENO only), tentative play (RR and ENO only), referencing experimenter (RR and ENO only), time spent playing (RR and ENO only), latency to vocalize, approach towards the stranger (Stranger only), avoidance of the stranger (Stranger only), gaze aversion (Stranger only), and verbal/nonverbal interaction with the stranger (Stranger only). Variables were all keyed in a consistent direction (e.g., long latencies to touch objects were keyed to reflect more BI).

Inhibitory Control/Executive Functioning Variable. The Tower of Patience and Snack Delay episodes were each coded for inhibitory control using a coding system adapted from Carlson (2005), which involved tallying the number of times a child failed to wait his or her turn during the episode. Tower of Patience consisted of 14 trials and Snack Delay consisted of seven trials. The composite global inhibitory control/executive control variable (α = .70, interrater ICC = .98, N = 8) was constructed by aggregating the scores from the two episodes.

Age 6 Follow-up Assessment Procedures

Laboratory Assessment. At age 6, the children participated in a new battery of laboratory tasks. The laboratory visit lasted approximately two hours, during which children participated in a standardized set of nine laboratory episodes that were adapted from the Lab-TAB (Goldsmith et al., 1995). Similar to the age 3 assessment, the nine episodes adapted from the Lab-TAB were

designed to elicit different emotional and behavioral expressions. Different episodes were used at the age 6 assessments to ensure that they were age-appropriate and to minimize familiarity effects. Between each episode, the child took a brief break to return to a neutral state prior to entering a new situation. The episodes are described below in the order that they were presented to the children.

Card Sorting. The child was shown cards depicting geometric figures varying in shape, number, and color, and were taught to sort the cards by color. The child sorted the cards for several timed trials that varied by outcome (contingent reward (erasers), noncontingent reward (stickers), and punishment (take away erasers)) and by the number of sorted cards required to obtain or avoid the contingency.

Mixed-Up Puzzles. The experimenter told the child to put together a puzzle that is "really easy"; however, the child was given the pieces from two similar but different puzzles, making it impossible to complete. The experimenter left the room for 3 minutes. When the experimenter returned, the child was told that the incorrect pieces were given to her/him and that it was impossible to put the puzzle together with them.

Story Time. The child was asked to tell a story to an unfamiliar research assistant, whom the experimenter described as a "story expert." The child was given a picture book, "*A Boy, a Dog, and a Frog"* by Mercer Mayer, and was asked to use the book to tell the story to the assistant, who would give the child a grade on how well s/he told the story. The child was given a maximum of four minutes to tell the story to the assistant. When the experimenter returned and asked the assistant about the child's performance, the child was praised by the assistant as an excellent story teller and received an A+.

Disappointing Toy. The child was shown three photographs of toys that varied in interest: a watering can, bunny hand puppets, and a remote-controlled car. The child was asked to choose the toy that they wanted to play with the most. The experimenter then left the photograph of that toy with the child while she left the room to get the toy. The experimenter returned after a brief period of time and told the child that the desired toy is currently being played with by another child. The experimenter gave the child one of the disappointing toys and left the child alone for two minutes to play with the undesirable toy. The experimenter returned with toy that the child had originally wanted and the child was given two minutes to play with the toy.

Picture Tearing. The child was shown a photo album by a research assistant. The assistant emphasized how special the photographs were to her/him, especially the photograph of an older couple, who the assistant described as her/his grandparents. The assistant left the room and the photo album with the child. The experimenter entered the room and told the child to rip up the picture of the assistant's "grandparents." The experimenter provided prompts to the child until the child either ripped the picture or two minutes elapsed. When the assistant returned to the room, s/he asked the child what had happened to the photograph. The assistant then reassured the child that there s/he has another copy of the photograph that was destroyed. The experimenter then apologized to the child for asking her/him to rip up the assistant's photograph.

Dress Up. The child was shown a variety of clothes and props (e.g., fireman, doctor) and was permitted to dress up in the items.

Kids' Club. The child was told that s/he was going to be interviewed for admission to a club "just for kids" (Erdley, Cain, Lomis, Dumas-Hines, & Dweck, 1997). An unfamiliar interviewer asked the child a series of questions under the pretense that s/he needed to determine whether s/he would get along well with the other club members. The interviewer told the child

that s/he was going to send the information to the club president by computer, so that that president could immediately decide whether the child would be admitted. After a brief delay, the assistant returned, stating that the president was not sure about whether the child should be admitted and that the president wanted to know more about the child before making a decision. The child was then allowed to choose whether s/he wanted to reapply to club by providing more information about her/himself. If the child chose to reapply, then the assistant asked the child further questions. The assistant then left the room to send the additional information. The child was then asked several questions by the experimenter to assess her/his attribution for the ambiguous rejection. The episode ended when the assistant returned with a certificate of membership, explaining that s/he had actually made it into the club from the very beginning and that the president just wanted to know more about the child.

Whoopee Cushion. The experimenter showed the child a remote-controlled electronic box that emits whooped cushion sounds. The child was allowed to test the noise using the remote control. The experimenter then invited the child to "trick" her/his mother with the whoopee cushion by hiding it under a chair. When the mother entered the room, the child asked the mother to sit in the chair and then used the remote control to activate the whoopee cushion.

Object Fear. The child was instructed to explore a room that was filled with fear-eliciting objects, including a box filled with plastic insects and from which cricket sounds were emitted, a cage with plastic rates inside it, and a large, fuzzy, black spider covered with a cloth.

Laboratory Coding Procedures. The coding methods and interrater reliability procedures that were used at the age 3 assessment were also employed with this data. Alphas and interrater reliabilities are presented below for the affective, behavioral, and behavioral inhibition variables.

Affective Variables. Positive Affect (PA) (α = .83, *ICC* = .95), Fear (α = .50, *ICC* = .68), Sadness (α = .72, *ICC* = .79), Anger (α = .63, *ICC* = .77).

Behavioral Variables. Interest (α = .69, *ICC* = .78), Activity Level (α = .72, *ICC* = .72), Anticipatory PA (α = .50, *ICC* = .53), Initiative (α = .75, *ICC* = .84), Sociability (α = .80, *ICC* =.84), Compliance (α = .48, *ICC* = .78), Impulsivity (α = .65, *ICC* = .76), Dominance (α = .70, *ICC* = .66), Warmth (α = .69, *ICC* = .77), Social Interest (α =.75, *ICC* =.73), Affiliative (α = .73, *ICC* = .75), Assertive (α = .76, *ICC* = .81), Domineering/Pushy (α = .55, *ICC* = .50), Avoidant (α = .67, *ICC* = .67, Socially Anxious (α = .31, *ICC* = .55). Due to low frequency and reliability, the variables of clingy and hostility were not used. The low frequency of these behaviors at age 6 was not surprising, as it is developmentally appropriate for these behaviors to decrease over time and as children learn how to self-regulate their emotions. Lastly, inhibitory control was not assessed during the age 6 Lab-TAB and could not be included in our age 6 model.

Behavioral Inhibition variable. BI was coded from two episodes: Story Telling and Object Fear (α =.51, *ICC* = .773). Similar to our age 3 BI variable, the age 6 BI variable did not contain any affective ratings from these episodes because the affective ratings were used to create the age 6 Lab-TAB fear variable.

Tester Impression Variables. In addition to the affective, behavioral, and BI variables described above, global ratings of affect based on all of the episodes were made by the experimenter at the conclusion of the laboratory assessment. The following variables were rated on a single 5-point Likert scale (1= rarely, 2= subtle or ambiguous signs, 3= mild, 4= moderate, 5=extreme): positive affect, anger, sadness, and fear. Interrater reliabilities are not available for these variables. Likewise coefficient alpha cannot be calculated for these variables because they are based on only one item. As noted above, the variables of clingy and hostility were not used as

indicators in our age 6 model due to low frequency and reliability, thus, the variables of sadness, anger, and fear were included in our age 6 model in order to have at least three indicators per factor.

Results

Preliminary Analyses. A number of variables were transformed to reduce kurtosis. Log transformations were applied to dominance, anger (Lab-TAB), sadness (Lab-TAB), impulsivity, domineering/pushy, and compliance. All variables were then standardized.

Deriving an Age 6 Structural Model of Temperament

Confirmatory Factor Analysis (CFA): Examining the Fit of the Age 3 Model to the Age 6 Sample. Confirmatory factor analysis (CFA) was conducted using maximum-likelihood estimation procedures using AMOS 18.0 with the age 6 sample (N = 440) to examine the fit of the same five-factor structure obtained using CFA in the age 3 sample. To assess model fit, the following criteria were used: (a) chi-square statistic; (b) the root-mean-square error (RMSEA; Steigher & Lind, 1980); and (c) the comparative fit index (CFI; Bentler, 1990). Since the chi-square statistic is often significant in moderate to large samples, less weight is given to it compared to the other fit indices. Based on recent discussion of the challenges of applying CFA in temperament and personality research (e.g., Marsh, Hau, & Wen, 2004; Marsh et al., 2010), Hopwood and Donnellan's recommended cutoff values of RMSEA < .10 and CFI > .90 were used to determine acceptable model fit. The target model consisted of five-factors, PA/Interest, Sociability, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint, and 16-indicator variables (see Figure 1). The same indicators (age 6 versions) were used in this model as the original age 3 model, with a few exceptions. The age 6 tester impression fear variable was used as an indicator for the Fear/Inhibition factor instead of the clingy variable, the age 6 tester

impression sadness and anger variables were used as indicators for the Dysphoria factor instead of hostility, and the domineering/pushy variable was used as an indicator for the Impulsivity vs. Constraint factor because we did not have an inhibitory control variable for the age 6 assessment. Fit indices for this model are presented in Table 1. The age 3 model was an acceptable fit to the age 6 data, with a RMSEA of .098 (CI = .089-.106) and CFI of .888.

Post Hoc Model Fitting. To improve the age 6 CFA model fit, the model estimates and modification indices (MIs) were examined. The model estimates indicated that the correlated paths between the latent factors of PA/Interest and Fear/Inhibition (r = .008, p = .890), PA/Interest and Impulsivity vs. Constraint (r = .08, p = .211), and Sociability and Fear/Inhibition (r = .04, p = .21), and the residuals of interest and initiative (r = .02, p =.223) were non-significant and subsequently removed from the model in the interest of parsimony. Based on the MIs, we made three methodologically or theoretically meaningful post hoc modifications (see Figure 2). First, we correlated the residuals between the tester impression sadness and fear indicators, as both of these indicators tap aspects of negative affect and were assessed using the same measure. Second, the tester impression sadness and Lab-TAB sadness indicators were correlated based on the justification that both variables tap sad affect in the laboratory. Third, we added a (negative) path from the Fear/Inhibition factor and the interest indicator. The rationale for this modification was based on the behavioral inhibition literature, which is characterized as fearfulness, hesitancy, and wariness to novel social and nonsocial stimuli (Kagan et al., 1984). Ratings of interest in our laboratory assessment were based on whether a child approached and engaged in play with the novel stimuli. Therefore, children high in inhibition and fear exhibited low approach and interest in unfamiliar situations.

The fit of the model incorporating these modifications is shown in Table 1. Model fit was improved and good, with a RMSEA = .083 (90% CI = .075-.092) and CFI =.918.

Structural Stability of Latent Factors Between Age 3 and 6. As demonstrated by the analyses above, the age 3 and age 6 models are both comprised of the five-higher order factors of PA/Interest, Sociability, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint, suggesting good structural stability from age 3 to 6. Furthermore, similar to the age 3 model, the age 6 model exhibited significant correlations between the latent Sociability and PA/Interest factors (.88), the Sociability and Impulsivity vs. Constraint factors (.28), and the Impulsivity vs. Constraint and Dysphoria factors (.66). Unlike the age 3 model, the covariance paths between the latent PA/Interest and Fear/Inhibition, PA/Interest and Impulsivity vs. Constraint, and Sociability and Fear/Inhibition were non-significant and removed from the model. Ordinarily, the next step would be to formally test structural invariance using SEM between the age 3 and 6 models, which would involve determining whether the factor variances, covariances, and means are the same across age 3 and 6. However, prior to testing structural invariance, measurement invariance must first be established, which involves examining how the same observed variables measure the latent construct over time. Unfortunately, we could not formally assess measurement and structural invariance because the indicators in the two models were not identical (e.g., clinginess and hostility were too infrequent and difficult to rate reliably at age 6 to include in the model).

Relative Stability of Factors Between Age 3 to 6. Although this study used SEM to estimate the relative stability of temperament traits, the cross-time bivariate correlations between the age 3 and 6 indicators/lower-order traits were also examined and presented in Table 3. Overall, the age 3 indicators/lower-order traits exhibited low to moderate stability with their age 6 counterparts.

SEM procedures were used to estimate the relative stability of the latent factors between the age 3 and 6 temperament models. The combined age 3 and 6 model is depicted in Figure 3 and yielded a good fit to the data, with a RMSEA of .065 (90% CI = .061-.070) and CFI = .894. More specifically, as shown in Table 2 and Figure 3, all five latent factors demonstrated significant relative stability from age 3 to 6. Stability estimates ranged from low to moderate (.17-.47). Furthermore, two significant temporal associations were observed between different latent factors, specifically between the latent age 3 Sociability and age 6 PA/Interest factors, and the latent age 3 Impulsivity vs. Constraint and age 6 Fear/Inhibition factors. Again, estimates of these paths ranged from low to moderate (.17-.38).

Discussion

Despite the continued debate surrounding the structure of temperament traits in young children, there remains an underlying assumption that temperament traits exhibit stability across a time. Researchers have relied primarily on parent-report measures to investigate the structural and relative stability of temperament traits, which are time-and cost-effective and allow one to draw upon a parent's vast knowledge of the child's emotional and behavioral reactions; however, they are also susceptible to a number of response and perceptual biases. Although more costly and time intensive, laboratory-observational measures have several advantages (e.g., standardized procedures, objective criteria) relative to parent-report measures and may provide an alternative perspective about the stability of temperament. Thus, given the dearth of extant studies examining the structural and relative stability of temperament in young children using a laboratory-observational measure, the present study sought to extend the literature as well as our own previous study of three-year-old children.

Structural stability of temperament traits from age 3 to 6. A few cross-sectional studies have examined the structural stability of temperament in young children (e.g., Rothbart et al., 2001; Halverson et al., 2003) using EFA with different-aged children. However, compared to longitudinal designs, these cross-sectional studies may be confounded by sample and developmental effects. Therefore, in the present study, we used a longitudinal design and CFA to examine the fit of our age 3 five-factor model on the full age 6 sample. With a few modifications, this model was a good fit to the age 6 sample and consisted of the same five-factors (i.e., PA/Interest, Sociability, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint) as the age 3 model. Although we were unable to test for measurement and structural invariance due to developmental changes in the frequency of a few indicators, this finding supports our initial hypothesis that the age 3 and 6 models would be comprised of similar higher-order dimensions and suggests a high level of structural stability from age 3 to 6. In what follows is a brief discussion of the most notable similarities and differences between the factor structure of our age 3 and 6 models.

Similar to the age 3 model, the age 6 model was comprised of two higher-order factors, PA/Interest and Sociability, that fall under the broad dimension of extraversion. More specifically, both PA/Interest factors were comprised of the core facet of extraversion, PA, and the appetitive, reward-seeking facets of anticipatory PA and interest. Similar to the age 3 model, all three indicators loaded highly on the age 6 PA/Interest factor, with interest having the highest loading, followed by PA and anticipatory PA. Further, in both models, impulsivity loaded (moderately) on the PA/Interest factor, which is deemed an essential facet of extraversion in several models (e.g., Depue & Collins, 1999; Eysenck & Eysenck, 1985), and the residuals between the PA and anticipatory PA indicators are correlated.

The age 3 and 6 Sociability factors were both comprised of the traits of sociability, dominance, and initiative, which are characterized as interpersonal surgent traits (i.e., traits related to agency and affiliation) in some models of extraversion (Depue & Collins, 1999). Akin to the age 3 model, all three indicators loaded highly on the age 6 Sociability factor, with the sociability indicator demonstrating the highest loading. Furthermore, the stability of the separate PA/Interest and Sociability factors from age 3 to 6 suggests that the core components of extraversion may not fully consolidate into one cohesive factor until later development. However, as demonstrated by the strong correlation (.88) between these two latent factors at age 6 and the relative stability estimates discussed in more detail below, there may be a trend toward convergence between the PA/Interest and Sociability factor over time. Alternatively, it is possible that the emergence of two different factors may be a reflection of our laboratory-observational measure, as most of our tasks involved some level of social interaction between the experimenter and child, resulting in a separate Sociability factor.

Analogous to the age 3 model, the age 6 model contained two higher-order factors, Dysphoria and Fear/Inhibition, that fall under the rubric of neuroticism/negative emotionality. The emergence of separate factors at both age 3 and age 6 are consistent with research suggesting that anger and sadness are associated because both emotions are elicited by goal blockage or loss, whereas fear is related to the anticipation of punishment or loss (Lewis & Ramsey, 2005; Putnam, Ellis, Rothbart, 2005). In contrast to the age 3 model, the age 6 Dysphoria factor included the tester impression variables of anger and sadness instead of hostility and included a correlated residual between the tester impression sadness and Lab-Tab sadness residual, as both variables tap sad affect in the laboratory task. However, similar to the age 3 Dysphoria factor, all of the indicators had moderate to high loadings on the age 6

Dysphoria factor, with the anger indicators demonstrating the highest loadings followed by the sadness indicators.

Analogous to the age 3 factor, the age 6 Fear/Inhibition factor included the facets of BI and fear, but clinginess was excluded from consideration at age 6 due to its low frequency. At both age 3 and 6, the indicators had moderate to high loadings on the Fear/Inhibition factor; however, at age 3, BI had the highest loading, whereas at age 6 the fear indicators had the highest loadings. Additionally, unlike the age 3 factor, the age 6 Fear/Inhibition factor was negatively associated with interest, which suggests that fearfulness/inhibition at age 6 is related to low approach and interest in the laboratory setting. Thus, the distinct Dysphoria and Fear/Inhibition factors at age 3 and 6 suggest that the common facets of neuroticism/negative emotionality remain separate in early childhood and my not coalesce until later childhood or adolescence. On the other hand, it is possible that this division reflects the differences between laboratory-observational and parent-report measures.

Both age 3 and 6 models included factors characterized as Impulsivity vs. Constraint. Specifically, both age 3 and 6 factors encompassed impulsivity, a facet related to impulse control and behavioral regulation, and compliance, a facet related to the ability demonstrate respect for authority and follow rules. With regard to the factor loadings, all of the indicators exhibited moderate to high loadings on the both age 3 and 6 Impulsivity vs. Constraint factors, with compliance exhibiting the highest loading. Additionally, the Impulsivity vs. Constraint factor was negatively associated with sociability at both age 3 and 6, which is consistent with the research suggesting that lower impulsivity and good behavioral control and constraint are related to appropriate behavior and successful social interactions (e.g., Eisenberg et al., 2004; Eisenberg et al., 2000; Lengua, 2003). Lastly, we did not assess the age 3 measure of inhibitory control at

age 6, and thus could not include it our age 6 model. Instead, our age 6 factor included the domineering/pushy facet, another indicator of behavior regulation and control characterized by demandingness, noncompliance, and arguing with the experimenter, which loaded highly on the factor.

Relative stability of temperament traits from age 3 to 6. Only a limited number of studies (e.g., Goldsmith & Campos, 1990; Belsky et al., 1996) have examined the relative stability of multiple laboratory-assessed traits within the same study; nonetheless, most of these studies used cross-time bivariate correlations to estimate stability which are susceptible to measurement error, had brief intervals in between assessment periods which maximizes stability, and used samples of infants and toddlers. The present study utilized SEM procedures (i.e., accounting for measurement error) to examine the relative stability of laboratory-assessed temperament traits over a three year interval from ages 3 to 6. Furthermore, unlike previous studies, the stability of these traits were all examined within the same structural model.

Overall, all five of the latent factors exhibited significant stability from age 3 to 6. As shown in Table 2, the latent factors of Sociability (.47), Dysphoria (.30), and Impulsivity vs. Constraint (.30) all exhibited moderate relative stability from age 3 to 6, whereas the latent factors of PA/Interest (.17) and Fear/Inhibition (.26) exhibited low relative stability.

Our estimates exhibit some similarities and differences from previous studies examining the stability of laboratory-assessed traits in young children. More specifically, although they utilized cross-time bivariate correlations and did not use structural models with higher-order latent factors, Durbin and colleagues (2007) also reported moderate to high relative stability for sociability, moderate stability for the Negative Emotionality traits of sadness and anger, and low stability for fear. In contrast to our PA/Interest factor, Durbin et al. (2007) found high stability

for the Positive Emotionality traits of PA, anticipatory PA, and interest. In the only study to date to use SEM to examine the relative stability of traits in young children, Majdandžić and van den Boom (2007) reported higher stability estimates for their interest and positive emotionality/exuberance factors compared to our PA/Interest factor. However, in contrast to our moderately stable Dysphoria factor, they reported non-significant stability for their sadness and fear factors and high stability for their anger factor.

Additionally, there were some significant temporal associations between different latent factors from age 3 to 6. Specifically, there was a moderate link (.38) between the latent age 3 Sociability and the age 6 PA/Interest factors, which suggests that sociability is associated with high PA at age 6. This association may account for the lower stability estimate between the age 3 and 6 PA/Interest factors and suggests a tendency for convergence over development, which is consistent with the single factor (e.g., extraversion) often found in adolescents and adult models of personality. Furthermore, this finding is consistent with those experimental studies demonstrating the causal influence of extraverted (i.e., talkative/sociable, dominant) behavior on (increased) positive affect (McNiel & Fleeson, 2006; McNiel, Lowman, & Fleeson, 2010).

Based on the strong association between these two factors, a structural model consisting of a combined PA/Interest-Sociability factor (i.e., four-factor model) was tested on the age 6 data. In order to compare this model with the five-factor model, we examined the Bayesian information criteria (BIC) and Akaike information criteria (AIC) for both models. A lower BIC and AIC are considered a better fit to the data. However, the four-factor model (BIC = 828.00, AIC = 664.17) did not fit the data better than the five-factor model with distinct PA/Interest and Sociability factors (BIC = 645.75, AIC = 473.72).

Interestingly, the latent age 3 Impulsivity vs. Constraint factor was significantly associated with the age 6 Fear/Inhibition factor (.17). Based on the cross-time correlations, it appears that compliance is strongly contributing to this association between these factors such that fearfulness and inhibition at age 6 is related to more noncompliance at age 3. One possible interpretation for this finding is that fearfulness, inhibition, and/or anxiety may manifest as noncompliance or impulsivity at age 3 but evolve into more "typical" expressions of fearfulness and inhibition at age 6. This finding is consistent with the treatment literature that emphasizes age-specific or developmental expressions of anxiety. For instance, compared to slightly older children, young preschoolers may not have developed the communication and cognitive skills necessary to articulate their fears or engage in appropriate self-control strategies taught to older children (Eisen & Kearney, 1995). Thus, fear and anxiety in young preschoolers may manifest itself through noncompliance, refusal to engage or play, outbursts/tantrums, and/or other inappropriate, disruptive behaviors (e.g., hitting, throwing toys) (Albano, Chorpita, Barlow, 2003; Campbell, 2006, Pincus, Eyberg, Choate, 2005).

Limitations and Future Directions

A few factors limit the interpretation of our results. First, the use of CFA/SEM with temperament and personality data has proven highly challenging due to frequent secondary-and cross-loadings across items (Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1994). Although several methodologically and theoretically meaningful post-hoc modifications were made in order to improve the age 6 model fit, it is noteworthy that we were able to derive a good-fitting model given the low success rate of prior CFA studies in this domain.

Second, as noted in the Methods and Results sections, we were unable to test for measurement and structural invariance because we did not have the same indicators at both assessment time points, which would have potentially provided a more stringent test of the structural stability of temperament. As such, future studies of temperament in young children should strive to utilize the same indicators at each time point in order to gain better estimates of structural stability. However, this will be challenging, as some tasks and indicators are more appropriate or frequent at some ages than others (e.g., clinginess is more frequent in three year-olds than six year-olds).

Third, Lab-TAB tasks were developed to elicit specific behaviors and emotions, which increases the chances of observing relevant responses. However, it raises the question of whether the child's emotion or behavior is largely attributable to situation-specific, rather than trait, influences. In order maximize cross-situational variance, most of the variables were averaged across all of the episodes before including them in the analyses. This approach also permits the inclusion of emotions or behaviors that are atypical in particular situations (e.g., fearfulness in an episode designed to elicit exuberance); these infrequent or atypical behaviors may be especially informative with regard to temperament.

Fourth, the participants in our sample were predominately White/European American and middle class. Although these socio-demographic variables are representative of the population in the geographic region, they may constrain the generalizability of our findings. Future studies should examine the structural and relative stability of temperament in young children utilizing a more ethnically and economically diverse sample.

Lastly, the primary goal of the current study was to build upon our previous work by estimating the structural and relative stability of laboratory-assessed temperament traits in a

longitudinal sample of young children. Thus, it was beyond the scope of the current study to examine gender differences in the structural and relative stability of temperament in young children; however, in a future study we hope to investigate gender invariance using a laboratory-observational measure to determine whether the structural and relative stability varies as a function of gender. As a starting point, Olino and colleagues (under review) recently examined gender differences in temperament traits assessed by multiple methods (laboratory-observational, maternal-report, and paternal-report) across several samples of preschool and early school-aged children. For laboratory –observational measures, they found that girls demonstrated higher levels of sociability and fear and lower levels of overall negative emotionality (NE), sadness, anger, and impulsivity compared to boys. (Olino et al.,under review).

In sum, this study aimed to extend our previous work by using an alternative approach to parent-report measures, a laboratory-observational-measure, to examine the structural and relative stability of temperament in young children. Using SEM procedures, we identified a similar five-factor structure in both our age 3 and 6 samples, suggesting good structural stability. Those variations in the structure at age 3 and age 6 may be attributed to differences in age or measurement. All five of the latent factors demonstrated significant relative stability between age 3 and 6. Additionally, two significant temporal associations were found between different latent factors (i.e., Sociability and PA/Interest and Impulsivity vs. Constraint and Fear/Inhibition), which may be indicative of potential convergence in later childhood or age-specific differences in the expression of traits (e.g., fear manifests as noncompliance in preschoolers). Thus, this represents one of the first studies to utilize SEM procedures to explore the structural and relative stability of laboratory-assessed traits in young children.

References

Albano, A. M., Chorpita, B. F., & David Barlow. Childhood anxiety disorders. In Eric J. March and Russell A. Barkley (Eds.). *Child Psychopathology,* 2nd *ed.* (pp. 279-329) New York: Guilford Press.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology, 5,* 115-129.

Belsky, J., Hsieh, K. H., Crnic, K. (1996). Infant positive and negative emotionality: One dimension or two? *Developmental Psychology, 32,* 289-298.

Byrne, B. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85*, 17-32.

Campbell, M. A. (2006) Anxiety in early childhood. *Educating Young Children, 12*, 31-33.

Carlson, S. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616.

Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry, 12*, 49-66.

Caspi, A., & Roberts. B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453-84.

Caspi, A., & Shiner, R. L. (2006). Personality development. In W. Damon, R. Lerner, & N. Eisenberg (Eds.). *Handbook of child psychology: Vol. 3. Social, emotional, and personality development.* , 6th *ed.* (pp. 300-364) New York: Wiley.

Carnicero, J.A.C, Pérez-López, Del Carmen González Salina, M., Martínez-Fuentes, M. T. (2000). A longitudinal study of temperament in infancy: Stability and convergence of measures. *European Journal of Personality, 14,* 21-37.

Church, A. T., & Burke, P. J., Jr. (1994). Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. *Personality Processes and Individual Differences, 66,* 93-114.

De Fruyt, F., Bartels, M., Van Leeuwan, K. G., De Clercq, B., and Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology, 91,* 538-552.

De Pauw, S. S. W. & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry and Human Development, 41,* 313-329.

De Pauw, S. S. W., Mervielde, I., & Van Leeuwen, K. G. (2009). How are traits related to problem behavior in preschoolers? Similarities and contrasts between temperament and personality. *Journal of Abnormal Child Psychology, 37,* 309-325.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417-440.

Durbin, C. E., Hayden, E. P., Klein, D. K., Olino, T. M. (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion, 7,* 188-199.

Dyson, M.W., Olino, T. M., Durbin, C. E., Goldsmith, H. H., & Klein, D. (in press). The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion*.

Erdley, C. A., Cain, K. M., Loomis, C. C., Dumas-Hines, F., & Dweck, C. S. (1997). Relations among children's social goals, implicit personality theories, and responses to social failure. *Developmental Psychology, 33,* 263–272.

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.

Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development, 72*, 1-21.

Fraley RC, Roberts BW. 2005. Patterns of continuity: a dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychol. Rev.* 112:60–74

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment, 23*, 337-353.

Goldsmith, H. H., & Campos, J. J. (1990). The structure of infant temperamental dispositions to experience fear and pleasure: A psychometric perspective. *Child Development, 61*, 1944-1964.

Goldsmith, H. H., & Gagne, J. R., (in press). Behavioral assessment in temperament. In Handbook of Temperament Eds M Zentner & R. Shiner

Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1995). *Laboratory Temperament Assessment Battery: Preschool Version.* Unpublished manuscript.

Halverson, C. F., Havil, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavopoulos, V., et al. (2003). Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences. *Journal of Personality, 71*, 995-1026.

Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript.

Hopwood, C. J. & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review, 14*, 332-346.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J. et al. (1997). Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18*, 667-683.

Jouriles, E. N., & Thompson, S. M. (1993). Effects of mood on mothers' evaluations of children's behavior. *Journal of Family Psychology, 6*, 300-307.

Kagan, J. (1998). Biology and the child. In W. Damon (Editor-in-Chief) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional and personality development*. (5^{th} ed., pp. 177-235). New York: Wiley.

Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia-Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development, 55*, 2212-222.

Kagan, J., Snidman, N., Kahn, V., & Towsley, S. (2007). The preservation of two infant temperaments into adolescence. *Monographs of the Society for Research in Child Development, 72*, vii-vii, 1-95.

Kochanska, G., Murray, K. T., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. Child Development, 67, 490–507.

Kandler C, Bleidorn W, Riemann R, Spinath FM, Thiel W, Angleitner A. 2010. Sources of cumulative continuity in personality: a longitudinal multiple-rater twin study. *Journal of Personality and Social Psychology*, 98, 995–1008.

Krueger, R. F., Johnson W. 2008. Behavioral genetics and personality: A new look at the integration of nature and nurture. In *Handbook of Personality: Theory and Research (3rd ed.)*, Eds. OP John, RW Robins, LA Pervin, pp. 287-310. New York: Guilford

Lemery, K. A., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and child temperament. *Developmental Psychology, 35,* 189-204.

Majdandzic, M., & van den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality, 75*, 121-167.

Mangelsdorf, S. C., Schoppe, S. J., & Buur, H. (2000). The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood. In V. J. Molfese, & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 121-140). Mahwah, NJ: Lawrence Erlbaum Associates.

Marsh, H. W., Hau, K. T., & Zonlin, W. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320-341.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471-491.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluation of university teaching. *Structural Equation Modeling, 16*, 439-476.

McCrae, R. R., & Costa, P. T., Jr. (1987). *Validation* of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81-90.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality: Trait structure as a human universal. *American Psychologist, 52,* 509-516.

McCrae, R. R., Zonderman, A. B., Costa, P.T., Bond, M., Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus procustes rotation. *Journal of Personality and Social Psychology, 70,* 352-566.

McNeil, J. M., & Fleeson, W. (2006). The causal effects of extraversion on positive affect and neuroticism on negative affect: Manipulating state extraversion and state neuroticism in an experimental approach. *Journal of Research in Personality, 40,* 529-550.

McNeil, J. M., Lowman, J. C., & Fleeson, W. (2010). The effect of state extraversion on four types of affect. *European Journal of Personality, 24,* 18-35.

Olino, T. M., Durbin, C. E., Klein, D. N., Hayden, E. P., & Dyson, M. W. (under review). Gender differences in young children's temperament traits: Comparisons across observational and parent-report methods.

Pedlow, R., Sanson, A., Prior, M., Oberklaid, F. (1993). Stability of maternally reported temperament from infancy to 8 years. *Developmental Psychology, 29,* 998-1007.

Pfiefer, M., Goldsmith, H. H., Davidson, R., J., & Rickman, M. (2002). Continuity and change in inhibited and uninhibited children. *Child Development, 73,* 1474-1485.

Pincus, D. B., Eyberg, S. M., & Choate, M. L. (2005). Adapting parent-child interaction therapy for young children with separation anxiety disorder. *Education and Treatment of Children, 28,* 163-181.

Prior, M. (1992). Childhood temperament. *Journal of Child Psychology and Psychiatry, 33,* 249-279.

Putnam, S. P., Rothbart, M. K., Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development, 17,* 387-405.

Reitan, R.M., & Wolfson, D. (1992). *Neuropsychological evaluation of older children*. Tucson, AZ: Neuropsychology Press.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3-25.

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., & Bates, J.E. (2006). Temperament in children's development. In W. Damon, R. Lerner, & Eisenberg (Eds.), *Handbook of child psychology, Sixth edition: Social, emotional, and personality development,* (Vol. 3, 1-99) New York: Wiley.

Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span*, (pp. 85-119). Hillsdale, NJ: Erlbaum.

Rothbart, M. K., & Goldsmith, H. H. (1985). Three approaches to the study of infant temperament. Developmental Review, 5 (3), 237–260.

Scarpa, A., Raine, A., Venables, P., & Mednick, S. (1995). The stability of inhibited/uninhibited temperament from ages 3 to 11 years in Mauritian children. *Journal of Abnormal Child Psychology, 23,* 607-618.

Saudino, K., J. (2003). Parent ratings of infant temperament - Lessons from twin studies. *Infant Behavior and Development 26,* 314-325.

Saudino, K. J. (2009). Do different measures tap the same genetic infleunces? A multi-method study of activity level in young twins. *Developmental Science, 12,* 626-633.

Stifter, C. A., Willoughby, M. T., & Towe-Goodman, N. (2008). Agree or agree to disagree? Assessing the convergence between parents and observers on infant temperament. *Infant and Child Development, 17,* 407-426.

Zeman, J., Klimes-Dougan, B., Cassano, M., & Adrian, M. (2007). Measurement issues in emotion research with children and adolescents. *Clinical Psychology: Science and Practice, 14,* 377-401.

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *European Journal of Developmental Science, 2,* 7-37.

Zoelch, C., Seitz, K., & Schumann-Hengsteler, R. (2005). From rag(bag)s to riches: Measuring the developing central executive. In W. Schneider, R. Schumann-Hengsteler, and B. Sadian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 39-69). Mahwah, NJ: Lawrence Erlbaum.

Table 1

Summary of Fit Indices for Lab-TAB Age 6 Confirmatory Factor Analysis Model

Model	χ^2	*df*	CFI	RMSEA	CI RMSEA (Lo 90-High 90)
Original Age 3 Model with Age 6 sample	486.90***	93	.888	.098	.089-.106
Modified Age 3 model with Age 6 sample	383.68***	94	.918	.083	.075-.092
Combined Age 3 and 6 Model	1166.49***	406	.894	.065	.061-.070

*** $p \leq .001$; Note: *df* = degrees of freedom; CFI = Comparative Fit Index; RMSEA = root-mean-square error of approximation; CI RMSEA= 90% confidence interval for root-mean-square of approximation

Table 2

Estimates of Relative Stability and Temporal Associations for Latent Factors from Age 3 to 6

From Age 3 to Age 6	*Stability Estimate*
Sociability to Sociability	.47***
PA/Interest to PA/Interest	.17***
Fear/Inhibition to Fear/Inhibition	.26***
Dysphoria to Dysphoria	.30***
Impulsivity vs. Constraint to Impulsivity vs. Constraint	.31***
Sociability to PA/Interest	.38***
Impulsivity vs. Constraint to Fear/Inhibition	.17***

***$p < .001$

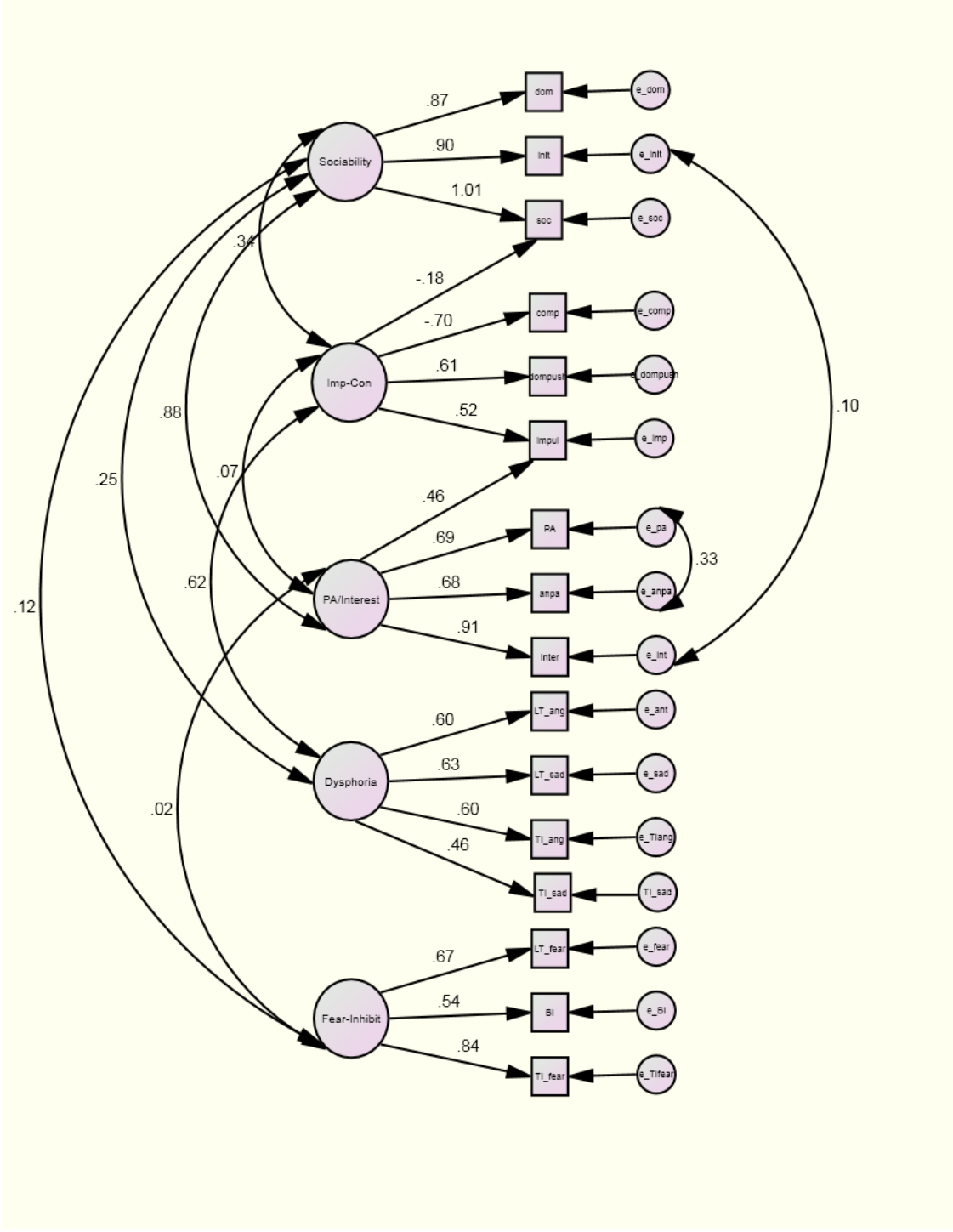
Table 3.

Cross-time Correlations between Age 3 and 6 Indicators.

	Age 6															
	PA/Interest			*Sociability*			*Dysphoria*				*Fear/Inhibition*			*Impulsivity vs. Constraint*		
Age 3	PA	AnPA	Inter	Soc	Domin	Initia	LT-Sad	TI-Sad	LT-Ang	TI-Ang	LT-Fear	TI-Fear	BI	Impul	Comp	Dom/ Push
PA	.40**	.29**	.23**	.16**	.09	.21**	.08	.01	.01	-.08	.04	-.03	-.09	.05	.08	-.06
AnPA	.29**	.22**	.17**	.10**	.07	.16**	.09	.01	.03	-.10*	-.02	-.07	-.14**	-.02	.09	-.01
Inter	.32**	.28**	.31**	.22**	.19**	.20**	-.06	-.06	-.01	-.04	-.04	-.10*	-.17**	.08	.07	.02
Soc	.32**	.29**	.35**	.39*	.37**	.41**	.05	.01	.11*	.06	.06	.01	.00	.21**	.03	.05
Domin	.30**	.29**	38**	.39**	.40**	.44**	.10*	.07	.15*	.14**	.07	.07	.02	.29**	-.05	.16**
Initia	.32**	.29**	.38**	.40**	.38**	.41**	.05	.06	.12*	.17**	.07	.06	.00	.30**	-.03	.09
LT-Ang	.07	.11*	.09	.11**	.13**	.13**	.08	.06	.11*	.09	.12*	.13**	.16**	.15**	.03	.13**
LT-Sad	.12*	.10*	.05	.06	.12*	.11*	.25**	.16**	.15*	.18**	.08	.12**	.17**	.16**	-.09	.14**
Hostile	.07	.08	.05.	.03	.02	.09	.10*	.03	.09	.08	.07	.02	.00	.15**	.00	.09
Fear	.00	-.01	-.01	.04	.05	.08	.09	.04	.05	.04	.23**	.22**	.20**	-.03	.00	-.02
BI	-.15**	-.10*	-.23**	-.17**	-.16**	-.17**	.03	.08	-.09	.03	.13**	.15**	.14**	-.16**	-.01	-.08
Clingy	-.14**	-.08	-.20**	-.12**	-.10*	-.06	.14**	.14**	-.02	.04	.12**	.14**	.16**	-.05	-.08	.02
Impul	.21*	.19**	.20**	.22**	.23**	.33**	.10*	.08	.13**	.04	.03	.05	-.04	.25**	-.03	.12*
Comp	.00	-.04	-.02	-.06	-.13**	-.12*	-.13**	-.13**	-.12*	-.29**	-.10*	-.16**	-.16**	-.23**	.15**	-.14**
IC	.03	.07	.06	.05	.05	.12*	.20	.06	-.06	-.04	.03	.05	-.06	.03	-.01	.03

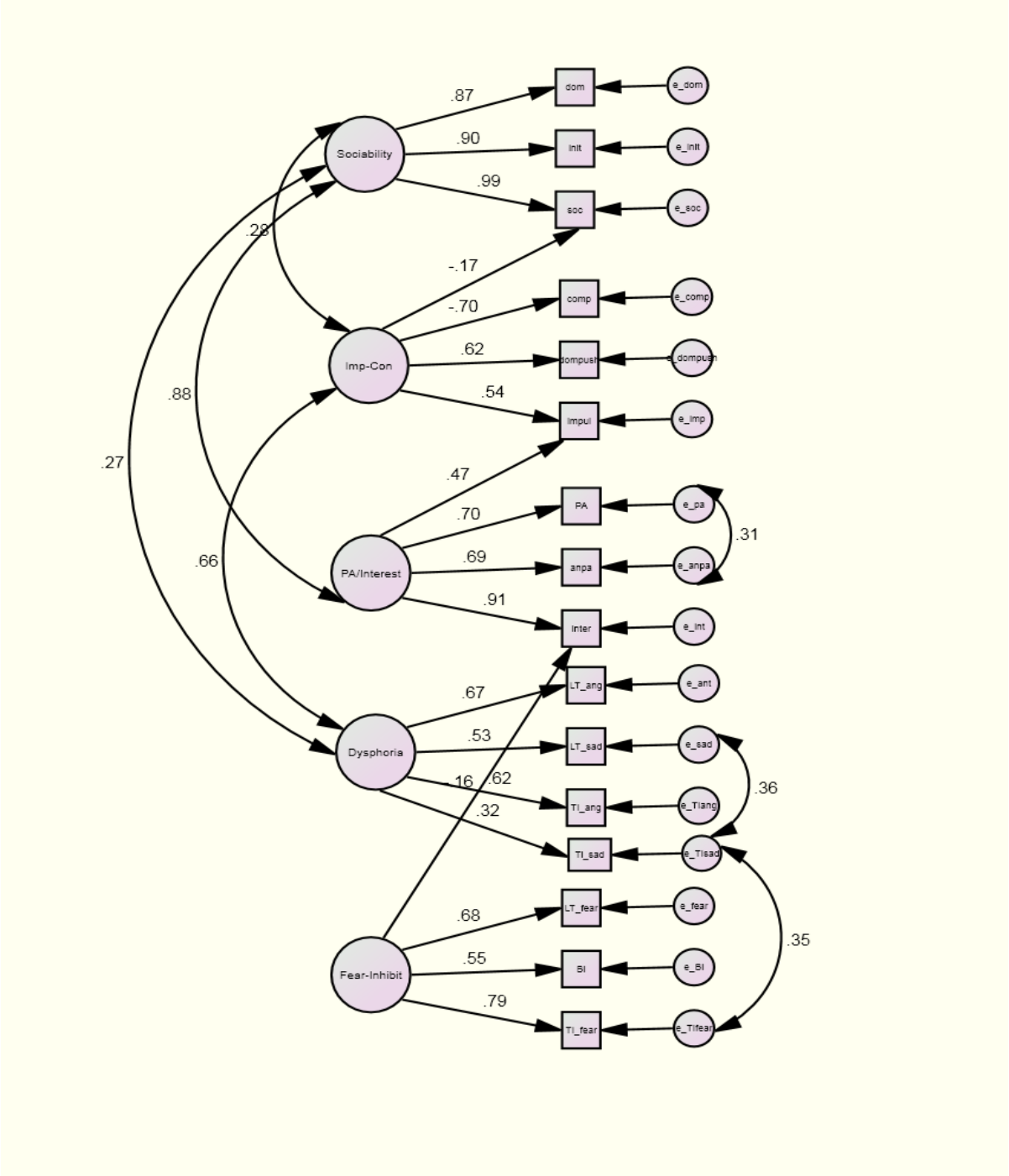
*p < .05, **p < .01, *** p < .001; PA =Positive Affect, AnPA = Anticipatory PA, Inter = Interest, Soc =Sociability, Domin = Dominance/surgency, Initia =initiative, LT-Sad = Lab-TAB Sadness, TI-Sad = Tester Impression Sadness, LT-Ang=Lab-TAB Anger, TI Ang = Tester Impression Anger, LT-Fear = Lab-TAB Fear, TI-Fear = Tester Impression Fear, BI =Behavioral Inhibition, Impul =Impulsivity, Comp = Compliance, Dom/Push = Domineering Pushy, IC =Inhibitory Control..

Figure 1. Original Age 3 CFA model with Age 6 sample



Note. Imp-Con=Impulsivity vs. Constraint Factor, dom=dominance, soc=sociability, init=initiative, comp=compliance, dompush=domineering/pushy, PA=positive affect, anpa=anticipatory PA, Inter=interest, LT_ang= Lab-TAB anger, LT_sad = Lab-TAB sad, TI_ang= Tester impression anger, TI_sad=Tester impression sad, LT_fear=Lab-TAB fear, BI=Behavioral Inhibition, TI_fear=Tester Impression fear, e=error

Figure 2. Modified Age 3 CFA model with Age 6 sample



Note. Imp-Con=Impulsivity vs. Constraint Factor, dom=dominance, soc=sociability, init=initiative, comp=compliance, dompush=domineering/pushy, PA=positive affect, anpa=anticipatory PA, Inter=interest, LT_ang= Lab-TAB anger, LT_sad = Lab-TAB sad, TI_ang= Tester impression anger, TI_sad=Tester impression sad, LT_fear=Lab-TAB fear, BI=Behavioral Inhibition, TI_fear=Tester Impression fear, e=error

Figure 3. Combined Age 3 and 6 model