HHS Public Access
Author manuscript
Psychol Assess. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:
Psychol Assess. 2015 December ; 27(4): 1388–1401. doi:10.1037/pas0000104.

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

Margaret W. Dyson,
Department of Psychiatry, University of California, San Diego

Thomas M. Olino,
Department of Psychology, Temple University

C. Emily Durbin,
Department of Psychology, Michigan State University

H. Hill Goldsmith,
Department of Psychology, University of Wisconsin-Madison

Sara J. Bufferd,
Department of Psychology, California State University San Marcos

Anna R. Miller, and
Adventist Rehabilitation Hospital of Maryland

Daniel N. Klein
Department of Psychology, Stony Brook University.

Abstract

It is generally assumed that temperament traits exhibit structural and rank-order stability over time. Most of the research on structural and rank-order stability has relied 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 rank-order stability of temperament traits in a community sample of young children (N = 447). Using structural equation modeling (SEM), we found that a similar five-factor structure consisting of the dimensions of Positive Affect/Interest, Sociability, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint provided an adequate fit to the data at both age 3 and 6 years, suggesting good structural stability. Moreover, all five latent factors exhibited significant, albeit modest, rank-order stability from age 3 to 6. In addition, there were significant heterotypic associations of age 3 Sociability with age 6 PA/Interest, and age 3 Impulsivity vs. Constraint with age 6 Fear/Inhibition.

Keywords

Temperament; children; stability; structural equation modeling

Correspondence concerning this article should be addressed to Margaret Dyson, Department of Psychiatry, University of California, San Diego, Child and Adolescent Services Research Center, 3020 Children's Way, MC 5033, San Diego, CA 92123. mdyson@ucsd.edu.

Individual differences in reactivity and regulation in young children have traditionally been studied within a temperament framework. Several research traditions have developed models of the structure of temperament traits (e.g., Thomas & Chess, 1977; Buss & Plomin, 1984; Rothbart, 1981). Most of these models concur that the structure of temperament traits in young children is multidimensional; however, disagreement exists about the number and nature of these primary trait dimensions (De Pauw, Mervielde, & Leeuwan, 2009; De Pauw & Mervielde, 2010). Despite debate regarding the structure and nature of traits, most models assume that temperament is fairly stable over time (Caspi & Roberts, 2005; Goldsmith et al., 1987; Pedlow, Sanson, Prior, & Oberklaid, 1993).

Traditional Models Child Temperament

In this section, we briefly review the three major multidimensional models of child temperament (i.e., Thomas & Chess, 1977; Buss & Plomin, 1984; Rothbart, 1981). While other influential models have been proposed (e.g., Kagan, Reznick, & Snidman, 1987), these have generally focused on a smaller number of traits. Thomas and Chess's (1977) model of temperament consists of 9 bipolar dimensions (i.e., sensory threshold, activity level, approach/withdrawal, distractibility, attention span/persistence, quality of mood, adaptability, rhythmicity, and intensity of reaction) and is posited to be stable from infancy to adulthood. Buss and Plomin (1984) proposed a psychobiological model comprised of the 4 dimensions of activity, impulsivity, sociability and emotionality. With a greater emphasis on attentional and self-regulatory mechanisms, Rothbart (1981) proposed a model comprising the dimensions negative affectivity, extraversion/surgency, and effortful control. All of these models were primarily derived from parent-reports using a deductive or “bottom up” approach based on salient dimensions of individual differences in young children or key mechanisms related to social and emotional behavior (e.g., self-regulation). Alternatively, some theorists have examined whether adult models of personality can be extended to depict temperament traits in childhood, as discussed below.

Common Dimensions of Child Temperament

The growing consensus about the structure of adult personality (i.e., Five Factor Model (FFM)) has encouraged developmental researchers to take a more inductive approach and develop a unified taxonomy for the structure of child temperament. Mervielde and Asendorpf (2000) identified shared themes in Thomas and Chess’ (1977), Buss and Plomin's (1984), and Rothbart's (1981) models and proposed that 4 dimensions encompass the key components of these 3 models: Extraversion, Emotionality, Activity, and Persistence (see De Pauw et al., 2009 and De Pauw & Mervielde, 2010). Caspi and Shiner (2006) also proposed a common taxonomy for temperament and personality, incorporating traits from the preschool years into adulthood. Their model consists of five higher-order dimensions using the FFM labels of Extraversion, Neuroticism, Conscientiousness, Agreeableness, and Openness to Experience. These higher-order trait dimensions encompass and serve to explain the covariation among the lower-order traits, which include many of the narrower trait dimensions from the different models of child temperament.

Drawing upon these theoretical models of temperament, we previously examined the structure of temperament traits in preschool-aged children using a laboratory observational measure. We hypothesized some broad similarities between our model and prior models of temperament based on parent-report measures. Using a two-stage factor analytic approach, we derived a five-factor model consisting of the higher order traits of Sociability, Positive Affect (PA)/Interest, Dysphoria, Fear/Inhibition, and Impulsivity versus Constraint. This model shared a number of similarities and some notable differences with previous models of temperament based on parent-report measures (De Pauw & Mervielde, 2010; see Dyson, Olino, Durbin, Goldsmith, & Klein, 2012 for a more detailed discussion of these similarities and differences). This model provided the empirical foundation for the present study.

Types of Stability in Young Children

In addition to the structure of temperament during early childhood, it is also important to consider the stability of these trait dimensions across time. An underlying assumption of most temperament models is that trait dimensions exhibit some stability over time. Developmental researchers propose at least four types of stability in longitudinal research (e.g., Caspi & Shiner, 2005; De Fruyt, Bartels, Leeuwen, Clereq, Decuyper, & Mervielde, 2006; Putnam, Rothbart, & Garstein, 2008): (a) *ipsative*, the degree to which the relative ordering of traits within an individual are preserved across time, (b) *mean-level*, the changes in the average trait level in the population, (c) *relative* or *rank order*, the degree to which the rank ordering of individuals on a trait is maintained over time, and (d) *structural*, the degree of continuity in the associations among traits across time. We assessed structural and relative stability using assessments in a large sample of young children who were assessed at age 3 and age 6. We could not evaluate ipsative stability because it requires more than two assessments for each child. We were also precluded from assessing mean-level stability because we used age-appropriate measures/tasks, which necessarily differed between assessment time points; hence, mean levels are confounded with differences in tasks.

Homotypic and Heterotypic Stability

When examining stability, homotypic and heterotypic stability should be differentiated. Extant research examining the rank-order stability of temperament in early childhood has primarily focused on homotypic stability, the stability of similar behaviors across time. However, maturation or normative development may impact the stability of temperament over time, especially in early childhood when changes are widespread and rapid (Caspi & Roberts, 2001; Putnam et al., 2008; Rothbart & Bates, 2006). As discussed below, lower rank-order 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, Hayden, Klein, & Olino, 2007; Roberts & DelVecchio, 2000; Rothbart & Bates, 2006) compared to later in life. Maturation may also influence the structural stability of temperament, as increasing coalescence or differentiation of traits may occur over the course of childhood (e.g., the number of traits may increase from infancy to preschool; Caspi, Roberts, & Shiner 2005; Eisenberg, 2000). In fact, several studies have found evidence of heterotypic stability of temperament dimensions in infancy and early childhood (e.g., Kagan, Snidman, & Arcus, 1998; Putnam et al., 2008). Thus, lower homotypic stability and greater heterotypic continuity may be expected in early childhood, compared with older ages, due to maturational or normative developmental changes (e.g., differential change in early childhood across children may create some instability in rank ordering of children on temperament traits). Furthermore, to assess traits consistently across time, measures must be adjusted to be sensitive to the child's developmental stage (e.g., Knight & Zerr, 2012; Putnam et al., 2008). Below, we discuss common methods of assessing temperament in early childhood, as well as the potential advantages and limitations of the approaches.

Methods of Assessing Temperament in Young Children

Parent-report questionnaires are the most common method of evaluating temperament in young children and have provided consistent support for the rank-order stability of childhood temperament (e.g., Pedlow, Sanson, Prior, & Oberklaid, 1993; Lemery, Goldsmith, Klinnert, & Mrazek, 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-reports are vulnerable to multiple perceptual and response biases, and thus represent a mixture of objective and subjective factors, including parents' personality, emotional state, psychopathology and desire to portray their child in a positive light (Durbin & Wilson, 2012; Mangelsdorf et al., 2000; Rothbart & Bates, 2006; Stifner, Willougby & Towe-Goodman, 2008). Rank-order stability estimates based on parent-report measures may also reflect the stability of parent expectations or schemas of temperament rather than the stability of the child's behavior (Durbin et al., 2007, Durbin & Wilson, 2012; Gagne, Hulle, Askan, Essex, & Goldsmith, 2011; Mangelsdorf et al., 2000; Saudino, 2003).

Laboratory-observational measures of child temperament have advantages relative to parent-reports. These measures allow the researcher to utilize standardized procedures to elicit specific behaviors and emotions and use objective criteria to code behavioral and emotional displays, circumventing the issue of parental interpretation (Durbin et al., 2007; Garstein, Bridgett, Rothbart, Roberston, Iddin, Ramsey et al., 2010; Majdandžić & van den Boom, 2007; Zeman, Klimes-Dougan, Cassano, & Adrian, 2007). 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 limitations. In addition to being expensive and time-intensive, they may be susceptible to 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, 2012).

Studies have consistently reported low associations between parent-report and laboratory-observational measures of temperament (e.g., Durbin et al., 2007; Goldsmith, Reiser-Danner, & Briggs, 1991; Majdandzic, & van den Boom, 2007; Saudino & Cherny, 2001; Stifter et al., 2008). However, both approaches show evidence of validity (Durbin et al., 2007; Mangelsdorf et al., 2000). Therefore, observational measures may provide different information about the structure of temperament than parent-report measures (Gagne et al., 2011), and generalization about the structural and rank-order stability of temperament from one method to the other might be infeasible.

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 included exploratory factor analysis (EFA) on groups of children of different ages. 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 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 EFA. 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. However, a limitation of these cross-sectional studies is that developmental and sample effects cannot be distinguished.

Longitudinal Studies of Structural and/or Rank-Order Stability of Temperament Based on Parents' or Observers' Reports

In comparison to cross-sectional designs, longitudinal designs, in which the same children are examined at several points across time, provide a more robust and sensitive approach to examining stability. Rank-order stability estimates of temperament traits in young children using parent-report measures generally fall in the moderate range (Roberts & DelVecchio, 2000). However, most of these studies have relied on bivarate cross-time correlations, which are vulnerable to attenuation due to measurement error (Roberts, Caspi, & Moffit, 2001). In a meta-analysis examining the rank-order stability of traits from infancy to late adulthood, Roberts and DelVecchio (2000) aggregated stability coefficients for temperament trait dimensions and adult personality traits. For the time interval from 3.0 to 5.9 years, they obtained a mean stability correlation of .52.

Finally, a few studies have used Confirmatory Factor Analysis (CFA) and structural equation modeling (SEM) within the context of longitudinal designs to examine both the structural and rank-order stability of parent reports of temperament in young children while accounting for measurement error. 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) and 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 (*rs* .44-.83). Also utilizing SEM, Lemery and colleagues (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 and found 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, stability was high (estimates in the .70s). In sum, the results of these studies indicate good structural stability and moderate rank-order stability over time, especially after the infancy period (Lemery et al., 1999; Pedlow et al., 1993).

Stability of Temperament Using Laboratory-Observational Measures

To our knowledge, no studies have examined the structural stability of multiple temperament traits in young children using only laboratory-observational measures. Although studies have used laboratory-observational measures (sometimes in conjunction with parent reports) to examine the rank-order stability of temperament, most have examined only a single or small number of traits, such as behavioral inhibition, fearfulness, and effortful control (Eisenberg, Edwards, Spinrad, Sallquist, Eggum & Reiser, 2013; Fox, Henderson, Rubin, Calkins, & Schmidt, 2001; Garstein, et al., 2010; Kagan, Snidman, Kahn, & Townsley, 2007; Kochanska, Murray, & Coy, 1997; Pfiefer, Goldsmith, Davidson, & Rickman, 2002).

Similarly, only a handful of studies have examined the rank-order stability of multiple temperament traits using laboratory-observational measures (Belsky, Hsieh, & Crinic, 1996; Carnicero, Perez-Lopez, Gonzalez-Salinas, & Martinez-Fuentes, 2000; Goldsmith & Campos, 1990; Rothbart, Derryberry, & Hershey, 2000). These studies generally report moderate stability for periods of up to a year. However, most of these study designs had only short intervals between assessments, providing upper-range stability estimates. In addition, most studies focused on infants and toddlers, so few data exist on stability of temperament in the preschool and early school age periods. Across the ages of 3, 5, and 7 years, Durbin and colleagues (2007) found moderate to high rank-order stability estimates for laboratory-assessed Positive Emotionality, which includes the lower-order traits of positive affect (PA; *rs* range from .59-.70), anticipatory PA (r =.41), sociability (*rs* range from .52-.62), interest/engagement (*rs* range from.37-.48), and activity level (r = .62), and Negative Emotionality, which includes the lower-order traits of negative affect (*rs* range from .57-.59), sadness (r = .52), anger (*rs* range from .30-.40) (all Pearson correlations). The lower-order trait of fear yielded lower stability estimates (*rs* range from .21-.23). (Stability estimates were available only from age 3 to 5 for anticipatory PA, activity level, and sadness). There was no evidence of heterotypic stability between Positive and Negative Emotionality.

We are aware of only one study that used SEM to account for measurement error in examining the rank-order stability of multiple temperament traits 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 positive emotionality/exuberance, interest, anger, and sadness based on a combination of laboratory observations and parent-report questionnaires. A structural model could not be fit for fear due to low intercorrelations between the fear episodes. SEM estimates (standardized betas) of rank-order stability (standardized betas) for the dimensions of exuberance/positive emotionality (.76), interest (1.00), and anger (.55) ranged from moderate to high. However, sadness (.34) did not demonstrate significant stability across time. As expected, rank-order 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 or heterotypic stability.

Rationale for the Present Study

There is a notable absence of studies examining the structural stability of temperament assessed with laboratory-observational measures. Moreover, only a limited number of studies addressed the rank-order stability of laboratory-assessed temperament traits, and most have focused on infants and single temperament traits, and used short time intervals between assessments. Considering the differences in approach and the low correlations between parent-report and laboratory-observational measures, a laboratory-observational measure might produce a different picture of the structural and rank-order stability of temperament in young children compared with parent-report measures.

As discussed above, we previously 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 et al., 2012). Our best-fitting model consisted of five-higher order dimensions, Sociability, Positive Affect/Interest, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint (see left side of Figure 2 below and Figure 2 in Dyson et al., 2012). We 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 applying the structural model of temperament derived at age 3 to the age 6 Lab-TAB data. We hypothesized that the age 3 model would be a good fit to our sample at age 6, providing evidence for structural stability. Our second aim was to estimate the rank-order homotypic stability of the temperament dimensions from age 3 to 6. Based on the research reviewed above (e.g., Durbin et al., 2007; Lemery et al., 1999; Majdandžić & van den Boom, 2007; Pedlow et al., 1993; Roberts & DelVecchio, 2000), we hypothesized moderate rank-order stability between our laboratory-assessed trait dimensions at ages 3 and 6. Additionally, given that early childhood represents a period of rapid development in which the manifestation of underlying traits may vary over time, we also examined the heterotypic stability of these traits across time.

Method

Participants and Demographics

The original sample consisted of 550 three-year old children who participated in a longitudinal study of temperamental emotionality. Participants were recruited through commercial mailing lists. Families with a three-year-old child 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. Following a detailed description of the study, written informed consent was obtained from all the families. The families were financially compensated for their participation

A total of 447 participants (54.0% male and 46.0 % female) had data for the age 3 and 6 laboratory assessments. Thus, analyses were based on these participants. The mean age of the children at the second assessment was 73.1 months (*SD* = 4.97). The sample was primarily White/European-American (89.2%). 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) years, 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.

Laboratory Assessment 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. The Lab-TAB provides standardized episodes designed to elicit a variety of emotions and behaviors (i.e., each construct is elicited by 3 to 4 episodes; see Appendix A for descriptions of Lab-TAB episodes used in this study), and can be scored at varying “grain” levels (e.g., global and micro-level coding) (Gagne, et al., 2011). Different episodes were used at the age 3 (12 episodes) and age 6 (9 episodes) assessments to ensure that they were developmentally age-appropriate (Knight & Zerr, 2012). 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), inhibitory control (age 3 only), and tester impression (age 6 only) 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. Coders were assigned to specific episodes and had to reach at least 80% agreement on all specific codes within the episode with an experienced rater before coding independently. To examine interrater reliability, videotapes were independently coded by a second rater. The median and ranges of alphas and interrater reliabilities for both age 3 and 6 variables are presented below for the affective, behavioral, behavioral inhibition, and inhibitory control variables.

Affective traits—Each instance 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 episodes. Within each episode, these intensity ratings were summed within each channel (facial, bodily, vocal) for each of the four affective traits. The intensity ratings were then averaged within each channel across all episodes, which resulted in scores for each of the three channels for each of the four affective traits. Each of these variables was then standardized. 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). The median and range for coefficient alpha for the PA, anger, sadness, and fear scales at the age 3 and 6 assessments were .75 (.63-.87) and .73 (.50-.83), respectively. The median and range of interrater ICCs (N=35) for PA, sadness, anger, and fear at the age 3 and 6 assessments were .77 (.64-.92) and .78 (.68-95), 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 was based on how engaged the child appeared in play. Anticipatory PA was based on PA that occurred in anticipation of a reward, reinforcer, or positive event. Initiative was based on the degree of passivity or assertiveness the child displayed in their interactions with others. Sociability was based on the child's attempts to engage and interact with the experimenter and the parent. Compliance 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 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): Domineering/Pushy was based on the degree to which the child made demands, was actively noncompliant, and argued with the experimenter or mother. Hostility was based on the degree to which the child directed physical or verbal aggression or angry comments at the experimenter or mother. Clinginess 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.

Dominance vs. submissiveness was rated on an 11-point Likert scale (−5 [extremely negative] to 5 [extremely positive]) because these traits are bivalent. Dominance was based on the degree of social potency demonstrated by the child in the interaction. High scores reflected dominant behavior, whereas negative scores indicated submissiveness and passivity.

Due to low frequency and reliability, clinginess and hostility were not used in the age 6 analyses. Potential explanations for the low frequency of these behaviors at age 6 is that negative emotionality may decrease across time, or it is developmentally appropriate for these behaviors to decrease over time as children learn to regulate their emotions. The median and range of coefficient alpha for all of the behavioral variables at the age 3 and 6 assessments were .71 (.60-83) and .70 (.48-.80), respectively. The interrater ICCs at the age 3 and 6 assessments were .79 (.51-.87) and .72 (.53-.84), respectively.

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). BI was coded from three episodes (Risk Room, Stranger Approach, and Exploring New Objects) at the age 3 assessment and two episodes (Story Telling and Object

Fear) at the age 6 assessment 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 variables at both age 3 and 6 did not contain any affective ratings from the designated episodes because the affective ratings were used to create the fear variable described above. The BI composite variables at age 3 (α =.74, interrater ICC = .90) and age 6 (α =.51, interrater ICC = .77) were constructed by combining the average standardized ratings of the following variables from the episodes: total number of objects touched, latency to touch objects, tentative play, referencing experimenter, time spent playing, latency to vocalize, approach towards the stranger, avoidance of the stranger, gaze aversion, and verbal/nonverbal interaction with the stranger. Variables were all keyed in a consistent direction (e.g., long latencies to touch objects were keyed to reflect more BI).

***Inhibitory control* (Age 3 Lab-TAB only)—**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 was constructed by aggregating the scores from the two episodes (α = .70, interrater ICC = .98). Inhibitory control was not assessed during the age 6 Lab-TAB and could not be included in our age 6 model.

***Tester impression variables* (Age 6 Lab-TAB only)—**Global ratings of anger, sadness, and fear, based on all of the episodes, were made by the experimenter at the conclusion of the laboratory assessment. These variables were rated on a single 5-point Likert scale (1=rarely, 2=subtle or ambiguous signs, 3=mild, 4=moderate, 5=extreme). Interrater reliabilities are not available for these variables, and coefficient alpha cannot be calculated because they are based on only one item. As clingy and hostility were not used as indicators in our age 6 model due to low frequency and reliability, the experimenter's ratings of sadness, anger, and fear were included in our age 6 model in order to have at least three indicators per factor.

Results

Data Normalization and Standardization

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

Deriving an Age 6 Structural Model of Temperament

Examining the fit of the age 3 model to the age 6 sample—We used CFA with maximum-likelihood estimation procedures in AMOS 18.0 to examine the fit of the same five-factor structure obtained in the age 3 sample (see Figure 2 in Dyson et al., 2012) with the age 6 sample (*N* = 447). To assess model fit, the following criteria were used: (a) chi-square statistic; (b) the root-mean-square error (RMSEA; Steiger & Lind, 1980); and (c) the comparative fit index (CFI; Bentler, 1990). Because 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 (2010) recommended using cutoff values of RMSEA < .10 and CFI > .90 for acceptable model fit. We used these cutoffs in our previous report (Dyson et al., 2012). The target model (based on the model at age 3) consisted of five-factors, PA/Interest, Sociability, Dysphoria, Fear/Inhibition, and Impulsivity vs. Constraint, and 16 indicator variables. The age 6 versions of the same indicators used by Dyson et al. (2012) were employed, with three 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. According to the Hopwood and Donnellan (2010) criteria, this model fell just shy of an acceptable fit to the age 6 data, χ^2 (93) = 491. 90, RMSEA = .098 (90% CI = .090-.107), and a CFI of .887.

Post hoc model fitting—To improve the age 6 CFA model fit, we iteratively examined the model results. First, we examined model estimates to identify non-significant paths to be trimmed from the model to increase parsimony. Second, after re-estimating the model, we examined modification indices (MIs) to identify whether any additional methodological influences should be included to improve model fit. The model estimates indicated that the correlated paths between the latent factors of PA/Interest and Fear/Inhibition (r = .03, p = .39), PA/Interest and Impulsivity vs. Constraint (r = .03, p = .34), and the residuals of interest and initiative (r = .02, p =.17) were non-significant; hence, we removed them from the model in the interest of parsimony. Based on the MIs, we made two post hoc modifications on methodological grounds (see Figure 1). 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 by the same rater using the same global rating scale. Second, the tester impression sadness and Lab-TAB sadness indicators were correlated based on the rationale that both variables tap sad affect in the same set of laboratory episodes. Fit of the revised model was good, with χ^2 (94) = 401.05, RMSEA = .086 (90% CI = .077-.094), and CFI =.914 (see Figure 1).

The age 3 and age 6 models both consisted 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 (.26), Sociability and Fear/Inhibition (.14), Sociability and Dysphoria (.25), and the Impulsivity vs. Constraint and Dysphoria factors (.66). Unlike the age 3 model, the covariance paths between the initiative and interest residuals, PA/Interest and Fear/Inhibition, and PA/Interest and Impulsivity vs. Constraint were non-significant and removed from the model. Ordinarily, the next step would be to formally test structural invariance between the age 3 and 6 models by 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 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).

Rank-order stability of factors from age 3 to 6—The cross-time bivariate correlations between the age 3 and 6 indicators/lower-order traits are presented in Table 1. 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 overall fit of the model for the age 3 and age 6 assessments, as well as the rank-order stability of the latent factors from ages 3 to 6. The combined age 3 and 6 model is depicted in Figure 2; fit indices were χ^2 (406) = 1188.09, CFI = .893, and RMSEA = .066 (90% CI = .061-.070). As shown in Figure 2, all five latent factors demonstrated significant homotypic stability from age 3 to 6 with estimates ranging from .19 to .49. In addition, results suggest some heterotypic stability among some of the latent factors. Specifically, the latent age 3 Sociability factor predicted the latent age 6 PA/ Interest factor (.37), and the latent age 3 Impulsivity vs. Constraint factor was associated with the latent age 6 Fear/Inhibition factor (.19).

Discussion

Most research on the structural and rank-order stability of temperament traits in young children has relied on parent-reports. We extended this literature by using a laboratory-observational measure. Below, we discuss those key findings related to structural and rank-order stability in the context of relevant literature.

Structural Stability of Temperament Traits from Age 3 to 6

We used a longitudinal design and CFA to examine the fit of our age 3 five-factor model (Dyson et al., 2012) at age 6. With a few modifications, this model was a relatively good fit to the age 6 data and consisted of the same five-factors (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 appropriateness of a few indicators, this finding supports our initial hypothesis that the age 3 and age 6 models would be comprised of similar higher-order dimensions and suggests there is a relatively high level of structural stability from age 3 to 6.

Similar to the age 3 model, the age 6 model included two higher-order factors, PA/Interest and Sociability, that fall under the broad dimension of extraversion. More specifically, at both ages, the PA/Interest factor included the emotional core 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, at both ages, impulsivity, deemed an essential facet of extraversion in several theoretical models (e.g., Depue &

Collins, 1999; Eysenck & Eysenck, 1985), loaded (moderately) on the PA/Interest factor. Furthermore, this relationship with the PA/Interest and impulsivity is consistent with research demonstrating strong approach-based positivity predicting high impulsivity and low constraint in early childhood (Rothbart et al., 2000).

The age 3 and 6 Sociability factors were both composed of the traits of sociability, dominance, and initiative, which are characterized as interpersonally surgent traits (i.e., traits related to agency) in some models of extraversion (Depue & Collins, 1999). Similar to the age 3 model, all three indicators loaded highly on the age 6 Sociability factor, with the sociability indicator demonstrating the highest loading. Although the PA/Interest and Sociability factors remained distinct and exhibited significant stability from age 3 to 6, these traits exhibited a strong correlation (.88) at age 6. Thus, one potential hypothesis for this strong association is that both latent factors are composed of moderately correlated (see Table 1) reward-seeking/agentic traits (i.e., interest, initiative, and dominance) described in some neural, approach-based models of extraversion (e.g., Behavior Activation System (BAS) or Behavioral Facilitation System (BFS; Depue & Collins, 1999; Gray, 1970), in which the mesolimbic dopamine system is believed to play a major role (Depue & Collins, 1999; Whiteside, Lamm, Helfinstein, & Fox, 2012).

Based on the association between PA/Interest and Sociability at age 6, we tested a structural model that included a combined PA/Interest-Sociability factor (i.e., a four-factor model). 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. The four-factor model (BIC = 16121.10, AIC = 15887.25) did not fit the data as well as the five-factor model with distinct PA/Interest and Sociability factors (BIC = 15048.05, AIC = 15810.01).

Both the age 3 and age 6 models contained two higher-order factors, Dysphoria and Fear/Inhibition, that fall under the rubric of neuroticism/negative emotionality. At both age 3 and age 6, all of the indicators of the age 6 Dysphoria factor had moderate to high loadings, with the anger indicators demonstrating the highest loadings followed by the sadness indicators. Again, similar to the age 3 model, at age 6 all indicators of the Fear/Inhibition factor had moderate to high loadings. 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, indicating that fearfulness/inhibition at age 6 is related to low approach and interest in the laboratory setting.

The emergence of separate Dysphoria and Fear/Inhibition 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 unpredictable threat or the anticipation of punishment (Lewis & Ramsey, 2005; Putnam, Ellis, Rothbart, 2005). These findings are also consistent with evidence indicating that fearful/anxious (internalizing) distress and irritable, hostile (externalizing) distress follow unique developmental trajectories, predict different outcomes, and often load on different factors at many ages (e.g., Caspi et al., 2005; Rothbart & Bates, 2006). Moreover, this distinction is consistent with biobehavioral models positing that fearful and inhibited behavior is associated with the Behavioral Inhibition System (BIS) (e.g., hyperresponsive amygdala) (Gray, 1991), which is regulated, at least in part, by the serotonin system, whereas anger/ irritability is linked to the BAS, whose input is largely dopaminergic (Rothbart & Posner, 2006).

Both the age 3 and 6 models included factors characterized as Impulsivity vs. Constraint. All of the indicators exhibited moderate to high loadings on this factor at both age 3 and 6, with compliance exhibiting the highest loading. Additionally, the Impulsivity vs. Constraint factor was negatively associated with sociability at both ages 3 and 6, which is consistent with research suggesting that good behavioral control and constraint and lower impulsivity are related to socially appropriate behavior and successful interpersonal interactions (e.g., Eisenberg et al., 2000, 2004). Substantial association existed between the Impulsivity vs. Constraint and the Dysphoria factors in both models. This association is consistent with research demonstrating links between regulatory capacity behaviors (e.g., inhibitory or effortful control) and negative emotions (Bridgett, Garstein, Putnam, McKay, Iddins, Robertson et al., 2009; Bridgett, Oddi, Laake, Murdock, & Bachmann, 2013; Kochanska, Coy, Tjebkes & Husarek, 1998; Putnam et al., 2008; Rothbart & Derryberry, 2002). For instance, deficits in inhibitory control and impulsivity may contribute to poor negative emotion regulation or high negative emotionality may impact the development of healthy regulatory behaviors. Furthermore, the strong association between these factors aligns with neurobiological conceptualizations that regulatory behaviors in early childhood are driven more by the bottom-up, emotion-generating processes of the well-developed limbic system, rather than top-down processes associated with the under-developed prefrontal cortex (e.g., Bridgett et al., 2009; Oscher & Gross, 2007).

Rank-order Stability of Traits from Age 3 to 6

Only a few studies (Belsky et al., 1996; Durbin et al., 2007; Goldsmith & Campos, 1990; Majdandzic & van den Boom, 2007) have examined the rank-order stability of multiple laboratory-assessed traits within the same study. Moreover, most of these studies used cross-time bivariate correlations to estimate stability, which are susceptible to measurement error, had brief intervals between assessment periods, and used infants and toddlers. The present study utilized SEM to examine the rank-order stability of laboratory-assessed temperament traits over a three-year interval from ages 3 to 6. Furthermore, unlike previous studies, we examined the homotypic and heterotypic stability of these traits within the same structural model.

All five of the latent factors exhibited significant rank-order homotypic stability from age 3 to 6. As shown in Figure 2, the latent factors of Sociability (.49), Dysphoria (.30), and Impulsivity vs. Constraint (.32) exhibited moderate stability, whereas the latent factors of PA/Interest (.19) and Fear/Inhibition (.24) exhibited lower stability.

Our findings have some notable similarities and differences to previous studies examining the stability of traits based on laboratory-observational measures. For instance, 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 rank-order stability for sociability, moderate stability for the Negative Emotionality traits of sadness and anger, and low stability for fear across the ages of 3, 5, and 7 years. Compared with Durbin et al.'s (2007) moderate-to-high stability estimates for the Positive Emotionality traits of PA, anticipatory PA, and interest, our cross-time correlations (see Table 1) for the lower-order traits of PA/Interest, PA, anticipatory PA, and interest, demonstrated low-to-moderate stability. In our latent variable analyses, the homotypic stability between the higher-order age 3 and 6 PA/Interest factors was lower. However, as discussed in more detail below, this may be due in part to the heterotypic path from age 3 Sociability to age 6 PA/Interest.

In the only study to date to use SEM to examine the rank-order stability of traits in young children (i.e., four year-old children), Majdandžić and van den Boom (2007) reported higher stability estimates for their interest and positive emotionality/exuberance variables than we observed for our PA/Interest factor. In addition to the heterotypic path from age 3 Sociability and age 6 PA/Interest, our lower estimates may be due to the longer time span (i.e., three years compared to seven months) between assessments. Where our Dysphoria factor was relatively stable, Majdandžić and van den Boom reported non-significant stability for sadness and fear, and moderate stability for anger; however, these differences may be partially attributable to our larger sample size.

Additionally, several significant heterotypic associations emerged between different latent factors from age 3 to 6. Others have also reported heterotypic stability of temperament dimensions in young children (e.g., Kagan et al., 1998; Putnam et al., 2008), which may reflect normal developmental maturation. There was a moderate (.37) link between the latent age 3 Sociability and age 6 PA/Interest factors, indicating that preschool sociability is associated with greater PA at age 6. This association may contribute to the lower homotypic stability between the age 3 and 6 PA/Interest factors and is consistent with experimental studies demonstrating that extraverted (i.e., talkative/sociable, dominant) behavior increases positive affect (McNiel & Fleeson, 2006).

The latent age 3 Impulsivity vs. Constraint factor was significantly associated with the latent age 6 Fear/Inhibition factor (.19). Based on bivariate cross-time correlations of the indicators (Table 1), compliance may be strongly contributing to this effect, as noncompliance at age 3 is associated with fearfulness and inhibition at age 6. One possible interpretation of this finding is that fearfulness, inhibition, and/or anxiety may manifest as noncompliance at age 3 but evolve into more typical expressions of fearfulness and inhibition at age 6. This finding is consistent with literature emphasizing age-specific expressions of anxiety and may also account for the lower stability estimates we found for our Fear/Inhibition factor from age 3 to 6. 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 (Kearney & Silverman, 1995). Instead, fear and anxiety in young preschoolers may be expressed through noncompliance, refusal to engage or play, outbursts/tantrums, and/or other inappropriate, disruptive behaviors (Albano, Chorpita, & Barlow, 2003; Pincus, Eyberg, & Choate, 2005). Furthermore, these behavioral displays may be more likely in the context of a laboratory setting as some children may find the novel setting stressful.

Limitations and Future Directions

Several factors should be considered in interpreting our results. First, the use of CFA/SEM with temperament and personality data has proven highly challenging due to frequent secondary loadings and cross-loadings across items (Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1994). As a result, it has been argued that less stringent fit criteria are appropriate in this area (Hopwood & Donnellan, 2010). After making several methodologically and theoretically meaningful post-hoc modifications, the age 6 rank-order stability model had a minimally acceptable CFI (.893), and the other fit indices showed an acceptable fit. Thus, it is notable that we were able to derive relatively good-fitting models given the low success rate of prior CFA/SEM studies in this domain.

Second, the goal for the present study was to assess structural and rank-order of stability in those participants who completed both the age 3 and age 6 laboratory assessments (i.e., the 447 participants who returned for the laboratory assessment at age 6). We considered deriving the age 6 factor structure taking the full age 3 sample into account (e.g., by running the originally specified model for the age 6 data incorporating the predicted paths from the age 3 data). However, we designed our current approach to avoid biasing similarities in structure by estimating the age 6 model initially independent of the age 3 data. In addition, there were no systematic differences between participants and non-participants at age 6 on age 3 variables of interest. Thus, we judged that it was not necessary to use age 3 variables to correct for attrition biases in the age 6 assessment. Nonetheless, we conducted secondary analyses using the complete age 3 *N* of 550 and obtained similar results (See Footnote 1).

Several limitations pertain to the laboratory tasks and indicator variables that we used. First, the Lab-TAB tasks were developed to elicit specific behaviors and emotions, which increases the chance of observing relevant responses. However, this design feature raises the question of whether the child's emotion or behavior is attributable to situation-specific, rather than trait, influences. In an attempt to minimize situation-specific variance, we averaged variables across episodes before including them in the analyses. Second, we were unable to formally test for measurement and structural invariance because we did not have identical indicators at both assessment occasions, which would have provided a more stringent test of the structural stability of temperament. Third, the use of different tasks and indicators at both occasions to assess the same latent factors might have led to an underestimation of the homotypic temporal associations between latent factors. Future studies should strive to utilize the same indicators at each time point for more rigorous tests of structural stability. However, this will be challenging, as some tasks and indicators are much more appropriate at some ages than others. Lastly, the internal consistency and interrater reliability of some of our indicator variables was only moderate, and reliability data were not available for several indicators (i.e., tester impression indicators). Nonetheless, the use of multiple indicators and latent variable analyses should mitigate the suboptimal reliability of specific indicators.

Consistent with most research on the stability of temperament in young children, we found modest-to-moderate stability for traits across time. In addition to the use of several different indicators at each occasion, another potential explanation for these modest temporal associations is that neurodevelopmental changes between ages 3 and 6 may alter the expression of emotions and behavior. For example, Rothbart and colleagues suggest that the maturation of the prefrontal cortex and the development of self-regulatory processes may impact stability because these processes modulate children's initial reactive responses (Posner & Rothbart, 2007; Rothbart & Bates, 2006). If early childhood represents a period of “plasticity,” this period may provide critical opportunities for prevention and early intervention. Thus, young children may be more responsive to environmental influences, such as fostering adaptive coping strategies or parenting interventions, that reduce susceptibility to later developing emotional and behavioral problems (Klein, Dyson, Kujawa, & Kotav, 2012). Thus, further research examining the influence of environmental factors on developmental trajectories of temperament traits in early childhood, and their implications for risk and resilience is warranted (Bridgett et al., 2009; Lengua & Wachs, 2012; Partridge & Learer, 2007).

Finally, several additional limitations should be considered with regard to the characteristics of our sample. Specifically, the participants in our sample were predominantly White/ European American and middle class. Although the sample was representative of the population in our geographic region, this may constrain the generalizability of our findings. Future studies should examine the structural and rank-order stability of temperament in young children utilizing a more ethnically and economically diverse sample. Additionally, we did not examine gender differences in the structural and rank-order stability of laboratory-assessed temperament traits in this sample. We plan to address these differences in future work (see Olino, Durbin, Klein, Hayden, & Dyson, 2013 for data on gender differences in the wave of assessments at age 3).

In conclusion, this study extended previous work by using laboratory observations as an alternative approach to parent-reports to examine the structural and rank-order stability of temperament in young children. Using SEM, we found that the same five-factor structure that fit the data when our sample was 3 years old continued to provide a good fit at age 6. In addition, all five of the latent factors demonstrated significant, albeit modest, rank-order stability between ages 3 and 6. Finally, heterotypic associations were found between two pairs of different latent factors (Sociability with later PA/Interest and Impulsivity vs. Constraint with later Fear/Inhibition).

Acknowledgments

This research was supported by the following grants: NIMH RO1 MH069942 (DNK) and a GCRC Grant no. M01-RR10710 to Stony Brook University from the National Center for Research Resources. HHG was supported by R01-MH101504 and P50-MH100031, with infrastructure support from P30-HD03352.

Appendix A

Laboratory-Temperament Assessment Battery (Lab-TAB) Episodes

Age 3 Lab-TAB Episodes

Risk Room

The child was left alone to explore a set of novel and ambiguous stimuli (e.g., large black box with eyes and teeth, a Halloween mask, small staircase).

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.

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 (e.g., pretend mice in a cage, sticky water-filled gel balls, and mechanical spider). 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

Author Manuscript

Author Manuscript

Author Manuscript

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. The experimenter returned with several small toys for the child to keep.

Age 6 Lab-TAB Episodes

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 and returned after 3 minutes. 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 using a picture book to an unfamiliar research assistant, whom the experimenter described as a “story expert.” 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. 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.

Object Fear

The child was instructed to explore a room that was filled with fear-eliciting objects (e.g., box filled with plastic insects and from which cricket sounds were emitted, a large, black spider covered with a cloth).

References

Albano, AM.; Chorpita, BF.; Barlow, D. Childhood anxiety disorders.. In: March, Eric J.; Barkley, Russell A., editors. Child Psychopathology. 2nd. Guilford Press; New York: 2003. p. 279-329.

Belsky J, Hsieh KH, Crnic K. Infant positive and negative emotionality: One dimension or two? Developmental Psychology. 1996; 32:289–298.d.

Bridgett DJ, Garstein MA, Putnam SP, McKay T, Iddins E, Robertson C, et al. Maternal and contextual influences and the effect of temperament development during infancy on parenting in toddlerhood. Infant Behavior & Development. 2009; 32:103–136. [PubMed: 19111913]

Bridgett DJ, Oddi KB, Laake LM, Murdock KW, Bachmann MN. Integrating and differentiating aspects of self-regulation: Effortful control, executive functioning, and links to negative affectivity. Emotion. 2013; 13:47–63. [PubMed: 22906086]

Caspi A, Roberts BW. Personality development across the life course: The argument for change and continuity. Psychological Inquiry. 2001; 12:49–66.

Caspi A, Roberts BW, Shiner RL. Personality development: Stability and change. Annual Review of Psychology. 2005; 56:453–84.

Caspi, A.; Shiner, RL. Personality development.. In: Damon, W.; Lerner, R.; Eisenberg, N., editors. Handbook of child psychology: Vol. 3. Social, emotional, and personality development. 6th ed.. Wiley; New York: 2006. p. 300-364.

Carlson S. Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology. 2005; 28:595–616. [PubMed: 16144429]

Carlson SM, Wang TS. Inhibitory control and emotion regulation in preschool age children’s theory of mind. Child Development. 2007; 72:1032–1053. [PubMed: 11480933]

Carnicero JAC, Pérez-López, Del Carmen González Salina M, Martínez-Fuentes MT. A longitudinal study of temperament in infancy: Stability and convergence of measures. European Journal of Personality. 2000; 14:21–37.

Church AT, Burke PJ Jr. Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. Personality Processes and Individual Differences. 1994; 66:93–114.

De Fruyt F, Bartels M, Van Leeuwan KG, De Clercq B, Mervielde I. Five types of personality continuity in childhood and adolescence. Journal of Personality and Social Psychology. 2006; 91:538–552. [PubMed: 16938036]

De Pauw SSW, Mervielde I. Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. Child Psychiatry and Human Development. 2010; 41:313–329. [PubMed: 20238477]

De Pauw SSW, Mervielde I, Van Leeuwen KG. How are traits related to problem behavior in preschoolers? Similarities and contrasts between temperament and personality. Journal of Abnormal Child Psychology. 2009; 37:309–325. [PubMed: 19165590]

Depue RA, Collins PF. Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extra-version. Behavioral and Brain Sciences. 1999; 22:491–569. [PubMed: 11301519]

Durbin CE, Hayden EP, Klein DK, Olino TM. Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. Emotion. 2007; 7:188–199.

Durbin CE, Wilson S. Convergent validity of and bias in maternal reports of child emotion. Psychological assessment. 2012; 24(3):647. [PubMed: 22149326]

Dyson MW, Olino TM, Durbin CE, Goldsmith HH, Klein D. The structure of temperament in preschoolers: A two-stage factor analytic approach. Emotion. 2012; 12:44–57. [PubMed: 21859196]

Eisenberg N, Edwards A, Spinrad TL, Sallquist J, Eggum ND, Reiser M. Are effortful and reactive control unique constructs in young children? Developmental Psychology. 2013; 49:2082–2094. [PubMed: 23398553]

Eisenberg N, Fabes RA, Guthrie IK, Reiser M. Dispositional emotionality and regulation: Their role in predicting quality of social functioning. Journal of Personality and Social Psychology. 2000; 78:136–157. [PubMed: 10653511]

Eisenberg N, Spinard TL, Fabes RA, Reiser M, Cumberland SA, Valiente C, Thompson M. The relations of effortful control and impulsivity to children's resiliency and adjustment. Child Development. 2004; 75:25–46. [PubMed: 15015673]

Fox NA, Henderson HA, Rubin KH, Calkins SD, Schmidt LA. Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. Child Development. 2001; 72:1–21. [PubMed: 11280472]

Gagne JR, Van Hulle CA, Aksan N, Essex MJ, Goldsmith HH. Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. Psychological Assessment. 2011; 23:337–353. [PubMed: 21480723]

Garstein MA, Bridgett DJ, Rothbart MK, Robertson C, Iddins E, Ramsey K, et al. A latent growth examination of fear development in infancy: Attributions of Maternal Depression and the risk for toddler anxiety. Developmental Psychology. 2010; 3:651–668.

Goldsmith HH, Campos JJ. The structure of infant temperamental dispositions to experience fear and pleasure: A psychometric perspective. Child Development. 1990; 61:1944–1964. [PubMed: 2083507]

Goldsmith, HH.; Gagne, JR. Behavioral assessment in temperament.. In: Zentner, M.; Shiner, R., editors. Handbook of temperament. Guildford; New York: 2012. p. 209-228.

Goldsmith, HH.; Reilly, J.; Lemery, KS.; Longley, S.; Prescott, A. Technical report. Department of Psychology, University of Wisconsin-Madison; 1995. Laboratory Temperament Assessment Battery: Preschool Version..

Goldsmith HH, Rieser-Danner, Loretta A, Briggs Stephen. Evaluating convergent and discriminant validity of temperament questionnaires for preschoolers, toddlers, and infants. Developmental Psychology. 1991:566–579.

Gray JA. The psychophysiological basis of introversion-extraversion. Behavior Research and Therapy. 1970; 8:249–266.

Halverson CF, Havil VL, Deal J, Baker SR, Victor JB, Pavopoulos V, et al. Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences. Journal of Personality. 2003; 71:995–1026. [PubMed: 14633056]

Hollingshead, AB. Four factor index of social status. Department of Sociology, Yale University; New Haven, CT.: 1975. Unpublished manuscript

Hopwood CJ, Donnellan MB. How should the internal structure of personality inventories be evaluated? Personality and Social Psychology Review. 2010; 14:332–346. [PubMed: 20435808]

Kagan, J. Biology and the child.. In: Damon, W.; Eisenberg, N., editors. Handbook of child psychology: Vol. 3. Social, emotional and personality development. 5th ed.. Wiley; New York: 1998. p. 177-235.

Kagan J, Reznick JS, Clarke C, Snidman N, Garcia-Coll C. Behavioral inhibition to the unfamiliar. Child Development. 1984; 55:2212–222.

Kagan J, Reznick JS, Snidman N. The physiology and psychology of behavioral inhibition in children. Child Development. 1987; 58:1459–1473. [PubMed: 3691195]

Kagan J, Snidman N, Arcus D. Childhood derivatives of reactivity in infancy. Child Development. 1998; 69:1483–1493. [PubMed: 9914634]

Kagan J, Snidman N, Kahn V, Towsley S. The preservation of two infant temperaments into adolescence. Monographs of the Society for Research in Child Development. 2007; 72:vii–vii. 1–95.

Kearney, CA.; Silverman, WK. Anxiety disorders.. In: Van Hasselt, VB.; Hersen, M., editors. Handbook of adolescent psychopathology: A guide to diagnosis and treatment. Lexington; Boston, MA: 1995. p. 435-464.

Klein, DN.; Dyson, MW.; Kujawa, AJ.; Kotov, R. Temperament and internalizing disorders.. In: Zentner, M.; Shiner, R., editors. The Handbook of Temperament. Guilford Press; New York: 2012.

Knight GP, Zerr AA. Informed theory and measurement equivalence in child development research. Child Development Perspectives. 2010; 4(1):25–30.

Kochanska G, Coy KC, Tjebkes TL, Husarek SJ. Individual differences in emotionality in infancy. Child Development. 1998; 64:375–390. [PubMed: 9586213]

Kochanska G, Murray K, Coy KC. Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. Child Development. 1997; 2:263–277. [PubMed: 9180001]

Lemery KA, Goldsmith HH, Klinnert MD, Mrazek DA. Developmental models of infant and child temperament. Developmental Psychology. 1999; 35:189–204. [PubMed: 9923474]

Lengua, LJ.; Wachs, TD. Temperament and Risk: Resilient and Vulnerable Responses to Adversity.. In: Zentner, M.; Shiner, R., editors. The Handbook of Temperament. Guilford Press; New York: 2012.

Lewis M, Ramsey D. Infant emotional and cortisol responses to goal blockage. Child Development. 2005; 76:518–530. [PubMed: 15784097]

Little, RJA.; Rubin, DB. Statistical analysis with missing data. Vol. 4. John Wiley & Sons, Inc.; Hoboken, NJ: 2002.

Majdandzic M, van den Boom DC. Multimethod longitudinal assessment of temperament in early childhood. Journal of Personality. 2007; 75:121–167. [PubMed: 17214594]

Mangelsdorf, SC.; Schoppe, SJ.; Buur, H. The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood.. In: Molfese, VJ.; Molfese, DL., editors. Temperament and personality development across the life span. Lawrence Erlbaum Associates; Mahwah, NJ: 2000. p. 121-140.

Marsh HW, Hau KT, Zonlin W. 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. 2004; 11:320–341.

Marsh HW, Lüdtke O, Muthén B, Asparouhov T, Morin AJ, Trautwein U, Nagengast B. A new look at the Big Five factor structure through exploratory structural equation modeling. Psychological Assessment. 2010; 22:471–491. [PubMed: 20822261]

Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, Trautwein U. Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluation of university teaching. Structural Equation Modeling. 2009; 16:439–476.

McCrae RR, Zonderman AB, Costa PT, Bond M, Paunonen SV. Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus procustes rotation. Journal of Personality and Social Psychology. 1996; 70:352–566.

McNeil JM, Fleeson W. 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. 2006; 40:529–550.

Mervielde I, Asendorpf JB. Hampson SE. Variable-centered versus person-centered approaches to childhood personality. Advances in personality psychology. Taylor & Francis, Philadelphia. 2000:37–76.

Ochsner, K.; Gross, J. Handbook of emotion regulation. Guilford Press; New York, NY US: 2007. The neural architecture of emotion regulation.; p. 87-109.

Olino TM, Durbin EC, Klein DN, Hayden EP, Dyson MW. Gender differences in young children's temperament traits: An observational study. Journal of Personality. 2013; 81:119–129. [PubMed: 22924826]

Partridge T, Lerner LV. A latent growth-cure approach to infant temperament. Infant and Child Development. 2007; 16:255–265.

Pedlow R, Sanson A, Prior M, Oberklaid F. Stability of maternally reported temperament from infancy to 8 years. Developmental Psychology. 1993; 29:998–1007.

Pfiefer M, Goldsmith HH, Davidson R, J. Rickman M. Continuity and change in inhibited and uninhibited children. Child Development. 2002; 73:1474–1485. [PubMed: 12361313]

Pincus DB, Eyberg SM, Choate ML. Adapting parent-child interaction therapy for young children with separation anxiety disorder. Education and Treatment of Children. 2005; 28:163–181.

Posner, MI.; Rothbart, MK. Educating the human brain. American Psychological Association; Washington, DC: 2007.

Putnam, SP.; Ellis, LK.; Rothbart, MK. The structure of temperament from infancy through adolescence.. In: Eliaz, A.; Angleitner, A., editors. Advances in research in temperament. Pabst Science Publishers; Lengerich, Germany: 2005.

Putnam SP, Rothbart MK, Gartstein MA. Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. Infant and Child Development. 2008; 17:387–405.

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

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

Rothbart MK. Measurement of temperament in infancy. Child Development. 1981; 52:569–578.

Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of temperament at three to seven years: The children's behavior questionnaire. Child Development. 2001; 72:1394–1408. [PubMed: 11699677]

Rothbart, MK.; Bates, JE. Temperament in children's development.. In: Damon, W.; Lerner, R.; Eisenberg, editors. Handbook of child psychology, Sixth edition: Social, emotional, and personality development. Vol. 3. Wiley; New York: 2006. p. 1-99.

Rothbart, MK.; Derryberry, D.; Hershey, K. Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years.. In: Molfese, VJ.; Molfese, DL., editors. Temperament and personality development across the life span. Erlbaum; Hillsdale, NJ: 2000. p. 85-119.

Rothbart, MK.; Posner, MI. Temperament, attention, and developmental psychopathology. In Handbook of Developmental Psychopathology, Revised. In: Cicchetti, D.; Cohen, DJ., editors. Wiley; New York: 2006. p. 167-88.

Saudino K, J. Parent ratings of infant temperament - Lessons from twin studies. Infant Behavior and Development. 2003; 26:314–325.

Stifter CA, Willoughby MT, Towe-Goodman N. Agree or agree to disagree? Assessing the convergence between parents and observers on infant temperament. Infant and Child Development. 2008; 17:407–426. [PubMed: 19936035]

Thomas, A.; Chess, S. Temperament and development. Bruner/Mazel; New York, NY: 1977.

Whiteside SP, Lynam DR. The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. Personality and Individual Differences. 2001; 30:669–689.

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

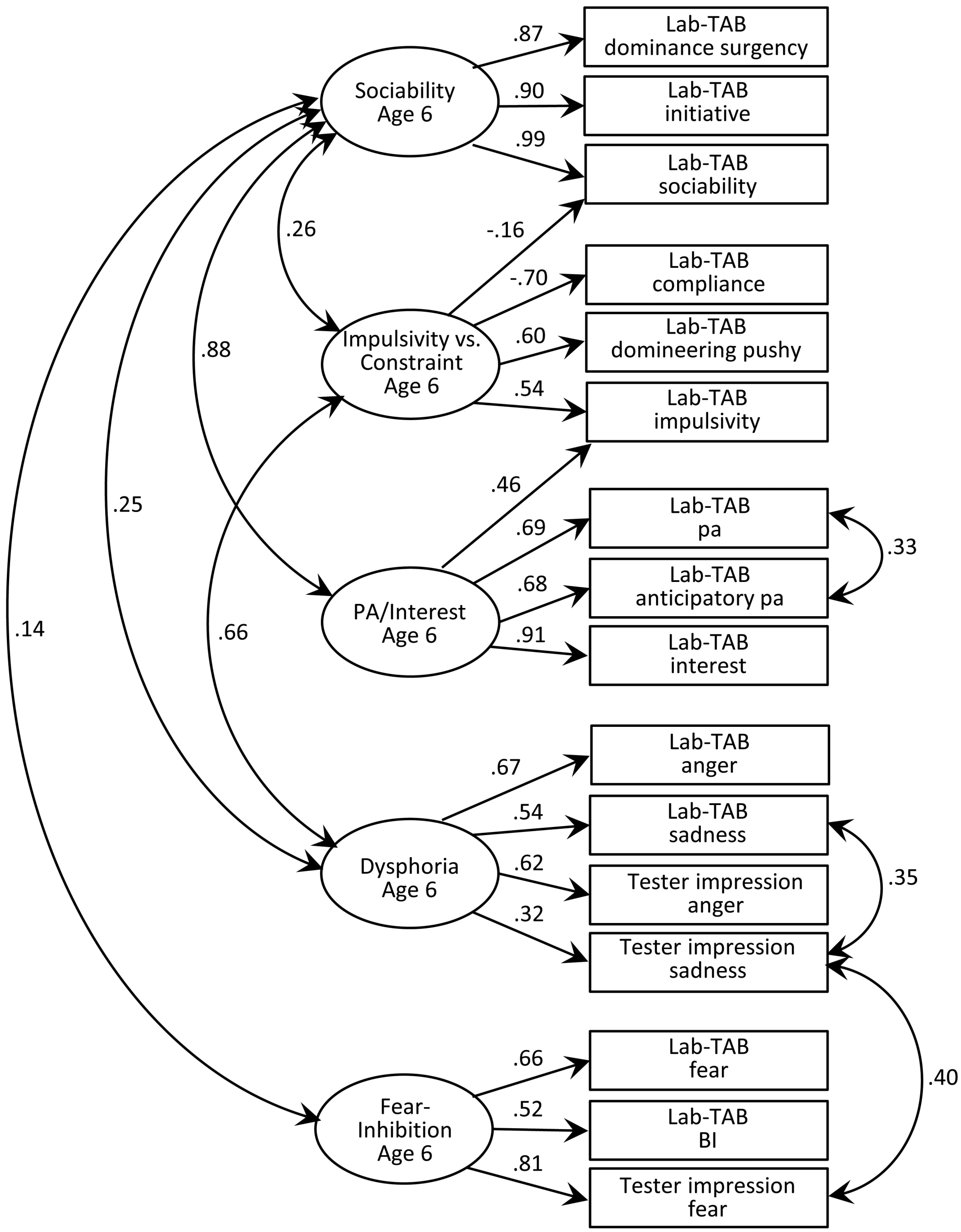

Figure 1.
Modified Age 3 CFA model with Age 6 sample[1]

[1]We also recalculated our two models with the original sample of 550 participants (at age 3 assessment). Using AMOS and maximum likelihood estimation procedures, both the age 6 CFA model (χ^2 (94) = 401.22, RMSEA = .077 (CI .069-.085), CFI =.913 and the combined age 3 to 6 model (χ^2 (406) = 1265.00, RMSEA = .062 (CI .058-.066), CFI =.895) indicated an acceptable fit to the data.

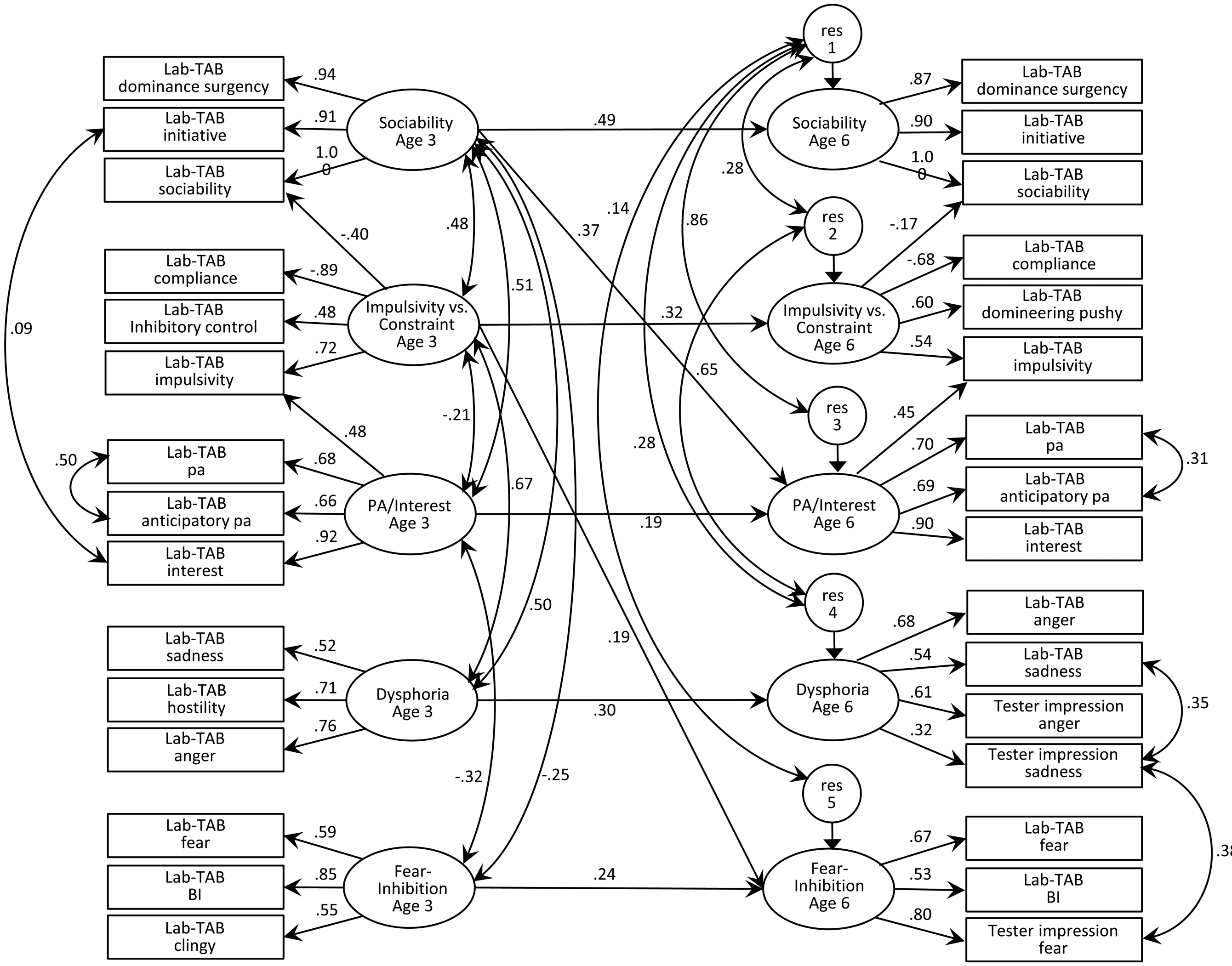

Figure 2.

Combined Age 3 and 6 model[1]

Table 1

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**	.17**	.10*	.22**	.08	.01	.01	−.08	.04	−.03	−.10*	.06	.07	−.06
AnPA	.29**	.22**	.17**	.11**	.08	.17**	.09	.01	.03	−.10*	−.02	−.08	−.15**	−.01	.07	−.02
Inter	.33**	.27**	.30**	.24**	.20**	.22**	−.05	−.06	.00	−.05	−.03	−.10	−.18**	.08	.05	.00
Soc	.33**	.28**	.35**	.40*	.39**	42**	.05	.00	.11*	.05	.07	.02	−.02	.21**	.01	.05
Domin	.30**	.29**	.38**	.40**	.41**	.44**	.09	.08	.14**	.13**	.07	.09	.00	.29**	−.06	.17**
Initia	.32**	.29**	.38**	.41**	.39**	.42**	.04	.06	.12*	.12*	.07	.07	−.02	.30**	−.04	.09
LT-Ang	.07	.12*	.10	.10*	.12*	.12*	.08	.07	.11*	.09	.12*	.12*	.17*	.15**	.02	.12**
LT-Sad	.11*	.10*	.05	.05	.11*	.10*	.24**	.16**	.15*	.18**	.08	.12**	.17**	.15**	−.09	.12**
Hostile	.07	.08	.06	.04	.03	.09	.10*	.03	.09	.06	.07	.02	.00	.15**	.00	.09
Fear	.00	−.01	−.02	.03	.04	.07	.08	.04	.04	.05	.23**	.21**	.20**	−.02	.00	−.02
BI	−.14*	−.11*	−.23**	−.17**	−.15**	−.17**	.03	.08	−.09	.04	.13**	.15**	.14**	−.15**	−.01	−.08
Clingy	−.14**	−.09	−.21**	−.12**	−.10*	−.07	.14**	.14**	−.01	.04	.12**	.14**	.16**	−.05	−.08	.01
Impul	.22*	.19**	.20**	.23**	.23**	.33**	.10*	.09	.13**	.04	.03	.06	−.05	.26**	−.03	.13*
Comp	.00	−.05	−.02	−.06	−.12**	−.11*	−.12**	−.13**	−.11*	−.19**	−.09	−.17**	−.15**	−.24**	.15**	−.15**
IC	.03	.07	.06	.05	.05	.12*	.02	.06	−.06	−.04	.03	.06	−.07	.04	−.01	.04

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.

* $p < .05$

** p < .01