



## Differential stability of temperament and personality from toddlerhood to middle childhood

Tricia K. Neppl<sup>a,\*</sup>, M. Brent Donnellan<sup>b</sup>, Laura V. Scaramella<sup>c</sup>, Keith F. Widaman<sup>d</sup>, Sarah K. Spilman<sup>a</sup>, Lenna L. Ontai<sup>e</sup>, Rand D. Conger<sup>f</sup>

<sup>a</sup> Institute for Social and Behavioral Research, Iowa State University, 2625 N. Loop Dr., Suite 2500, Ames, IA 50014, United States

<sup>b</sup> Psychology Social Science, Michigan State University, 101 Angell Building, East Lansing, MI 48824, United States

<sup>c</sup> Department of Psychology, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148, United States

<sup>d</sup> Department of Psychology, University of California, Davis, 265 Young Hall Davis, California 95616, United States

<sup>e</sup> Human and Community Development, University of California, Davis, 2327 Hart Hall Davis, CA 95616, United States

<sup>f</sup> Human and Community Development, University of California, Davis, 1361 Hart Hall Davis, CA 95616, United States

### ARTICLE INFO

#### Article history:

Available online 8 May 2010

#### Keywords:

Temperament  
Personality  
Stability  
Big Three  
Toddlerhood  
Early childhood  
Middle childhood

### ABSTRACT

This prospective, longitudinal investigation examined differential consistency of three core dimensions of individuality from toddlerhood through middle childhood. Data came from 273 families who participated with their child at least once during three developmental periods: toddlerhood (2 years), early childhood (3–5 years), and middle childhood (6–10 years). Both mothers and fathers reported on attributes of their child using subscales from the Toddler Behavior Assessment Questionnaire, the Child Behavior Questionnaire, and the Iowa Personality Questionnaire. Reports were used as indicators of the latent “Big Three” dimensions of positive emotionality, negative emotionality, and constraint at each of the three developmental periods. Results pointed to consistency in these broad dimensions of temperament and personality from toddlerhood to middle childhood.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

Dimensions of temperament have a long history in psychology (see e.g., Clark & Watson, 2008; Rothbart & Bates, 2006) and are increasingly emphasized in contemporary descriptions of social development (e.g., Caspi, Roberts, & Shiner, 2005; Caspi & Shiner, 2008; Mervielde, De Clercq, De Fruyt, & van Leeuwen, 2005; Roberts & Pomerantz, 2004; Rothbart & Bates, 2006; Sanson, Hemphill, & Smart, 2004). Indeed, individual differences in temperament are thought to influence many developmental outcomes (Rothbart & Bates, 2006; Sanson et al., 2004; Shiner & Caspi, 2003). For instance, children who are high in positive emotionality and self-regulation show higher levels of social competence and growth in social support. On the other hand, children who exhibit signs of high negative emotionality and low levels of self control have more social difficulties and higher levels of externalizing behavior problems (Caspi & Shiner, 2008; Eisenberg et al., 1996, 1997; Rothbart & Bates, 2006). In addition to social relationships, individual differences in temperament seem to play an important role in how well individuals adapt to the challenges of education and work as well as influence physical and psychological health throughout the life span (Clark & Watson, 2008; Mervielde et al., 2005; Shiner & Caspi,

2003). All in all, a complete understanding of temperament is central to understanding social development in childhood, adolescence, and beyond (Caspi et al., 2005; Rothbart, Ahadi, & Evans, 2000; Shiner & Caspi, 2003). However, questions remain as to the longitudinal consistency of temperament in the early years, especially from a multi-informant and multi-wave perspective. Accordingly, the current investigation examines continuity in key dimensions of temperament across three developmental periods: toddlerhood, the early childhood years, and middle childhood.

### 2. The differential stability of traits of temperament and personality

In light of the contention that early temperament plays a foundational role in future personality development (e.g. Caspi & Shiner, 2008; Rothbart & Bates, 2006), there has been increasing interest in the differential stability (or consistency) of temperament (e.g., Pedlow, Sanson, Prior, & Oberklaid, 1993; Roberts & DelVecchio, 2000). Differential stability reflects the degree to which the relative ordering of individuals on a particular dimension is consistent over time. Differential stability is also referred to as rank-order consistency (or stability) because it captures the degree of preservation of the rank-ordering of individuals on a given attribute (see Roberts & DelVecchio, 2000). For instance, questions of differential consistency concern whether toddlers who are

\* Corresponding author.

E-mail address: [tnepp@iastate.edu](mailto:tnepp@iastate.edu) (T.K. Neppl).

relatively happier and more cheerful than their peers develop into kindergartners who are relatively happier and more cheerful than their peers. The most common way to assess differential or rank-order stability is to correlate measures of temperament taken on the same sample across an appreciable interval of time.

Evidence of differential stability plays an important role in the understanding of psychological constructs. A construct with a relatively high level of differential consistency is considered more trait-like, whereas a construct with a relatively low level of differential consistency is more state-like (see e.g., Kenny & Zautra, 2001, p. 243). Accordingly, the degree of differential stability in temperamental characteristics can play an important role in reconciling recent debates about the nature of personality development (see e.g., Caspi & Roberts, 2001; Lewis, 2001a). For example, a strong contextual perspective on personality development (e.g., Lewis, 2001a) argues that individual characteristics are largely dependent on momentary environmental conditions. This view therefore predicts very little rank-order stability of temperamental characteristics across early periods of development as the social environments of children tend to expand dramatically from toddlerhood to middle childhood (see e.g., Lewis, 2001b, p. 111). Lewis (2001b) even argued that the “[stability] of personality characteristics is quite low for the first 30–40 years of life” (p. 111).

The emerging developmental perspective on individual differences in temperament and personality (e.g., Caspi et al., 2005; Donnellan & Robins, 2009; Roberts & Pomerantz, 2004; Shiner, 2009) offers a counter-point to the strong contextualist viewpoint. This perspective holds that there is a moderate degree of differential consistency in temperament and personality that tends to increase across the life span because of the dynamic interplay that occurs between individuals and their social contexts. According to this perspective, differential consistency increases with age due to the increasing agency that accompanies development, enabling individuals to select, modify, and create social contexts consistent with their individual dispositions (i.e., the cumulative continuity principle of personality development; Caspi et al., 2005). Increasing consistency with age also emerges from the accumulation of person–environment transactions whereby aspects of temperament affect social responses which, in turn, reinforce and accentuate those very aspects of temperament (i.e., the corresponsive principle of personality development; Caspi et al., 2005). For instance, a lively and cheerful toddler may evoke energetic and engaging responses from caregivers, experiences that may reinforce those initial temperamental dispositions. Likewise, an anger-prone toddler may tend to experience greater conflict with peers and caregivers, experiences that may accentuate negative affectivity. To the extent that these sorts of processes play out on a broad scale over the course of development, the developmental perspective predicts increasing differential stability with age. The main point, however, is that early emerging individual differences do persist over time.

Previous research has evaluated the differential stability of temperamental attributes (see e.g., Rothbart & Bates, 2006, pp. 125–126). Most notably, Roberts and DelVecchio (2000; see also Ferguson, 2010) conducted a meta-analytic review of the literature on differential stability of personality traits and included 59 studies that examined attributes of temperament in children ages 0–11.9. To facilitate comparisons of stability coefficients from different developmental periods, they estimated coefficients using a common retest interval of 6.7 years. Results suggest an increase in differential stability from the earliest age interval (0–2.9 years) to the next period (3–5.9 years) of .31–.49. The differential stability of temperament/personality when assessed between the ages of 6 and 11.9 was estimated to be .43. Importantly, they found that different traits tended to have generally similar levels of stability and any differences were small. All in all, these meta-analytic results

indicate that differential stability in temperamental attributes is evident even in very young children, contradicting a strict contextualist perspective on personality development.

Further evidence for differential stability in early temperament was demonstrated by Pedlow et al. (1993) using latent variable modeling of maternal reports to assess the stability of temperament across intervals starting at 4–8 months up to 88–99 months. Pedlow et al. concluded that continuity of temperament over time was substantial. The use of latent variable modeling reduced the attenuating effects of measurement error on estimates of consistency and may have contributed to the magnitude of the stabilities. However, a limitation of the Pedlow et al. study, which applies to much of the existing literature on the stability of temperament, is that the evidence of rank-order stability may also reflect the stability of mothers’ perceptions of their children in addition to the actual stability of the temperamental attributes in the child. One way to address this limitation is to derive estimates of stability based on variance that is shared between two knowledgeable raters of the same child (e.g., mothers and fathers). Such an approach would provide an indication of the degree of upward bias (if any) in estimates of differential stability that occurs when using a single rater of each child.

### 3. Organizing dimensions of temperament to study differential stability from toddlerhood to middle childhood

In addition to the possibility that estimates of differential stability can reflect stability in the perceptions of children’s attributes rather than actual stability in those attributes per se, other complications affect the study of differential stability of temperament across the early years of the life span. One major concern is that the content of temperament measures often changes with the age of the child. Questions that can meaningfully measure attributes of toddlers may not be suitable for older children. Studying temperamental characteristics across substantial intervals of childhood using exactly the same set of items for all measurement occasions ignores developmental differences in the frequency of expressed behaviors across time (e.g., Pedlow et al., 1993). Instead, researchers must strive to measure developmentally appropriate instantiations of the same underlying trait at each age. As a result, studying the stability of temperament across time requires theoretical models that specify which core dimensions of temperament should be evident at all ages so that meaningful estimates of consistency across developmental periods can be obtained. In other words, researchers need to draw on theoretical conceptualizations to identify the latent attributes to be assessed at different developmental periods.

Working taxonomies of personality traits can guide the study of the differential stability of temperament because they provide a conceptual basis for focusing on a common set of underlying dimensions at each developmental period. The most famous structural model of personality in adults is the five-factor model (see John, Naumann, & Soto, 2008) which includes five broad personality domains. Three-factor models have also been identified in the literature and there is increasing recognition that these models include many of the same traits that appear in the five-factor model but at a higher level of abstraction or at a higher level on the personality hierarchy (John et al., 2008; Markon, Krueger, & Watson, 2005; Tackett, Krueger, Iacono, & McGue, 2008; Watson, Clark, & Harkness, 1994). The three-factor model proposed by Tellegen (1985) is perhaps the most widely recognized because it is similar to Eysenck’s work on personality structure and serves as the underpinning of the widely used Multidimensional Personality Questionnaire (see Clark & Watson, 2008). The “Big Three” perspective organizes dimensions of personality into three “superfactors” comprised of Positive Emotionality, Negative Emotionality, and

Constraint. These superfactors are thought to form the core of temperament and personality across the life span (Tellegen, 1985).

Positive Emotionality is defined as an individual's engagement with the environment. High scorers tend to be extraverts who seek out others and are active, enthusiastic, and confident. Negative Emotionality involves perceiving the environment as threatening or distressing. Individuals scoring high on Negative Emotionality are more likely to display high levels of negative emotions and have a broad range of emotional problems. Finally, Constraint reflects the tendency to behave in a controlled manner. Highly constrained individuals likely avoid risk and plan carefully; such individuals also are more controlled by the long-range implications of their behavior rather than the immediate thrill of an action (Clark & Watson, 2008). In both child and adult models, Positive Emotionality and Negative Emotionality are expected to operate independently of one another (Rothbart et al., 2000) and are more stable than lower-order traits (Goldsmith, Lemery, Askan, & Buss, 2000).

The Big Three model also captures many of the temperamental characteristics studied in children (Putnam, Ellis, & Rothbart, 2001). For example, studies using parent-reported temperament of children ages 3–8 have demonstrated three higher-order factors that closely resemble the Big Three Model (Rothbart & Ahadi, 1994; Rothbart, Ahadi, & Hershey, 1994). These three temperament factors include Surgency, as defined by approach and pleasure; Negative Reactivity, as defined by anger and frustration; and Effortful Control, which includes inhibitory control and attentional focusing. Theoretically, Surgency during childhood corresponds to Positive Emotionality in adulthood, Negative Reactivity during childhood relates well with adulthood Negative Emotionality, and Effortful Control overlaps in many ways with adult Constraint (Rothbart & Ahadi, 1994). Likewise, Tackett et al. (2008) recently reported that the Big Three dimensions were identifiable in middle childhood using items that were derived from Tellegen's Multidimensional Personality Questionnaire. All in all, there is emerging evidence that the Big Three Model is a useful approach for organizing core individual differences in childhood and in adulthood. Given this convergence we draw upon the Big Three model to guide the present investigation and we use terms related to traits of temperament and traits of personality more or less interchangeably.

#### 4. Present investigation

The goal of the present study was to consider the stability of the Big Three traits from toddlerhood to middle childhood using a multi-rater approach. Specifically, this prospective, longitudinal investigation examined the continuity of temperament at three developmental periods: toddlerhood (when the focal children were 2 years of age), early childhood (when the focal children were 3–5 year years of age), and middle childhood (when the focal children were 6–10 year years of age). The current paper contributes significantly to the existing literature given the span of time covered by the present analyses as well as the three wave design. In particular, evidence regarding how individual differences in middle childhood are meaningfully related to temperamental variation in early childhood is needed given that research about individuality in middle childhood is fairly limited (but see Shiner, 1998; Tackett et al., 2008).

Various parent-reported measures of childhood temperament were used to create latent variables for the Big Three at each developmental period. The Big Three were conceptualized by drawing on perspectives from Clark and Watson (2008), Rothbart and Bates (2006) and Tellegen (1982). Positive Emotionality captures aspects of temperament related to zest, ambition, happiness, and sociability. Negative Emotionality captures anger, aggressiveness, and a disposition to experience unpleasant emotions easily. Finally, Con-

straint captures aspects of childhood temperament linked with effortful control, the willingness to follow rules, and the avoidance of risks.

Parent reports were used to measure children's temperamental characteristics. Although this measurement approach is not without its drawbacks, Goldsmith (1996) argued that parents have intricate knowledge of their children's behavior across a variety of settings over time. Parents can draw on rare but important behaviors when making judgments about the characteristics of their child (Caspi & Shiner, 2008) and can cover a broad range of traits. Thus, the quickest and most cost-effective way to measure children's temperament may be parent report. However, individual parents vary in how well they understand the questions being asked, in their psychological state at the time the measure is administered, in their recall of the child's behavior, in their relations with other children to use as a comparison to their own child, and in their familiarity with their own child's behavior (Goldsmith, 1996). To help overcome these potential drawbacks with ratings from any single parent, we modeled the stability in perceptions of the child's temperament that are shared by both parents (i.e., the shared variance in temperament ratings that presumably reflects agreement between mothers and fathers about the characteristics of the child). In sum, this is one of the first studies to follow children from toddlerhood to middle childhood using indicators of the Big Three model taken from both mothers and fathers.

#### 5. Method

##### 5.1. Participants

Data come from the Family Transitions Project (FTP), a longitudinal study of 559 target youth and their families. The FTP represents an extension of two earlier studies: The Iowa Youth and Families Project (IYFP) and the Iowa Single Parent Project (ISPP). In the IYFP, data from the family of origin ( $N = 451$ ) were collected annually from 1989 through 1992. Participants included the target adolescent, their parents, and a sibling within 4 years of age of the target adolescent. These two-parent families were originally recruited for a study of family economic stress in the rural Midwest. When interviewed in 1989, the target adolescent (an individual who later become one of the parents of the focal children in this study) was in seventh grade ( $M$  age = 12.7 years; 236 females, 215 males). Participants were recruited from both public and private schools in eight rural Iowa counties. Schools in communities of 6500 or less provided names and addresses of seventh grade students and their parents to serve as the initial sampling frame. Eligible families were sent a letter explaining the project and then were contacted via telephone and asked to participate. Families without telephones were contacted in person.

Due to the underlying demographics of rural Iowa, there were few minority families (less than 1% of the population in those counties); therefore, all IYFP participants were Caucasian. Seventy-eight percent of the eligible families agreed to participate. The families were primarily lower middle- or middle-class. Thirty-four percent of the families resided on farms, 12% lived in nonfarm rural areas, and 54% lived in towns with fewer than 6500 residents. In 1989, parents averaged 13 years of schooling and had a median family income of \$33,700. Families ranged in size from 4 to 13 members, with an average size of 4.94 members. Fathers' average age was 40 years, while mothers' average age was 38.

The ISPP began in 1991 when the target adolescent was in ninth grade ( $M$  age = 14.8 years), the same year of school in 1991 for the IYFP cohort of target youth. Participants included the target adolescents, their single-parent mothers, and a sibling within 4 years of age of the target adolescent ( $N = 108$ ). Telephone screeners identi-

fied families headed by a mother who had experienced divorce within two years prior to the start of the study. All but three eligible families agreed to participate. The participants were Caucasian, primarily lower middle- or middle-class, one-parent families that lived in the same general geographic area as the IYFP families. Measures and procedures for the IYFP and ISPP studies were identical, with the exception that ISPP fathers did not participate in the in-home interviews. ISPP families participated in three waves of data collection (1991–1993).

In 1994, the families from the ISPP were combined with the families from the IYFP to create the FTP. At that time, the focal adolescents from both studies were in the 12th grade. In 1994, target youth participated in the study with their parents as they had during earlier years of adolescence. Beginning in 1995, the target adolescent (one year after completion of high school) participated in the study with a romantic partner or friend. In 1997, when the targets averaged 21 years of age, the study was expanded to include the first-born child of the target. A child of one of the original FTP targets was eligible to participate in the study when he/she was at least 18 months of age. By 2005, children in the FTP ranged in age from 18 months to 13 years old. Thus, the FTP has followed the target from as early as 1989 through 2005 ( $M$  target age = 29.07 years), with a 92% cumulative retention rate.

The present report includes 273 target participants with an eligible child. Eligible children were the biological child of the target participant who participated in the study at least once by 2005. The report also includes the target's romantic partner (spouse, cohabitating partner, or boy/girlfriend) who was the other biological parent, step-parent, or parental figure to the target's child. Assessments occurred during three developmental periods. Time 1 included 228 children ranging from 18 to 29 months of age ( $M$  = 23.92 months; boys = 123). Time 2 included 222 children between the ages of 3 and 5 years old ( $M$  = 3.18 years). Since the same child could participate at age 3–5, we included data only from the first time a child was assessed during that time period to assure that we were not counting data from the same child within that age range multiple times. A total of 190 3-year-olds, 24 4-year-olds, and 8 5-year-old children participated at Time 2 (120 boys). Time 3 included 125 children between the ages of 6 and 10 years old ( $M$  = 6.19 years). Again, we included data only from the first time a child was assessed during that time period, and 111 6-year-olds, 7 7-year-olds, 5 8-year-olds, 1 9-year-old, and 1 10-year-old child participated (75 boys).

For the purpose of this study, the data were classified as mother- and father-report rather than by status of target- and romantic-partner report. Therefore, the mother in this report could be either the target or the target's romantic partner. At Time 1, there were 221 mothers (133 target participants) and 198 fathers (92 target participants). Time 2 included 211 mothers (132 targets) and 186 fathers (89 targets), and Time 3 included 116 mothers (80 targets) and 101 fathers (45 targets; see Table 1).

The relationship status between the target and their romantic partner could change across waves; therefore, the same romantic partner may not have participated during all three time points. For example, 20 target parents did not have a romantic partner who participated at any of the three time periods. Fifty-one targets had a romantic partner at Time 1 only, 10 targets had a romantic partner at Time 2 only, and 13 targets had a romantic partner only at Time 3. There were 48 targets with a romantic partner participating at all three developmental time points, 91 targets with a romantic partner at Times 1 and 2, 27 targets with a partner at Times 2 and 3, and 4 targets with a partner at Times 1 and 3. Of the 273 families, 87% were married to or cohabitating at Time 1, 84% were married or cohabitating at Time 2, and 84% were married or cohabitating at Time 3. The majority of these

**Table 1**  
Sample demographic characteristics,  $N = 273$ .

	Time 1	Time 2	Time 3
<b>Targets</b>	225	221	125
Female/mothers	133	132	80
Male/fathers	92	89	45
Average age	25.79 years	25.81 years	26.82 years
Age range	20–30 years	20–30 years	24–30 years
<b>Children</b>	228	222	125
Female	105	102	50
Male	123	120	75
Average age	23.92 months	3.18 years	6.19 years
Age range	18–29 months	3–5 years	6–10 years
<b>Romantic partners</b>	194	176	92
Female/mothers	88	79	36
Male/fathers	106	97	56
Average age	27.21 years	27.06 years	28.26 years
Age range	19–43 years	18–42 years	21–43 years
<b>Reporter</b>	419	397	217
Mother report	221	211	116
Father report	198	186	101
% Married	87	84	84
% Both partners biological parents of child	98	95	91

cohabitating partners were the other biological parent of the child (see Table 1).

## 5.2. Procedures

From 1997 through 2005, each target parent, their romantic partner, and the target's first-born child were visited in their home each year by a trained interviewer. During the visit, the target parent and his/her romantic partner completed a number of questionnaires, some of which included measures of parenting and child temperament. Parents completed questionnaires that were appropriate for their child's developmental level. For instance, when children were 2-years of age, parents independently completed the Toddler Behavior Assessment Questionnaire (TBAQ; Goldsmith, 1996), they completed the Child Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) when their children were 3–5 years of age, and the Iowa Personality Questionnaire (IPQ; Donnellan, Conger, & Burzette, 2005), a "Big Three" measure based on the Multidimensional Personality Questionnaire (MPQ), when children were 6–10 years of age. It is important to note, however, that there are indications in the current literature of convergence between the TBAQ and the CBQ (Goldsmith, Buss, & Lemery, 1997) and that MPQ-type items can be used to provide an informant-report of childhood personality (Tackett et al., 2008). These findings support our measurement strategy as described below.

## 5.3. Measures

### 5.3.1. Toddlerhood temperament (Time 1)

During the toddler period (i.e., 18–29 months of age), children's temperamental characteristics were assessed using parent reports on the TBAQ (see Goldsmith, 1996 for measure construction, validity, and reliability). The TBAQ measures several temperamental dimensions, but only the pleasure, anger proneness, and interest/persistence subscales were used in this analysis. All TBAQ items were rated on a 7-point Likert scale, ranging from never (1) to always (7). The means, standard deviations, and minimum and maximum scores for these construct indicators, as well as all remaining study variables, are provided in the Appendix.

**5.3.1.1. Positive emotionality.** During the toddler years, positive emotionality was measured using 19 items from the pleasure subscale of the TBAQ. Goldsmith (1996) defines pleasure as children's use of smiling, laughter, or other positive vocalizations in a variety of familiar situations. Sample items include: "when given a wrapped package or new toy in a bag, how often did your child squeal with joy" and "when being gently rocked or hugged, how often did your child giggle". One of the TBAQ pleasure items was dropped due to a poor loading with the rest of the items in the subscale. Scores were computed by averaging the 18 items separately for mothers and fathers. Coefficients alpha ( $\alpha$ ) were computed to estimate internal consistency (alpha = .88 for both mother and father report).

**5.3.1.2. Negative emotionality.** Negative emotionality was measured using the anger proneness subscale of the TBAQ. Anger proneness is defined as crying, protesting, hitting, or any other sign of anger in situations involving conflict with others (Goldsmith, 1996). Parents responded to 28 items regarding how likely their children were to react with negative emotions to a variety of situations. Sample items include: "when you turned off the television set because it was bedtime, dinnertime, or time to leave, how often did your child throw a tantrum" and "when you removed something your child should not have been playing with, how often did she/he scream". Scores were computed by averaging the 28 items separately for mothers and fathers. Coefficients alpha were used to estimate internal consistency; both mothers' and fathers' ratings demonstrated high internal consistency ( $\alpha = .91$  for mother report;  $\alpha = .87$  for father report).

**5.3.1.3. Constraint.** Constraint during the toddler years was measured using 22 items from the interest/persistence subscale of the TBAQ. According to Goldsmith (1996), interest/persistence is the duration of task engagement in ongoing solitary play or any other activity. Sample items include: "while coloring by her/himself, how often did your child continue to color alone for 20 min or more" and "when you told your child that she/he would have to play alone for a short time, how often did she/he require constant encouragement to remain constructively occupied". Two interest/persistence items were dropped due to poor loadings with the rest of the items in the subscale. Scores were computed by averaging the 20 items separately for mothers and fathers. Scores were internally consistent ( $\alpha = .84$  for mother report;  $\alpha = .76$  for father report).

### 5.3.2. Early childhood temperament (Time 2)

During early childhood (i.e., 3–5 years of age), temperamental characteristics were assessed using parent reports on the CBQ (Rothbart et al., 2001). The CBQ was developed to assess individual differences in positive emotional reactivity, such as pleasure, as well as negative emotional reactivity, as in sadness. It also assesses self-regulatory dimensions of attention and inhibitory control (Rothbart et al., 2001). Mothers and fathers rated each item on a 7-point Likert scale, ranging from extremely untrue (1) to extremely true (7).

**5.3.2.1. Positive emotionality.** Three different subscales of the CBQ were used to measure positive emotionality: smiling and laughter (13 items), approach/anticipation (13 items), and shyness (13 items; reversed scored). Smiling and laughter is defined as the amount of positive affect in response to changes in stimuli. Parents responded to questions regarding how likely their children were to react with positive emotion to a variety of situations (i.e., "within the past 6 months, my child often laughs out loud in play with other children"). Approach/anticipation is defined as the amount of excitement for pleasurable activities (i.e., "within the past

6 months, when my child sees a toy she/he wants, she/he gets very excited about getting it"). Finally, shyness is defined as slow or inhibited approach in uncertain situations (e.g., "within the past 6 months, my child sometimes prefers to watch rather than join the other children playing"). Scores were computed by averaging the 39 items separately for mothers and fathers. Both mothers' and fathers' ratings demonstrated high internal consistency ( $\alpha = .88$  for mother report;  $\alpha = .91$  for father report).

**5.3.2.2. Negative emotionality.** During the early childhood years, negative emotionality was measured using three subscales of the CBQ. These included anger/frustration (13 items), falling reactivity and soothability (13 items; reversed), and sadness (12 items). Anger/frustration is defined as the amount of negative affect related to an interruption of an ongoing activity (e.g., "my child has temper tantrums when she/he doesn't get what she/he wants"). Falling reactivity and soothability is the rate of recovery from distress, excitement, or general arousal. For example, "my child is easy to soothe when she/he is upset". Finally, sadness is defined as the amount of negative affect and lowered mood due to suffering and disappointment. Parents answered items such as "within the past 6 months, my child cries sadly when a favorite toy gets lost or broken" and "my child seems to feel depressed when unable to accomplish some task". Scores were computed by averaging the 38 items separately for mothers and fathers. Both mothers' and fathers' ratings were internally consistent ( $\alpha = .81$  for mother report,  $\alpha = .82$  for father report).

**5.3.2.3. Constraint.** Constraint during the early childhood years was measured using six subscales of the CBQ: attentional focusing (nine items), high intensity pleasure (13 items; reversed), low intensity pleasure (13 items), inhibitory control (13 items), impulsivity (13 items; reversed), and attentional shifting (five items). Attentional focusing is defined as the tendency to maintain focus during challenging tasks (e.g., "when picking up toys or other jobs, my child usually keeps at the task until it's done"). High intensity pleasure is defined as the amount of enjoyment in situations involving high stimulus intensity (e.g., "my child likes going down high slides or other adventurous activities"), whereas low intensity pleasure is the amount of enjoyment in situations involving low stimulus intensity (e.g., "my child enjoys just sitting in the sunshine"). Inhibitory control is the capacity to suppress inappropriate approach responses under instruction or uncertainty. For example, "my child can lower his/her voice when asked to do so". Impulsivity is defined as the speed of response to a situation (e.g., "my child usually rushes into an activity without thinking about it"). Finally, attentional shifting is the ability to transfer focus from one task to another task (e.g., "my child can easily shift from one activity to another"). Scores were computed by averaging the 66 items separately for mothers and fathers. The scores were internally consistent ( $\alpha = .84$  for mother report;  $\alpha = .86$  for father report).

### 5.3.3. Temperament in middle childhood (Time 3)

During the middle childhood years (i.e., 6–10 years of age), temperamental characteristics were assessed using parent reports on the IPQ (Donnellan et al., 2005) which is a relatively brief measure conceptually based on the Multidimensional Personality Questionnaire (MPQ; Tellegen, 1982). Therefore, both instruments assess aspects of normal personality that coalesce under three broad dimensions: Positive Emotionality, Negative Emotionality, and Constraint. Donnellan et al. (2005) found that the 42-item IPQ can be used as a reasonable alternative to the MPQ when time constraints place limits on the length of questionnaires. The IPQ items were constructed from the informant version of the MPQ developed by Tellegen, which asks informants to rate paragraph descriptions based on content analyses of the MPQ scales. The IPQ items

were constructed by modifying these attribute descriptions for use in a self-report format (Donnellan et al., 2005). Mothers and fathers rated their child on a 5-point Likert scale, ranging from your child is not at all like this (1) to your child is extremely high on this trait (5). The scoring conventions for the IPQ follow those that have been described in the MPQ literature (Donnellan et al., 2005).

**5.3.3.1. Positive emotionality.** During the middle childhood years, four subscales from the IPQ were used to measure positive emotionality. These subscales include achievement (four items), social closeness (three items), social potency (three items), and well-being (four items). Achievement is defined as working hard and being ambitious (e.g., “your child is not at all ambitious; your child is extremely ambitious and strives for perfection”). Social closeness is defined as being warm and sociable (e.g., “your child is not at all sociable and likes being alone; your child is extremely sociable and likes being with people”). Social potency is someone who enjoys being a leader (e.g., “your child is not a leader at all; your child is a natural leader and other children defer to him/her”). Finally, well-being is defined as feeling happy and cheerful (e.g., “your child is not at all happy and cheerful; your child is extremely happy and cheerful”). Scores were computed by averaging the 14 items separately for mothers and fathers. Cronbach’s alpha coefficients were computed to estimate internal consistency. Items were internally consistent ( $\alpha = .84$  for mother report;  $\alpha = .88$  for father report).

**5.3.3.2. Negative emotionality.** Negative emotionality during the middle childhood years was measured using three subscales of the IPQ: aggression (six items), alienation (five items), and stress reaction (five items). Aggression is defined as hurting others for one’s own advantage (e.g., “your child is not at all tough and does not take advantage of others; your child is extremely tough and takes advantage of others”). Alienation is defined as feeling betrayed or mistreated. For example, “your child believes that people often make things difficult for him/her; your child doesn’t believe that people make things difficult for him/her”. Finally, stress reaction is being tense, nervous, or anxious (e.g., “your child is not at all tense or worried; your child is extremely tense or worried” and “your child remains calm even in difficult situations; your child is easily upset about things”). Scores were computed by averaging the 16 items separately for mothers and fathers. Scores were internally consistent ( $\alpha = .84$  for mother report;  $\alpha = .87$  for father report).

**5.3.3.3. Constraint.** Constraint was measured using three subscales of the IPQ: control (four items), harm avoidance (three items), and traditionalism (six items). Control is defined as being reflective and careful (e.g., “your child is careful and thinks before she/he acts; your child is extremely impulsive and acts without thinking”). Harm avoidance means avoiding excitement and danger (e.g., “your child is not at all adventurous and prefers safe activities; your child is extremely adventurous and takes risks”). Finally, traditionalism is defined as having high moral standards (e.g., “your child wants people to think she/he is a nice person; your child doesn’t care what people think of him/her”). Scores were computed by averaging the 13 items separately for mothers and fathers. Responses were moderately internally consistent ( $\alpha = .59$  for mother report;  $\alpha = .77$  for father report).

#### 5.3.4. Control variables

To ascertain whether the stability of temperamental attributes was moderated by outside social or background characteristics, age of parents, gender of child, and parental relationship status were examined as control variables in subsequent analyses. The inclusion of these control variables was not expected to influence

the results. To be sure, evidence suggests that gender may be related to individual differences in temperament in terms of mean-level ratings (see Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006 for a meta-analysis). For example, in a recent study containing one sample between the ages of 3 and 12 months and another sample at 18–36 months, Putnam, Gartstein, and Rothbart (2006) found that caregivers rated female children lower in fear and higher in effortful control than male children. However, little evidence suggests that gender is a potent moderator of differential stability, especially for adults (Ferguson, 2010; Roberts & DelVecchio, 2000). In terms of relationship status, there is some evidence indicating that background socioeconomic conditions are related to parent–parent agreement on reports of internalizing and externalizing problems (e.g., Duhig, Renk, Epstein, & Phares, 2000) which could affect our estimates of stability if such effects extend to temperament. All in all, an evaluation of the statistical consequences of these control variables will enhance confidence in the robustness of the results.

## 6. Results

### 6.1. Overview

Structural equation models (SEMs) were used to test study hypotheses. SEMs and zero-order correlations among latent constructs were estimated using the AMOS 17.0 with Full Information Maximum Likelihood (FIML) estimation (Arbuckle, 1997, 2003). Due to overlap in the raters at each time point, the residuals for mother-reported temperament were allowed to correlate with each other across waves. The same procedure was used for father-reported temperament. Specifying rater-specific correlations reduces bias in the latent variable stability coefficients that can occur when these parameters are not specified (e.g., Cole, Ciesla, & Steiger, 2007). FIML was used because it is one of the most widely recommended approaches for dealing with missing data in longitudinal research (Allison, 2003; Arbuckle, 2003; Widaman, 2006). Studies indicate that it provides better estimation of model parameters than ad hoc procedures, such as listwise or pairwise deletion. Before estimating the SEMs, preliminary correlational analyses were conducted to examine the continuity of temperament across time and across reporter. Specifically, the continuity of positive emotionality, negative emotionality, and constraint were examined separately for mother and father report. Correlations were also computed to examine the relations of positive emotionality, negative emotionality, and constraint across mother and father reports. The following sections describe the results.

### 6.2. Correlational analyses

Table 2 provides the correlation coefficients for mother and father reports of positive emotionality, negative emotionality, and constraint during toddlerhood, early childhood, and middle childhood. Correlations demonstrated consistency across time and across reporter. For example, mother report of temperament as assessed by the TBAQ was statistically and significantly related to mother report of early childhood temperament assessed by the CBQ ( $r = .39, p < .001$  for positive emotionality;  $r = .48, p < .001$  for negative emotionality;  $r = .17, p < .05$  for constraint). Mother report on the CBQ also was significantly correlated with mother report of temperament during middle childhood as assessed by the IPQ. In addition, mother report on the TBAQ was statistically and significantly related to reports of temperament using the IPQ for both positive and negative emotionality. For constraint, however, the relation between the toddler years and middle childhood was not significant. Similar patterns of associations emerged regarding

**Table 2**  
Correlations among mother and father report of Positive Emotionality (PE), Negative Emotionality (NE), and Constraint (CN)  $N = 273$ .

	Mother PE			Father PE			Mother NE			Father NE			Mother CN			Father CN		
	TBAQ	CBQ	IPQ	TBAQ	CBQ	IPQ	TBAQ	CBQ	IPQ	TBAQ	CBQ	IPQ	TBAQ	CBQ	IPQ	TBAQ	CBQ	IPQ
<i>Mother PE</i>																		
TBAQ	–																	
CBQ	.39***	–																
IPQ	.21*	.39***	–															
<i>Father PE</i>																		
TBAQ	<b>.29***</b>	.11	.23*	–														
CBQ	.10	<b>.39***</b>	.22*	.44***	–													
IPQ	.13	.22*	<b>.49***</b>	.36**	.30**	–												
<i>Mother NE</i>																		
TBAQ	–.06	.07	–.14	.01	.04	–.13	–											
CBQ	–.11	–.06	–.29**	.01	.00	–.19*	.48***	–										
IPQ	–.05	.06	–.50***	–.26*	–.09	–.36***	.28**	.52***	–									
<i>Father NE</i>																		
TBAQ	.05	.08	.04	.00	.03	–.20	<b>.40***</b>	.20**	.16	–								
CBQ	–.07	–.06	–.15	–.01	–.04	–.21*	.12	<b>.35***</b>	.32**	.44***	–							
IPQ	.06	–.01	–.18*	–.21*	–.23*	–.49***	.20*	.22*	<b>.37***</b>	.29*	.33**	–						
<i>Mother CN</i>																		
TBAQ	.29***	.31***	.22*	.13*	.12	.15	–.12*	.02	–.03	–.13*	.04	.08	–					
CBQ	.09	–.05	.18*	.04	–.10	.08	–.32***	–.54***	–.39***	–.21**	–.19*	–.09	.17*	–				
IPQ	.24*	.02	.40***	.33**	.01	.23*	–.24*	–.33***	–.51***	–.07	–.30**	–.27**	.12	.52***	–			
<i>Father CN</i>																		
TBAQ	.27***	.10	.23*	.41***	.27***	.28*	–.15*	.05	–.26*	–.13*	.02	.24*	<b>.41***</b>	.07	.01	–		
CBQ	.09	–.09	.17	.04	–.20**	.22*	–.23**	–.31***	–.35**	–.43***	–.41***	–.26**	.14*	<b>.48***</b>	.45***	.20**	–	
IPQ	.22*	.01	.14	.31**	.12	.31**	–.35**	–.30**	.19*	–.30*	–.37***	–.53***	.13	.31**	<b>.44***</b>	.18	.49***	–

Coefficients in bold are mother–father agreement correlations for the Big Three traits.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

fathers' reports of positive emotionality, negative emotionality, and constraint (see Table 2).

Although the results of the correlational analyses indicated continuity of temperamental characteristics across time within reporter, evidence of continuity of temperament across reporter also emerged. For example, mother report of positive emotionality during early childhood (CBQ) was significantly correlated with father report during middle childhood (IPQ) and vice versa ( $r = .22$ ,  $p < .05$ ). For negative emotionality, father report of the TBAQ was significantly correlated with mother report of the CBQ and IPQ (see Table 2). Mother report of constraint during the toddler years (TBAQ) was significantly related to father report during early childhood (CBQ). Similarly, mother report of constraint during early childhood (CBQ) was related to father report of constraint during the middle childhood years (IPQ) and vice versa (see Table 2).

Significant relations were also found between mother and father reports of temperament within the same developmental period. For example, mother report of positive emotionality as assessed by the TBAQ was significantly related to father report of positive emotionality using the TBAQ ( $r = .29$ ,  $p < .001$ ). Significant relations between mom and dad report of the TBAQ also were found for negative emotionality ( $r = .40$ ,  $p < .001$ ) and constraint ( $r = .41$ ,  $p < .000$ ). Similarly, mother report of early childhood temperament using the CBQ was significantly correlated with father report of temperament using the CBQ. Finally, mother report of middle childhood temperament using the IPQ was statistically and significantly related to father report of temperament as assessed by the IPQ (see Table 2).

Due to the significant correlations between reporters, mothers and fathers were used as separate indicators within the same model in the analyses described below. To be sure, all models were first run separately by mothers and fathers. The results were essentially identical. Therefore, the analyses described below include both reporters within the same model.

### 6.3. Structural equation models

Standardized stability coefficients were used to test for continuity in positive emotionality, negative emotionality, and constraint. For all three models, the latent factor variances were fixed at 1.0 at each measurement occasion. Model fit was evaluated using the chi-square test of exact fit, the CFI, and the RMSEA. All discussed models passed the chi-square test (i.e., chi-square was not statistically significant at  $p < .05$ ) and were therefore judged adequate

based on the most conservative test for evaluating model fit in SEM. In addition, because the sample size drops over time, all three models were evaluated in an additional set of analyses using only those participants with complete data at all three points in time. The estimated parameters using this approach were essentially the same as the models reported here. We elected to report the estimates from the FIML models as they maximized sample sizes and reflect the best methodological practices (i.e., avoiding the use of listwise deletion for analyses; Widaman, 2006).

### 6.4. Continuity in positive emotionality

For positive emotionality, an initial model (not shown) examined the continuity from the toddler years as a direct predictor to middle childhood but not including early childhood as a mediator. The results showed that positive emotionality during toddlerhood predicted positive emotionality during middle childhood ( $b = .60$ ,  $t = 2.94$ ). This model fit the data with  $\chi^2(5) = 2.70$ ;  $p = .75$ ; Comparative Fit Index (CFI) = 1.00; Root Mean Square Error of Approximation (RMSEA) = .00. The full model presented in Fig. 1 considers the continuity and stability of positive emotionality from toddlerhood to early childhood to middle childhood. As shown in Fig. 1, the Positive Emotionality factor measured during the toddler years was statistically and significantly associated with the early childhood period latent factor ( $b = .37$ ,  $t = 2.32$ ). Similarly, the Positive Emotionality latent construct measured during the early childhood years was statistically and significantly associated with the same latent construct measured during the middle childhood years ( $b = .44$ ,  $t = 2.54$ ). As shown in Fig. 1, the model fit the data.

### 6.5. Continuity in negative emotionality

For Negative Emotionality, an initial model (not shown) examined continuity from the toddler years as a direct predictor to middle childhood but not including the path to early childhood. Results showed that Negative Emotionality during toddlerhood predicted Negative Emotionality during middle childhood ( $b = .51$ ,  $t = 2.63$ ). This model fit the data with  $\chi^2(5) = 4.62$ ;  $p = .46$ ; CFI = 1.00; RMSEA = .00. As with Positive Emotionality, the full model presented in Fig. 2 considers the continuity and stability of Negative Emotionality from toddlerhood through the middle childhood years. The Negative Emotionality latent factor measured during toddlerhood was statistically and significantly associated with the latent factor during early childhood ( $b = .46$ ,  $t = 3.36$ ).

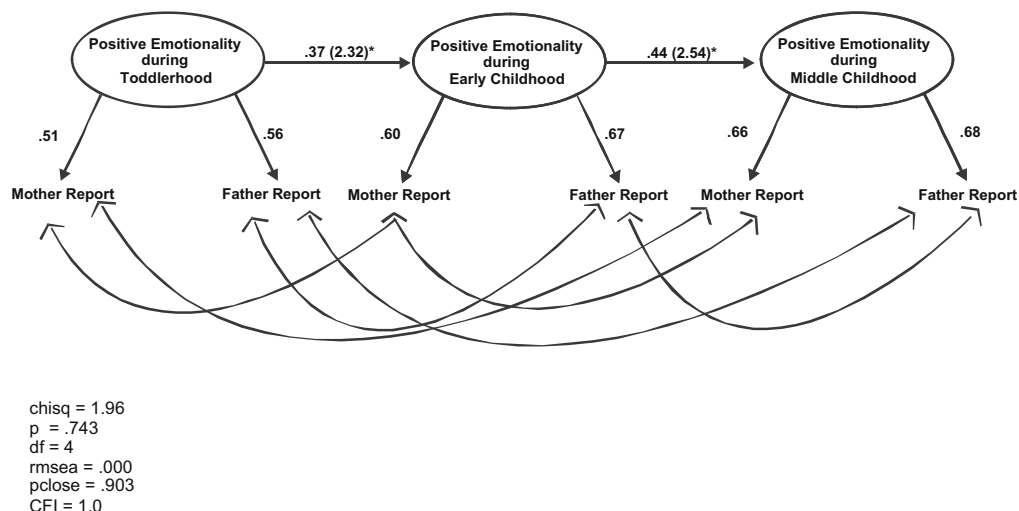


Fig. 1. Stability of Positive Emotionality from Toddlerhood to Middle Childhood.



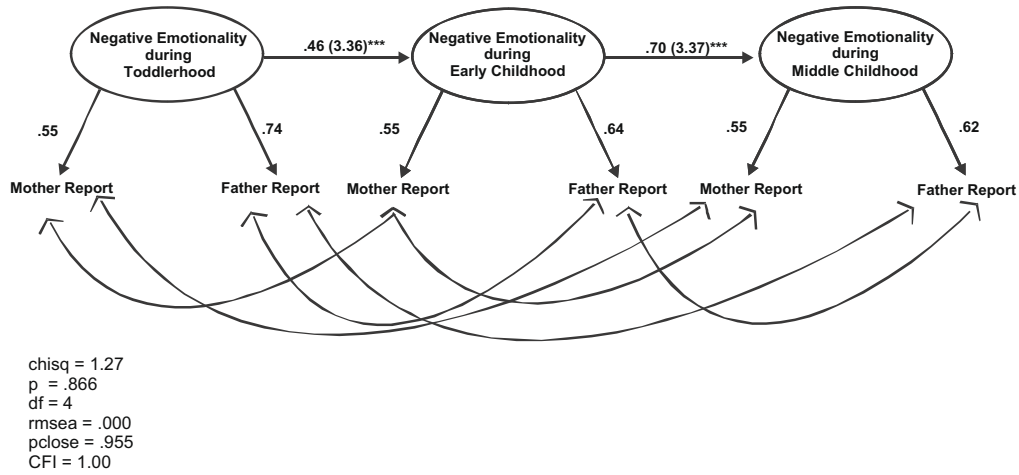


Fig. 2. Stability of Negative Emotionality from Toddlerhood to Middle Childhood.

Similarly, the latent construct for Negative Emotionality as measured during the early childhood years was statistically and significantly associated with the same latent construct measured during middle childhood ( $b = .70, t = 3.37$ ). Like the model in Fig. 1, this model also fit the data well.

6.6. Continuity in constraint

For Constraint, an initial model (not shown) examined continuity from toddlerhood as a direct predictor to the middle childhood years but not including the path to early childhood. The results showed that Constraint during toddlerhood did not predict Constraint during middle childhood ( $b = .17, t = 0.96$ ). This model also fit the data with  $\chi^2(5) = 3.33; p = .65; CFI = 1.00; RMSEA = .00$ . The full model presented in Fig. 3 considers the continuity in constraint from toddlerhood through middle childhood. As demonstrated in Fig. 3, the Constraint latent factor measured during the toddler years was statistically and significantly associated with the early childhood latent factor ( $b = .26, t = 2.00$ ). Similarly, the Constraint latent construct measured during the early childhood years was statistically and significantly associated with the same latent

construct measured during middle childhood (IPQ;  $b = .74, t = 5.24$ ). As shown in Fig. 3, this model fit the data.

6.7. Control variables

In order to rule out the possibility that continuity of temperament was affected by control variables, all three models were re-estimated adding controls for gender, age of parents, and relationship status. The results were essentially identical from those reported above. Thus, adding these covariates did not change the substantive interpretations of the results.

7. Discussion

Although a number of studies have found evidence for stability of personality in adults, relatively fewer studies have examined the continuity of temperament beginning in toddlerhood through middle childhood. Moreover, studies examining differential stability of temperament have most often relied on a single informant and examined stability using only two waves of data. To be sure, Caspi and Shiner (2008) emphasized the importance of using multiple

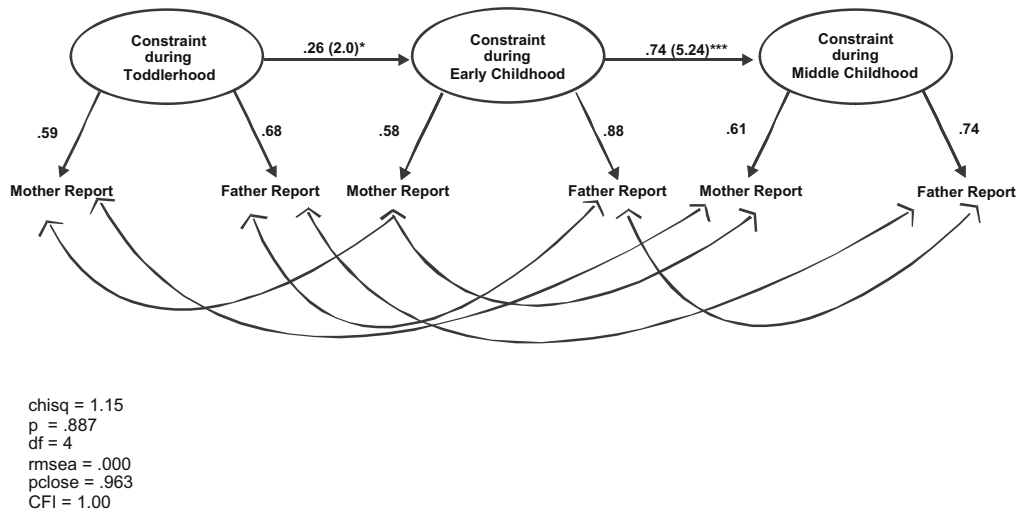


Fig. 3. Stability of Constraint from Toddlerhood to Middle Childhood.

raters when designing studies to assess trait characteristics. Thus, the current study provides a much needed perspective on the differential stability of the “Big Three” dimensions of temperament from toddlerhood through middle childhood using multi-informant data covering a substantial period of development marked by significant changes in the social contexts of children.

Latent variable models provided evidence for differential longitudinal consistency in the Big Three dimensions of temperament, consistent with a developmental perspective on individual differences (see Caspi et al., 2005). Specifically, Positive Emotionality measured during toddlerhood was associated with Positive Emotionality during the early childhood years. In turn, Positive Emotionality measured during the early childhood years was related to Positive Emotionality measured during middle childhood. The same results were true for the models measuring Negative Emotionality and Constraint.

These findings are noteworthy for three reasons. First, these results are consistent with previous research demonstrating moderate stability of temperament and personality from childhood to early adulthood, especially after age 3. However, we observed consistency when modeling variance that is shared between two informants. Thus, these results are not simply an artifact in the stability of the perceptions of any single parent or reporter. Second, stability was evident when using different instruments to measure temperamental characteristics at the three developmental periods. That is, we employed different measures of temperament designed to be developmentally appropriate for children of different ages during the course of the study. Finally, we provided data that individual differences in middle childhood are meaningfully related to individual differences in early childhood thus providing needed data on the origins of personality in middle childhood. Taken as a set, these findings contradict a strict contextualist perspective on the stability of individual difference. As it stands, the current findings are best interpreted as demonstrating continuity in temperament and personality and therefore support the emerging developmental perspective on individuality across life span.

The current analyses would not have been possible without making reference to an explicit structural model of temperament which focused our attention on a core set of three dispositions that may have different behavioral manifestations at different periods of development. These dimensions were assessed with different measures and the middle childhood assessment was even derived from an adult measure of the Big Three (the TBAQ and the CBQ share a more explicit common history; see Goldsmith et al., 1997, p. 894). This difference in measurement traditions makes the current evidence of consistency even more impressive. Moreover, our findings converge with those recently reported by Tackett et al. (2008) as both investigations suggest that items tied to constructs linked with the Multidimensional Personality Questionnaire can be used to assess individual differences in middle childhood. Tackett et al. (2008) provided evidence for links between personality in middle childhood and personality in late adolescence, whereas we provided evidence that the measures of Big Three in middle childhood are meaningfully linked with measures of the Big Three in early childhood. All in all, the current results underscore the developmental salience of the Big Three and add to a growing recognition that this framework can be applied to the study of individual differences across the life span.

In addition to providing general evidence for differential stability in the Big Three, there were trends related to connections between individual differences in toddlerhood and middle childhood. For both Positive and Negative Emotionality, evidence of a direct association between the toddler years and middle childhood emerged. This association was almost as large as the association from toddler to early childhood or early childhood to middle childhood. A statistically significant, direct association

between toddlerhood and middle childhood did not, however, occur for Constraint. This could be due to the fact that the TBAQ interest scale was the only indicator of Constraint which may have narrowed the construct and limited its predictive validity. From a developmental perspective it might make sense that researchers may struggle to measure Constraint-linked attributes in toddlerhood given that such attributes might become more elaborated as children develop language skills to aid in regulation. It might also be easier to observe outward manifestations of Constraint at later ages as there are increasing demands on children to delay gratification and exhibit self control. This conclusion is consistent with other findings from prior studies that have found increasing stability in Constraint at older ages (at least during the very early years of the life span). For example, during infancy, Kochanska and her colleagues found a modest relation between attentional control at 9 months of age and compliance with maternal demands at 14 months. However, during the preschool years, this relation was much stronger (Kochanska, Murray, & Harlan, 2000; Kochanska, Tjebkes, & Forman, 1998).

Other developmental differences in the degree of stability also emerged. That is, the magnitude of the stability coefficients from early childhood to the middle childhood years was higher than those from toddlerhood to early childhood for all three dimensions of temperament. This finding, too, is consistent with studies that have found that rank-order stability of individual differences increases with age (Ferguson, 2010; Fraley & Roberts, 2005; Roberts & DelVecchio, 2000). This finding may point to the effects of cumulative continuity from toddlerhood through middle childhood such that behavioral tendencies become increasingly stable as the consequences of individual variation build-up over time further reinforcing those tendencies and attributes (see e.g., Caspi et al., 2005). It is perhaps noteworthy, however, that such processes are evident fairly early in the life span.

Several limitations to this study also are worthy of comment. First, the lack of racial, ethnic, and geographic sample diversity may limit the generalizability of the results. Some might argue that an additional limitation is the use of parent report to measure temperament. Observational methods using standard procedures to elicit certain emotions have been developed as an alternative to parent report. That is, if a standardized observational measure is used, some might argue that it might yield more objective indices of individual behavior. These observational studies may also produce lower stability coefficients than questionnaire measures (Majdandzic & van den Boom, 2007). However, the stability coefficients found in this study are similar to those found in other studies using observational measures. For example, Durbin, Hayden, Klein, and Olino (2007) used laboratory measures to assess the stability of positive and negative emotionality in children ages 3–7. They found coefficients ranging from .45–.70, which are comparable to those found in the current study. Moreover, concerns could also be raised about limitations in laboratory procedures because these typically involve very brief samplings of relevant child behaviors. In our view, both parental reports and observational tasks are useful approaches for assessing temperament, and neither approach is categorically superior to the other.

In summary, this study used a prospective, longitudinal research design to help advance our understanding of the stability of temperament and personality from toddlerhood to middle childhood. The results suggest that there is increasing continuity in the Big Three dimensions during this period of development. The next steps are now to help identify factors that may affect the stability of these dimensions and to further clarify how temperament impacts social development. It seems clear that children respond to socialization experiences in different ways, perhaps because of their individual characteristics. As it stands, there seems to be good reason to further integrate the study of

personality and temperament across the life span to better understand the antecedents and consequences of individuality (see Caspi et al., 2005; Shiner, 2009).

### Acknowledgments

This research is currently supported by grants from the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, and the National Institute of Mental Health (HD047573, HD051746, and MH051361). Support for earlier years of the study also came from multiple sources, including the National Institute of Mental Health (MH00567, MH19734, MH43270, MH59355, MH62989, and MH48165), the National Institute on Drug Abuse (DA05347), the National Institute of Child Health and Human Development (HD027724), the Bureau of Maternal and Child Health (MCJ-109572), and the MacArthur Foundation Research Network on Successful Adolescent Development Among Youth in High-Risk Settings.

### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jrp.2010.04.004.

### References

- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*, 545–557.
- Arbuckle, J. L. (1997). *AMOS user's guide version 3.6*. Chicago: Small Waters.
- Arbuckle, J. L. (2003). *AMOS 5.0 update to the AMOS user's guide*. Chicago: Small Waters.
- Caspi, A., & Roberts, B. W. (2001). Personality development across the life span: 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*, 453–484.
- Caspi, A., & Shiner, R. L. (2008). Temperament and personality. In M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, E. Taylor, & A. Thapar (Eds.), *Rutter's child and adolescent psychiatry* (5th ed.). London: Blackwell.
- Clark, L. A., & Watson, D. (2008). Temperament: An organizing paradigm for trait psychology. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 265–286). New York: Guilford Press.
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*, 381–398.
- Donnellan, M. B., Conger, R. D., & Burzette, B. G. (2005). Criterion-related validity, self-other agreement, and longitudinal analyses for the Iowa personality questionnaire: A short alternative to the MPQ. *Journal of Research in Personality, 39*, 458–485.
- Donnellan, M. B., & Robins, R. W. (2009). The development of personality across the life span. In P. J. Corr & G. Matthews (Eds.), *Cambridge handbook of personality* (pp. 191–204). Cambridge University Press.
- Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice, 7*, 435–453.
- Durbin, C. E., Hayden, E. P., Klein, D. N., & Olino, T. M. (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion, 7*, 388–399.
- Eisenberg, N., Fabes, R. A., Guthrie, I. K., Murphy, B. C., Poulin, R., & Shepard, S. (1996). The relations of regulation and emotionality to problem behavior in elementary school children. *Development and Psychopathology, 8*, 141–162.
- Eisenberg, N., Fabes, R. A., Shepard, S. A., Murphy, B. C., Guthrie, I. K., Jones, S., et al. (1997). Contemporaneous and longitudinal prediction of children's social functioning from regulation and emotionality. *Child Development, 68*, 642–664.
- Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin, 132*, 33–72.
- Ferguson, C. J. (2010). A meta-analysis of normal and disordered personality across the life span. *Journal of Personality and Social Psychology, 98*, 659–667.
- Fraleigh, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual difference in psychological constructs across the life course. *Psychological Review, 112*, 60–74.
- Goldsmith, H. H. (1996). Studying temperament via construction of the toddler behavior questionnaire. *Child Development, 67*, 218–235.
- Goldsmith, H. H., Buss, K., & Lemery, K. S. (1997). Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of the environment. *Developmental Psychology, 33*, 891–905.
- Goldsmith, H. H., Lemery, K. S., Askan, N., & Buss, K. A. (2000). Temperamental substrates of personality development. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 1–32). Hillsdale, NJ: Erlbaum.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). The Big Five trait taxonomy: History, measurement, and conceptual perspectives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and Research* (3rd ed., pp. 114–158). New York: Guilford Press.
- Kenny, D. A., & Zautra, A. (2001). New methods for the analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change. Decade of behavior* (pp. 243–263). American Psychological Association.
- Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220–232.
- Kochanska, G., Tjebkes, T. L., & Forman, D. R. (1998). Children's emerging regulation of conduct: Restraint, compliance, and internalization from infancy to the second year. *Child Development, 52*, 83–90.
- Lewis, M. (2001a). Issues in the study of personality development. *Psychological Inquiry, 12*, 67–83.
- Lewis, M. (2001b). Continuity and change: A reply. *Psychological Inquiry, 12*, 110–112.
- Majdandzic, M., & van den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality, 75*, 121–167.
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139–157.
- Mervielde, I., De Clercq, B., De Fruyt, F., & van Leeuwen, K. (2005). Temperament, personality, and developmental psychopathology as childhood antecedents of personality disorders. *Journal of Personality Disorders, 19*, 171–205.
- Pedlow, R., Sanson, A., Prior, M., & Oberklaid, F. (1993). Stability of maternally reported temperament from infancy to 8 years. *Developmental Psychology, 29*, 998–1007.
- Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasziw Angleitner (Ed.), *Advances in research on temperament* (pp. 165–182). Lengerich, Germany: Pabst Science.
- Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The early childhood behavior questionnaire. *Infant Behavior & Development, 29*, 386–401.
- 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.
- Roberts, B. W., & Pomerantz, E. M. (2004). On traits, situations, and their integration: A developmental perspective. *Personality and Social Psychology Review, 8*, 402–416.
- Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology, 103*, 55–66.
- Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78*, 122–135.
- Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21–39.
- 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 W. Damon, & R. M. Lerner (Editors-in-Chief) & N. Eisenberg (Vol. Ed.), *Social, emotional, and personality development. Handbook of child psychology* (Vol. 3, 6th ed., pp. 99–166). New York: Wiley.
- Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*, 142–170.
- Shiner, R. L. (1998). How shall we speak of children's personalities in middle childhood? A preliminary taxonomy. *Psychological Bulletin, 124*, 308–332.
- Shiner, R. L. (2009). The development of personality disorders: Perspectives from normal personality development in childhood and adolescence. *Development and Psychopathology, 21*, 715–734.
- Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry, 44*, 2–32.
- Tackett, J. L., Krueger, R. F., Iacono, W. G., & McGue, M. (2008). Personality in middle childhood: A hierarchical structure and longitudinal connections with personality in late adolescence. *Journal of Research in Personality, 42*, 1456–1462.
- Tellegen, A. (1982). Brief manual for the multidimensional personality questionnaire. Unpublished manuscript, University of Minnesota, Minneapolis.
- Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorder* (pp. 681–706). Hillsdale, NJ: Erlbaum.
- Watson, D., Clark, L. A., & Harkness, A. R. (1994). Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology, 103*, 18–31.
- Widaman, K. F. (2006). Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development, 71*, 42–64.