Longitudinal Studies of Anger and Attention Span: Context and Informant Effects

Jungmeen Kim, **Kirby Deater-Deckard**, **Paula Y. Mullineaux**, and **Ben Allen**
Virginia Polytechnic Institute and State University

Abstract

This study examined stabilities of informant and context (home vs. classroom) latent factors regarding anger and attention. Participants included children from the National Institute of Child Health and Development Study of Early Child Care and Youth Development who were measured at 54 months, first grade, and third grade. Latent factors of anger and attention span were structured using different indicators based on mothers', fathers', caregivers', teachers', and observers' reports. We used structural equation modeling to examine the autoregressive effects within a context (stability), the concurrent associations between home and classroom contexts, and informant effects. The results indicated that for both anger and attention (1) there were significant informant effects that influenced stability in a context, (2) there was higher stability in home context than nonhome context, and (3) stability within a context increased over time. The findings suggested that anger was more prone to context effects and informant effects than attention.

Individual differences are evident in most aspects of psychological functioning spanning personality, cognitive performance, motivation and emotion, beliefs, and motor behaviors. One of the keys to understanding how individual differences unfold lies in the examination of behavioral indicators related to temperament and personality traits (i.e., stable, biologically influenced individual difference attributes) and psychopathology in early childhood and how those behaviors change into and through middle childhood. Understanding processes of stability and developmental changes in specific facets of child behavior can be greatly enhanced by improving the reliability and validity of the assessment of those facets, using multiple measures that involve observations by multiple informants. In the current longitudinal study, we performed structural equation modeling analyses using multimethod and multi-informant data to investigate the effects of contexts and informants on stability and change in two indicators of child behavior that are related to individual differences in temperament, personality and psychopathology: dispositional anger/ frustration and attention span.

Temperament includes behaviors that vary widely across individuals, are readily observed beginning early in life, are somewhat stable over time and across settings, and include a biological foundation (Rothbart & Bates, 1998; Sanson, Hemphill, & Smart, 2004; Strelau, Zawadzki, & Piotrowska, 2001). Temperament is one of the bases of personality, the latter being a much broader array of attributes including perceptions and interests (Keogh, 2004; Rothbart, Ahadi, & Evans, 2000), and temperament and personality in childhood both contribute to risk for behavioral and emotional problems (Hampson, 2008). Typically, developmental theorists state that stable behaviors represent underlying dispositions or traits. However, the evidence for this assumption is mixed. The stability of behavioral traits is typically moderate to substantial (rs = .30 to .60), and there is evidence for method/

Correspondence concerning this article should be addressed to Jungmeen Kim, Department of Psychology, Virginia Tech, 109 Williams Hall (0436), Blacksburg, VA 24061. jungmeen@vt.edu .

informant and situational effects (Sanson et al., 2004). These informant effects are seen as modest to moderate correlations (.15 to .45) between maternal and observer ratings and between mothers' and fathers' ratings on measures of temperament, personality, and maladjustment (Bornstein, Gaughran, & Segui, 1991; Carnicero, Perez-Lopez, Salinas, & Martinez-Fuentes, 2000; De Los Reyes & Kazdin, 2005; Hayden, Klein, & Durbin, 2005; Seifer, Sameroff, Barrett, & Krafchuk, 1994). This modest to moderate agreement between informants might indicate that they rely on different knowledge bases when judging children's behavior or that they are reliably reporting context-specific behaviors (Mangelsdorf, Schoppe, & Buur, 2000).

There remains a long-standing debate about which methods and informants are most valid and reliable for investigating a range of mechanisms linking behavioral traits with other important psychological outcomes. One approach that has the potential to address this debate is the use of multi-informant and method constructs with good internal and external validity (Karp, Serbin, Stack, & Schwartzman, 2004). Such an approach can not only elucidate systematic method/rater biases (if they exist), but can also promote the development of new assessment tools with incrementally superior psychometric properties. In a similar vein, psychometric theories assert that well-constructed multi-item scales are better predictors of criteria on average than are single items (Mathijssen, Koot, Verhulst, De Bruyn, & Oud, 1998; Nunnally & Bernstein, 1994). It is believed that a multimethod and informant composite score—though typically less internally consistent than a single-informant score—has better predictive validity and is more likely to yield results that will be replicated (Rushton, Brainerd, & Pressley, 1983). However, using multi-informant composites constrains the data by obscuring the examination of the variance in behavior that may or may not overlap across different settings and informants' reports.

In the current study, by developing a multi-informant measurement model that distinguishes child behavior in multiple contexts (home and child-care/school) and testing its sensitivity to detecting time and informant variants, we attempted to derive accurate estimates of the most likely "trait" and "state" components of variance in some behavioral indicators of anger/ frustration and attention span. The measures of anger and attention used in the current study involved behavioral characteristics that can be seen as behavioral indicators related to temperament facets of negative affectivity (anger) and effortful control (attention). More generally, many longitudinal studies of a broad range of child developmental outcomes have the same data structure: Parents (and sometimes observers) report on child behavior in the home context, and teachers (and sometimes different observers) report on child behavior in the child-care or school context. The development of models that can address potential context and informant effects is important, especially because context and informant effects are always confounded in the longitudinal data in this common design (i.e., parents remain as the same informants across multiple waves of data collection, but typically different teachers are informants across the same waves of data collection). Thus, our broader goal was to develop measurement models that could be applied to many developmental outcomes in longitudinal studies with multiple informants and contexts represented.

The present study examined longitudinal factor models that utilize multi-informant scores in an effort to describe and predict individual differences from 54 months to 9 years of age. More specifically, we analyzed the first two phases of data from the National Institute of Child Health and Development (NICHD) Study of Early Child Care and Youth Development (SECCYD), a multisite, on-going longitudinal study that began in 1991 with an ethnically and socioeconomically diverse sample of 1,364 children, their parents, and child-care providers. The study included measures that tap into two important behavioral indicators (anger and attention) for cognitive and social-emotional development that were

measured at home, child-care, and school environments by multiple informants involving mothers, fathers, caregivers, teachers, and observers.

We used confirmatory factor analysis models to examine measurement structures of anger and attention at 54 months, first grade, and third grade. Though the derivation of a multi-informant measurement model of anger and attention is an important first step, the more important second goal of our research was to identify systematic effects of contexts on stability and change in these indicators across different age periods (spanning 4 to 9 years old). The existing research on temperament described above—nearly all of it based on mono-method/informant assessments of constructs—suggests moderate to substantial short-term (1–2 years) stability, with more modest estimates of stability the longer the time period. In light of this literature, we hypothesized that individual differences in the multi-informant constructs of anger and attention would show moderate to high stability within the same context across time (1–2 years) as well as moderate to high correlations between contexts within a measurement time point.

Furthermore, we expected to find evidence of method and informant effects, requiring us to strike a balance between maximizing the internal consistency and the external validity of the multi-informant constructs. Accordingly, we introduced informant effects to longitudinal factor models in order to identify and distinguish consistent informant effects from reliable cross-informant variance (i.e., trait variance) when considering the context effects—that is, the role of the caregiving environment in the home and child-care/school setting. In doing so, we investigated systematic effects of informants on estimates of variance in the trajectories of cross-informant variance, manifested in stability and change in context effects.

METHOD

Participants

We examined the public data sets of the NICHD SECCYD (http://www.nichd.nih.gov/research/supported/seccyd/datasets.cfm). Data collection began in 1991 in nine states (Arkansas, California, Kansas, New Hampshire, North Carolina, Pennsylvania, Virginia, Washington, and Wisconsin) and included 1,364 children (52% male) and their families when the children were 1 month of age. The sample included participants from four racial categories: White (80%), Black (13%), Asian (2%), and other (5%; e.g., American Indian, Inuit, or Aleutian). At the time of the child's birth, all mothers were at least 18 and not more than 46 years old ($M = 28.11$, $SD = 5.63$). The current analyses included measures taken when the children were 54 months of age as well as when they were in first and third grades.

Measures

We used mothers', fathers', caregivers' or teachers' (depending on the child's age), and observers' (trained paid research staff) ratings on items pertaining to several key indicators of anger/frustration and attention span/distractibility. The items were selected based on face validity from a variety of instruments (see Table 1 for an overview) that are described here.

Parent and caregiver or teacher ratings of children's problem behavior were gathered at each time point. The Child Behavior Checklist (CBCL; Achenbach, 1991) and the closely related Teacher Report Form (TRF; Achenbach, 1991) were scored using a 3-point Likert-type scale: 0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very true or often true*. From the CBCL, we included two items for anger (easily frustrated; has temper tantrums or hot temper) and one item for attention (can't concentrate or pay attention for a long time). From the TRF, we included two items for anger (demands must be met immediately; temper

tantrums or hot temper) and three items for attention (fails to finish things he/she starts; can't concentrate/pay attention for long; inattentive, easily distracted).

The Child Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) was used to assess parental and caregiver reports of child temperament based on context-specific ratings. The CBQ items were scored on a 7-point Likert-type scale (1 = *extremely untrue of your child* to 7 = *extremely true of your child*) and yields three higher-order factors and 13 subscales. We used two of the subscales, anger/frustration (Cronbach's α = .76) and attentional focusing (Cronbach's α = .74).

We used several items from the parent and teacher versions of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990), which were rated on a 3-point Likert-type scale: 0 = *never* to 3 = *very often*. We included two parent-rated items (controls temper when arguing with other children; controls temper when in a conflict situation with parent) from the self-control scale (Cronbach's αs: .82 for mothers and .77 for fathers). Additionally, we used two parallel teacher-rated items (controls temper when arguing with peers; controls temper in conflict situations with adults) also from the self-control scale (Cronbach's α = .91). One parent-rated item pertaining to attention span/persistence (completes tasks within a reasonable time [reversed]) from the Cooperation scale (Cronbach's αs = .78 for mothers and .76 for fathers) was used.

Parents and teachers also completed a questionnaire based on the Disruptive Behaviors Disorders (DBD) Rating Scale (Pelham, Gnagy, Greenslade, & Milich, 1992), which assessed perceptions about children's behavior. The 26 behavior items were scored on a 4-point Likert-type scale: 0 = *not at all*, 1 = *just a little*, 2 = *pretty much*, and 3 = *very much*. Two items from the Oppositional Defiant Disorder Scale (often loses temper; often is angry and resentful) were used for anger across parents' and teachers' ratings (Cronbach's αs = .85 and .93, respectively). For attention, three items from the Inattentive Categorical Score (often is easily distracted; often fails to pay close attention; often has difficulty continuously paying attention) were used to measure attention across parents' and teachers' ratings (Cronbach's αs = .86 and .91, respectively).

In addition to parent and caregiver or teacher ratings, several observational coding systems also were utilized. The Observational Record of the Caregiving Environment (ORCE; Arnett, 1989) assessed child behavior and quality of care observed in the child-care setting. In addition to ratings of the frequency key behaviors, global ratings of child behavior were completed using a 4-point Likert-type scale (1 = *not at all characteristic* to 4 = *highly characteristic*). We used the global ratings of child aggression/angry affect for anger and child attention (reversed) for attention. Interrater reliability estimates based on repeated measures analysis of variance (ANOVA; i.e., the unbiased estimate of the reliability of the mean of $k = 2$ measurements after taking into account differences in the raters, $k =$ number of raters, described by Winer, 1971) for these ratings were .80 or above.

The Classroom Observation System (COS) was developed by the SECC Steering Committee for the NICHD SECCYD (2006). The COS captured discrete child behaviors and interactions with others in the classroom using a global 7-point Likert-type scale (1 = *uncharacteristic* to 7 = *extremely characteristic*). For anger, we used two of these scales (negative behavior with peers and negativity toward the teacher) and four scales were used for attention (off-task inappropriate, off-task unoccupied, spaced out/disengaged, and attention [reversed]). Interrater reliability estimates ranged from .70 to .99 for these behavioral composites.

The Unstructured Peer Observation (UPO) also was developed by the SECC Steering Committee for the NICHD SECCYD (2006). It measures a child's interaction with peers

during their recess, the least structured time of the school day. The frequency of specific behaviors was observed for 30 s and then documented for 30 s. This cycle was repeated until the end of recess (approximately 20 min). Ratings of negative affect also were recorded using a 7-point Likert-type scale: 1 = *uncharacteristic* to 7 = *extremely characteristic*. We used one of the behavioral frequency ratings—other negative behavior (verbal aggressive and nonaggressive acts intended to hurt or annoy a peer)—and one of the global ratings—negative mood (anger, hostility, and aggression). Both of these items were from the Negative Dyadic/Aggressive Play composite (Cronbach's α = .68). Interrater reliability estimates based on repeated measures ANOVA ranged from .54 to .99 for the behavioral scales and ranged from .42 to .96 for the global ratings.

The Friendship Interaction Coding (FIC) was developed by the SECC Steering Committee for the NICHD SECCYD (2006). This observational measure captured social behavior between the focus child and a friend during three structured play sessions. The social interaction items were rated on a 5-point Likert-type scale (1 = *low* to 5 = *very high*) for each of the three structured play conditions. We used two of the social interaction items during the child–friend play interactions for anger: contribution of the focus child to negative interactions and negative mood of the focus child. Contribution to negative interactions was characterized by whining, demanding and controlling behavior, negative affect, and anger by the target child (Cronbach's α = .66). The focus child's negative mood was characterized by expressions of discontent, boredom, anger, frustration, or hostility (Cronbach's α = .56). Interrater reliability estimates based upon repeated measures ANOVA ranged from .59 to .89 for these behavioral composites.

Statistical Analysis

We used confirmatory factor analyses (CFA) models via structural equation modeling (SEM; Bollen, 1989) to estimate measurement models of anger and attention. CFA (Bollen, 1989) is a powerful data reduction technique for theoretically informed structures, as it allows the use of a small number of latent variables based on the covariation among a set of observed variables. In CFA models (Long, 1983), the researcher imposes constraints upon the model, which may be necessary for theoretical or statistical reasons in order to determine (1) which pairs of common factors (i.e., latent constructs) are correlated, (2) which observed variables are affected by which common factors, (3) which observed variables are affected by a unique factor (i.e., errors in the variables), and (4) which pairs of unique factors are correlated (e.g., informant effects).

Two criteria were used to determine empirical indices of a certain latent psychological construct (Nunnally & Bernstein, 1994). First we considered *content validity* by examining the adequacy with which a specified domain of content was sampled. Each item or scale that is said to comprise the latent construct must stand on its own as an adequate representation of that construct—demonstrated empirically as showing moderate and significant bivariate correlations with other potential items/scales being considered for that construct. Next we conducted a series of factor analyses to test *construct validity*. More specifically, we tested measurement models for the latent context factors (i.e., home and nonhome) based on the scores reported by multiple informants to examine the factorial composition of child behavior. The latent factors of the home context consisted of the scores reported by mothers and fathers, and the latent factors of the nonhome context consisted of the scores reported by observers and caregivers (54 months) or teachers (first and third grades). Items were reverse scored if necessary so that higher scores indicated higher levels of dispositional anger and better attention span.

To examine stability and changes in anger and attention, we estimated longitudinal factor analysis models that specified within-occasion correlations between home and nonhome

context factors and between-occasion auto-regressive effects for the same context factor (i.e., stability of contexts). We used the Analysis of Moment Structures program (AMOS; Arbuckle, 2007) that estimated parameters incorporating full information maximum likelihood (FIML) methods that allowed data from all individuals to be included regardless of their pattern of missing data and were more appropriate than the more commonly used methods such as mean substitution. With FIML, each occasion is treated as a separate line of data, with individual and group effects estimated based on all occasions with valid data on the dependent variable.

RESULTS

We estimated zero-order bivariate Pearson correlations between all of the study variables separately for anger and attention at each occasion (54 months, first grade, and third grade; available upon request). For anger, interitem correlations ranged from .31 to .50 for the home context (3 items) and −.01 to .60 for the nonhome context (6 items) at 54 months, .18 to .48 for the home context (6 items) and −.00 to .54 for the nonhome context (8 items) in first grade, and .16 to .61 for the home context (10 items) and .13 to .77 for the nonhome context (8 items) in third grade. For attention, interitem correlations ranged from .29 to .50 for the home context (3 items) and .16 to .71 for the nonhome context (5 items) at 54 months, .19 to .40 for the home context (4 items) and .16 to .80 for the nonhome context (5 items) in first grade, and .17 to .65 for the home context (10 items) and .19 to .82 for the nonhome context (8 items) in third grade.

Testing Longitudinal Factor Models

In the multitrait–multimethod design (Campbell & Fiske, 1959), different informants represent possible systematic errors that are referred to as method factors. Figures 1 and 2 present longitudinal factor models of context and informant effects for anger and for attention, respectively. We introduced the determinants of the method factors to gain insights into the potential for confounding or spurious effects caused by systematic measurement errors (see Bollen & Paxton, 1998). As can be seen in Figures 1 and 2, when there were three or more manifest variables that were provided by the same informant, a latent factor was introduced to represent the particular informant. Such a latent method factor encompasses the total influences from a single informant that do not directly fall under the definition of the construct under study. When there were only two manifest variables provided by the same informant (thus a latent informant factor could not be constructed), informant effects were estimated by correlating between the two unique factors.

We tested systematic error ("biases") due to informant effects by comparing two models— one with and one without informant/method factors. If there are no systematic method/ informant variances (i.e., no significant method/informant factors), then the fit of the simpler model should be essentially the same as the more complex model (with informant effects). Stabilities within the same informants were estimated through autocorrelations over time (between the adjacent times) for unique factors for the same informant (see McArdle & Nesselroade, 1994).

Anger—In the longitudinal factor models for anger, the fit for the model without informant effects was moderate, $\chi^2(773) = 3,418.84$, $p = .00$, cumulative fit index (CFI) = .78, root mean squared error of approximation (RMSEA) = .05. In a subsequent model, seven latent factors that represented informant effects were introduced, representing different sources of informants: caregiver at 54 months, mother and father in first and third grades, and teacher in first and third grades. As can be seen in Figure 1, several correlations between unique

factors were added to account for informant effects within a time point: for mother ratings at 54 months and observer ratings at 54 months, first grade, and third grade. For observer ratings, informant effect correlations were introduced only between the manifest variables that were provided by the same observer. To consider the stability of informant effects, we attempted to estimate correlations between unique factors at 54 months and the latent informant factors in first grade for the same informant and correlations between the latent informant factors in first grade and in third grade for mother and father, respectively. This model failed to converge, so we estimated longitudinal correlations between unique factors of the corresponding items within the same informant between two adjacent occasions (four correlations for mothers and four correlations for fathers; e.g., mothers' report of a CBCL anger item between 54 months and first grade).

The model with informant effects provided an adequate fit, $\chi^2(731) = 1{,}371.23$, $p = .00$, CFI $= .95$, RMSEA $= .03$. Compared to the model without informant effects, the model with informant effects provided a significantly better fit, $\Delta\chi^2 = 2{,}047.61$, $\Delta df = 42$, $p < .05$. As can be seen in Figure 3, a closer examination of the significant coefficients in the best-fitting model (the model with informant effects) indicated that all of the within-context stability coefficients were significant. There was a strong stability in anger in the home context between 54 months and first grade ($B = .88$, $SE = .07$, $\beta = .87$, $p < .05$) and between first grade and third grade ($B = 1.07$, $SE = .07$, $\beta = .96$, $p < .05$). Similarly, there was a significant stability, with weaker magnitude compared to home context, in anger measured in the nonhome context (preschool to school) between 54 months and first grade ($B = .33$, $SE = .05$, $\beta = .62$, $p < .05$) and between first grade and third grade ($B = .69$, $SE = .06$, $\beta = .62$, $p < .05$). With respect to concurrent correlations between home and nonhome contexts, anger measured in the home context was significantly associated with anger measured in the nonhome context at 54 months ($r = .44$, $p < .05$), in first grade ($r = .27$, $p < .05$), and in third grade ($r = .42$, $p < .05$).

We observed that the stability within a context and the cross-context correlations became stronger by considering informant effects that may have been confounded with context effects. That is, when we compared the anger longitudinal factor models with and without informant effects, stabilities for the home context increased from $\beta = .81$ to $\beta = .87$ for 54 months–first grade and from $\beta = .85$ to $\beta = .96$ for first grade–third grade. Stabilities for the nonhome context increased from $\beta = .26$ to $\beta = .62$ for 54 months–first grade and from $\beta = .31$ to $\beta = .62$ for first grade–third grade. A similar trend was found for the cross-context correlations: The correlation between home and nonhome contexts increased from $r = .31$ to $r = .44$ at 54 months, from $r = .19$ to $r = .27$ in first grade, and from $r = .25$ to $r = .42$ in third grade. As for stability of informant effects, autocorrelations over time for the unique factors ranged from $r = .19$ to $r = .41$ for mother reports and from $r = .17$ to $r = .52$ for father reports.

Attention—The fit for the model without informant effects was acceptable, $\chi^2(553) = 2{,}006.95$, $p = .00$, CFI $= .90$, RMSEA $= .04$. To test informant effects, we introduced five latent factors that represented method or informant effects in a subsequent model (see Figure 2). These latent informant factors represented different sources of informants: Caregiver at 54 months, mother and father in third grade, and teacher in first and third grades. Several correlations between unique factors were added to account for informant effects: mother ratings at 54 months and observer ratings in first and third grades. Stability of informant effects was estimated by correlations among unique factors for the corresponding items for mother (three correlations) and for father (three correlations), respectively, between 54 months and first grade and between first grade and third grade. The model with informant effects provided a significantly better fit compared to the model without informant effects, $\chi^2(519) = 940.33$, $p = .00$, CFI $= .97$, RMSEA $= .02$; $\Delta\chi^2 = 1{,}066.62$, $\Delta df = 34$, $p < .05$.

In Figure 4, the results of the best-fitting model (including informant effects) revealed that there was significant stability in attention measured in the home context between 54 months and first grade ($B = .62$, $SE = .05$, $\beta = .79$, $p < .05$) and between first grade and third grade ($B = 1.15$, $SE = .06$, $\beta = .87$ $p < .05$). Stability of attention measured in nonhome contexts was indicated by a significant autoregressive coefficient between 54 months and first grade ($B = .48$, $SE = .05$, $\beta = .68$, $p < .05$) and between first grade and third grade ($B = 1.27$, $SE = .08$, $\beta = .76$, $p < .05$). In addition, all of the correlations between home and nonhome contexts were significant ($r = .67$, $p < .05$ at 54 months; $r = .81$, $p < .05$ in first grade; and $r = .56$, $p < .05$ in third grade).

Similar to anger, the stabilities of home and nonhome contexts and the cross-context correlations became stronger by adding informant effects in the longitudinal factor model for attention. This was evidenced by increases in within-context stabilities after introducing informant effects. For the home context, stabilities increased from $\beta = .76$ to $\beta = .79$ for 54 months–first grade and from $\beta = .80$ to $\beta = .87$ for first grade–third grade. Within-context stabilities for the nonhome context increased from $\beta = .47$ to $\beta = .68$ for 54 months–first grade and from $\beta = .58$ to $\beta = .76$ for first grade–third grade. The cross-context correlations between home and nonhome contexts also increased: from $r = .53$ to $r = .67$ at 54 months, from $r = .65$ to $r = .81$ in first grade, and from $r = .51$ to $r = .56$ in third grade. As for stability of informant effects, autocorrelations over time for the unique factors ranged from .10 to .35 for mother reports and from .08 to .33 for father reports.

It seems that the informant effects on the estimates of stability were stronger for anger than attention, as evidenced by the larger increases in stability coefficients after accounting for the informant effects. The average increase (comparing effects before and after inclusion of informant effects) in the within-context stability coefficients was .21 for anger, whereas the average increase in the within-context stability coefficients was .12 for attention.

DISCUSSION

We investigated stability and change in anger/attention from 4 to 9 years of age in an effort to more adequately describe and better understand the nature of context and informant effects in studying the development of effortful control of attention and regulation of anger specifically and the role of context on estimates of stability of individual differences more generally. Attention regulation (effortful control) and anger (negative affectivity) have been shown to play a major role in children's social adjustment and school functioning. It is ideal to use independent, multisource assessments, because reliance on single-informant data can inflate any associations between indicators of temperament and personality and developmental outcomes due to shared method variance (e.g., Sanson et al., 2004). To our knowledge, the current study is the first to examine context effects and simultaneously consider method factors in the development of dispositional anger and attention from early childhood into and through middle childhood.

We used multimethod multi-informant scores based on a large, representative national sample to model context-specific stability (i.e., home vs. child care/school). In doing so, we attempted to test systematic informant effects on the stability of anger and attention within a context, as suggested by Kraemer et al. (2003) who differentiated factor weights for "trait," "perspective" (i.e., characteristics of the informant), and "context" (i.e., factors related to circumstances that might influence the subject's expression of the trait). We tested alternative statistical models with and without the determinants of method factors (systematic measurement errors due to a particular informant) in examining contextual influences on the stability in dispositional anger and attention span. The temporal stability in anger and attention was substantial for parents' ratings assessed in the home context,

whereas the temporal stability in anger and attention was moderate to substantial for caregivers', teachers', and observers' ratings assessed in the nonhome context (e.g., daycare, school). These findings in general parallel prior research suggesting that temperament and personality dimensions are fairly stable over time and across settings (Sanson et al., 2004; Strelau et al., 2001). In the current study, significant improvement in model fits after considering informant effects further suggested that significant variances in the stability of behaviors were attributable to systematic errors introduced by informant characteristics.

Our data suggested that anger was more prone to context effects and informant effects than was attention. The discrepancy in stabilities between home and nonhome contexts was greater for anger than attention, indicating that anger was more affected by context-specific effects (e.g., Mangelsdorf et al., 2000). In addition, changes in temporal stability in anger (compared to attention) were more greatly affected by partialling out informant effects. Specifically, whereas there were slight increases in stability in attention for home and nonhome contexts, notable changes were observed in within-context stability in anger, especially for the child-care/school context. In particular, dispositional anger assessed in daycare and school seems to be more vulnerable to informant bias effects compared to dispositional anger assessed in the home context that involves mother and father reports. This is probably because different informants were involved in rating child behavior at various ages, because their nonparental caregivers and school teachers typically changed from one year to the next.

For both anger and attention, regardless of context, there was a tendency that temporal stability increased over time. This was particularly true within the nonhome context, in which stability for attention was lower between 54 months and first grade than between first and third grades, reflecting transitions from day-care to school settings between 54 months and first grade. This finding is consistent with prior research that has shown increasing stability in attention span in the transition to and through elementary school, accompanied by increases in genetic influences (Deater-Deckard, Petrill, Thompson, & DeThorne, 2006).

As expected, the correlations between behavior in the home and child-care/school contexts were significant and high for both anger and attention, although the consistency across contexts was higher for attention than for anger. Correlations between home and nonhome contexts were medium to large (ranging from .56 to .81) for attention and moderate (ranging from .27 to .44) for anger from 4 to 9 years of age. These findings suggest that attention may be considered more trait-like, whereas anger may be considered more transient or state-like (e.g., Strelau, 2001).

Limitations and Conclusions

The main strength of this study was its analysis of multi-informant and multimethod scores based on a large, representative normative sample. Although secondary analysis offers opportunities for increasing the informational value as well as being a relatively low-cost way to ask original research questions (Bullock, 2007), we were limited as to the available variables in constructing the measurement models of anger and attention. One consequence was that we were unable to consistently construct latent informant factors for anger and attention across all occasions. Consequently, when we had less than three items for a particular informant, we estimated informant effects through correlations between unique factors for the items that were reported by the same informant. Another consequence was that some of the items used in the current analysis were rather broadly defined indicators of anger and attention. For example, observational data (such the COS) included items that had somewhat compromised content validity. In addition, some of the observational measures had less than optimal psychometric properties (e.g., the low internal consistency and interrater reliability of the UPO and the FIC coding). Nevertheless, we believe that

observational data made a valuable contribution to our measurement model by adding an independent informant to parents' and teachers' reports.

Second, because we examined items and scales from various informants across different instruments at any given point in time, only some of the items that were used to construct the latent factors of anger and attention were consistent across the three occasions. Therefore, our longitudinal factor models did not permit testing of factor invariance over time. This is a common problem in longitudinal studies of early and middle childhood, in which the particular assessment instruments that are used for younger versus older children tend to change. Despite a lack of complete overlap among the variables across occasions, our data showed strong temporal stability for the anger and attention latent factors.

Using multi-informant multimethod assessments is strongly encouraged in personality, temperament, and developmental psychopathology research (e.g., Sanson et al., 2004). Assuming adequate internal consistency, analyzing composites that are comprised of different informants' reports produces a substantial reduction in the number of statistical effects being estimated, which maximizes power while minimizing the study-wide Type 1 error rate. In addition, the predictive validity of multi-informant composite scores also is optimized compared to single-informant scores, because any effects of random measurement error are minimized (Kane & Case, 2004; Rushton et al., 1983). However, before advocating the usage of composite scores, it is important to examine the variances in behavior that represent the separate and combined influences of the individual's actual traits (or characteristics), the context in which the individuals were observed, and the biases of the different informants (Kraemer et al., 2003).

Our approach permitted examination of informant- or context-specific sources of variance in the development of anger and attention observed by multiple adults. Disaggregating and controlling for such variance will contribute to improving reliability and predictive validity of a construct. At the same time it also is likely that at least some of the variance attributable to contexts and informants could yield further insights if it were examined in its own right. The present results suggest that researchers who study development of personality and psychopathology need to bear in mind that stability in a psychological construct may vary depending on the context and that these context-specific effects are influenced by (and often confounded with) informant effects. This is a very important consideration, given that most longitudinal studies that include multiple informant sources also unfortunately confound stability of informant and stability of context. That is, most children live in the same household with the same parent or parents throughout childhood, and the stability of their behavior in the home context will in part reflect stability arising from having the same informants (mothers and fathers) reporting on those children's behaviors repeatedly. In contrast, measures of children's behaviors in the classroom context will reflect year-to-year changes in the actual setting as well as changes in informant (i.e., different teachers each year).

Furthermore, it would be informative to apply the approach we used in this study to questions about personality stability later in life. For example, it could be important to consider context and informant effects in studying the stability of personality traits among college students using self-reports versus others' ratings (such as family and peer ratings). Our hope is that the approach to testing longitudinal measurement models presented in the current study will provide a framework from which personality and temperament researchers can build longitudinal models that address this confound when estimating stability of individual differences in behaviors across diverse contexts.

Acknowledgments

REFERENCES

Achenbach, TM. Manual for the Teacher's Report Form and 1991 Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.

Arbuckle, J. Amos (Version 16.0.1) [Computer software]. Spring House, PA: Amos Development Corporation; 2007.

Arnett J. Caregivers in day-care centers: Does training matter? Journal of Applied Developmental Psychology. 1989; 10:541–552.

Bollen, KA. Structural equations with latent variables. New York: John Wiley & Sons; 1989.

Bollen KA, Paxton P. Detection and determinants of bias in subjective measures. American Sociological Review. 1998; 63:465–478.

Bornstein MH, Gaughran JM, Segui I. Multimethod assessment of infant temperament: Mother questionnaire and mother and observer reports evaluated and compared at five months using the Infant Temperament Measures. International Journal of Behavioral Development. 1991; 14:131–151.

Bullock M. Secondary analysis: Extending the value of data. Journal of Applied Developmental Psychology. 2007; 28:383.

Campbell DT, Fiske DW. Convergent and discriminant validation by multitrait-multimethod matrix. Psychological Bulletin. 1959; 56:81–105. [PubMed: 13634291]

Carnicero JAC, Perez-Lopez J, Salinas MDCG, Martinez-Fuentes MT. A longitudinal study of temperament in infancy: Stability and convergence of measures. European Journal of Personality. 2000; 14:21–37.

De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. Psychological Bulletin. 2005; 131:483–509. [PubMed: 16060799]

Gresham, FM.; Elliott, SN. The social skills rating system. Circle Pines, MN: American Guidance Service; 1990.

Hampson SE. Mechanisms by which childhood personality traits influence adult well-being. Current Directions in Psychological Science. 2008; 17:264–268. [PubMed: 19809527]

Hayden EP, Klein DN, Durbin CE. Parent reports and laboratory assessments of child temperament: A comparison of their associations with risk for depression and externalizing disorders. Journal of Psychopathology and Behavioral Assessment. 2005; 27:89–100.

Kane M, Case SM. The reliability and validity of weighted composite scores. Applied Measurement in Education. 2004; 17:221–240.

Karp J, Serbin LA, Stack DM, Schwartzman AE. An observational measure of children's behavioural style: Evidence supporting a multi-method approach to studying temperament. Infant and Child Development. 2004; 13:135–158.

Keogh, BK. Temperament in the classroom: Understanding individual differences. Baltimore: Paul H. Brookes; 2004.

Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. American Journal of Psychiatry. 2003; 160:1566–1577. [PubMed: 12944328]

Long, JS. Confirmatory factor analysis: A preface to LISREL. Newbury Park, CA: Sage; 1983.

Mangelsdorf, SC.; Schoppe, SJ.; Buur, H. The meaning of parent reports: A contextual approach to the study of temperament and behaviour problems in childhood. In: Molfese, VJ.; Molfese, DL.,

editors. Temperament and personality development across the life span. London: Erlbaum; 2000. p. 121-140.

Mathijssen JJJP, Koot HM, Verhulst FC, De Bruyn EEJ, Oud JHL. The relationship between mutual family relations and child psychopathology. Journal of Child Psychology and Psychiatry. 1998; 39:477–487. [PubMed: 9599776]

McArdle, JJ.; Aber, MS. Patterns of change within latent variable structural equation models. In: von Eye, A., editor. Statistical methods in longitudinal research. Vol. Vol. 1. San Diego, CA: Academic Press; 1990. p. 151-224.

McArdle, JJ.; Nesselroade, JR. Using multivariate data to structure developmental change. In: Cohen, SH.; Reese, HW., editors. Life-span developmental psychology: Methodological contributions. Hillsdale, NJ: Erlbaum; 1994. p. 223-267.

NICHD SECCYD. Manuals of operation. 2006. Retrieved from http://secc.rti.org/

Nunnally, JC.; Bernstein, IH. Psychometric theory. 3rd ed.. New York: McGraw-Hill; 1994.

Pelham WE, Gnagy E, Greensdale KE, Milich R. Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. Journal of the American Academy of Child and Adolescent Psychiatry. 1992; 31:218.

Rothbart MK, Ahadi SA, Evans DE. Temperament and personality: Origins and outcomes. Journal of Personality & Social Psychology. 2000; 78:122–135. [PubMed: 10653510]

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: Damon, W.; Eisenberg, N., editors. Handbook of child psychology: Vol. 3, Social, emotional and personality development. 5th ed.. New York: Wiley; 1998. p. 105-176.

Rushton JP, Brainerd CJ, Pressley M. Behavioral development and construct validity: The principle of aggregation. Psychological Bulletin. 1983; 94:18–38.

Sanson A, Hemphill SA, Smart D. Connections between temperament and social development: A review. Social Development. 2004; 13:142–170.

Seifer R, Sameroff AJ, Barrett LC, Krafchuk E. Infant temperament measured by multiple observations and mother report. Child Development. 1994; 65:1478–1490. [PubMed: 7982363]

Strelau J. The concept and status of trait in research on temperament. European Journal of Personality. 2001; 15:311–325.

Strelau, J.; Zawadzki, B.; Piotrowska, A. Temperament and intelligence: A psychometric approach to the links between both phenomena. In: Collis, JM.; Messick, S., editors. Intelligence and personality: Bridging the gap in theory and measurement. Mahwah, NJ: Erlbaum; 2001. p. 61-78.

Winer, BJ. Statistical principles in experimental design. 2nd ed.. New York: McGraw-Hill; 1971.
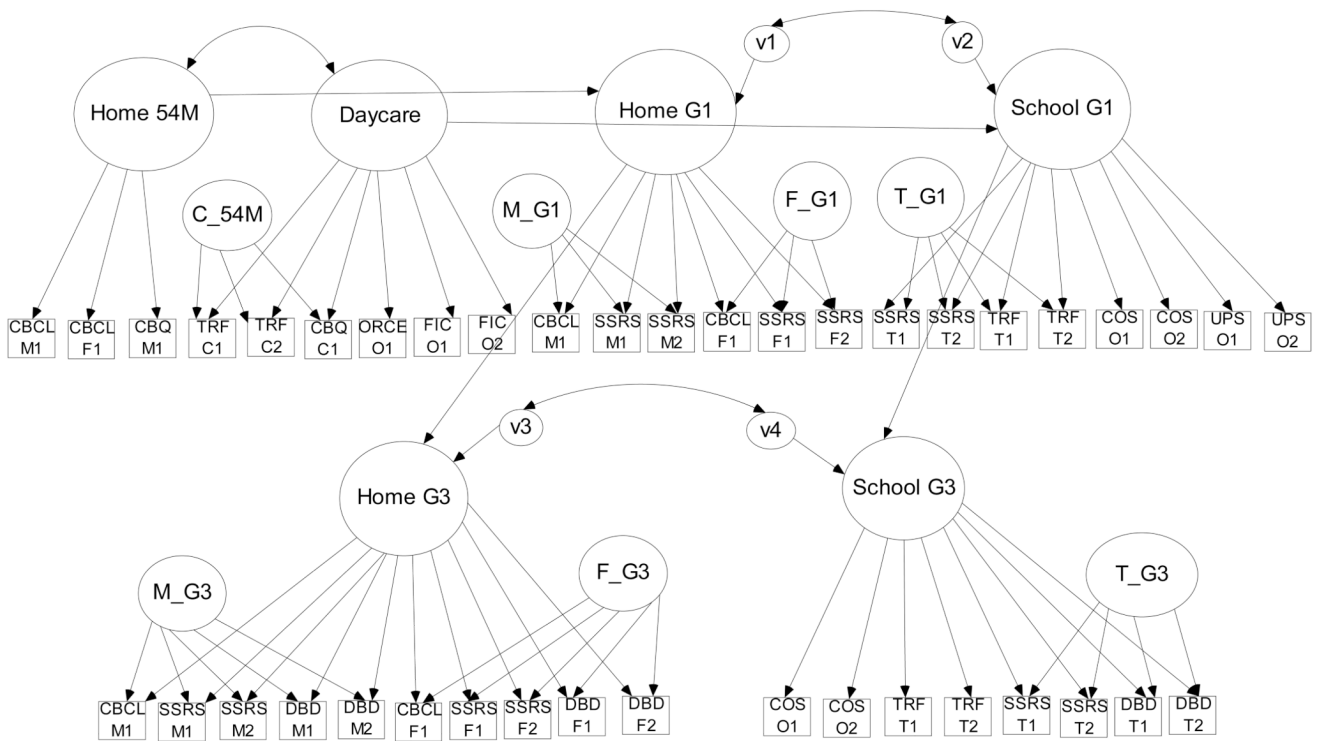
Figure 1.
Longitudinal factor model of context and informant effects for anger from 54 months to third grade. For graphical simplicity, unique factors and correlations across time points were omitted from the figure. CBCL = Child Behavior Checklist; TRF = Teacher Report Form; CBQ = Child Behavior Questionnaire; SSRS = Social Skills Rating System; DBD = Disruptive Behaviors Disorders Rating Scale; ORCE = Observational Record of the Caregiving Environment; COS = Classroom Observation System; UPO = Unstructured Peer Observation; FIC = Friendship Interaction Coding. M = mother; F = father; C = caregiver; T = teacher; O = observer; 54M = 54 months; G1 = first grade; G3 = third grade.
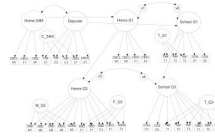
Figure 2.
Longitudinal factor model of context and informant effects for attention from 54 months to third grade. For graphical simplicity, unique factors and correlations across time points were omitted from the figure. CBCL = Child Behavior Checklist; TRF = Teacher Report Form; CBQ = Child Behavior Questionnaire; SSRS = Social Skills Rating System; DBD = Disruptive Behaviors Disorders Rating Scale; ORCE = Observational Record of the Caregiving Environment; COS = Classroom Observation System; UPO = Unstructured Peer Observation; FIC = Friendship Interaction Coding. M = mother; F = father; C = caregiver; T = teacher; O = observer; 54M = 54 months; G1 = first grade; G3 = third grade.
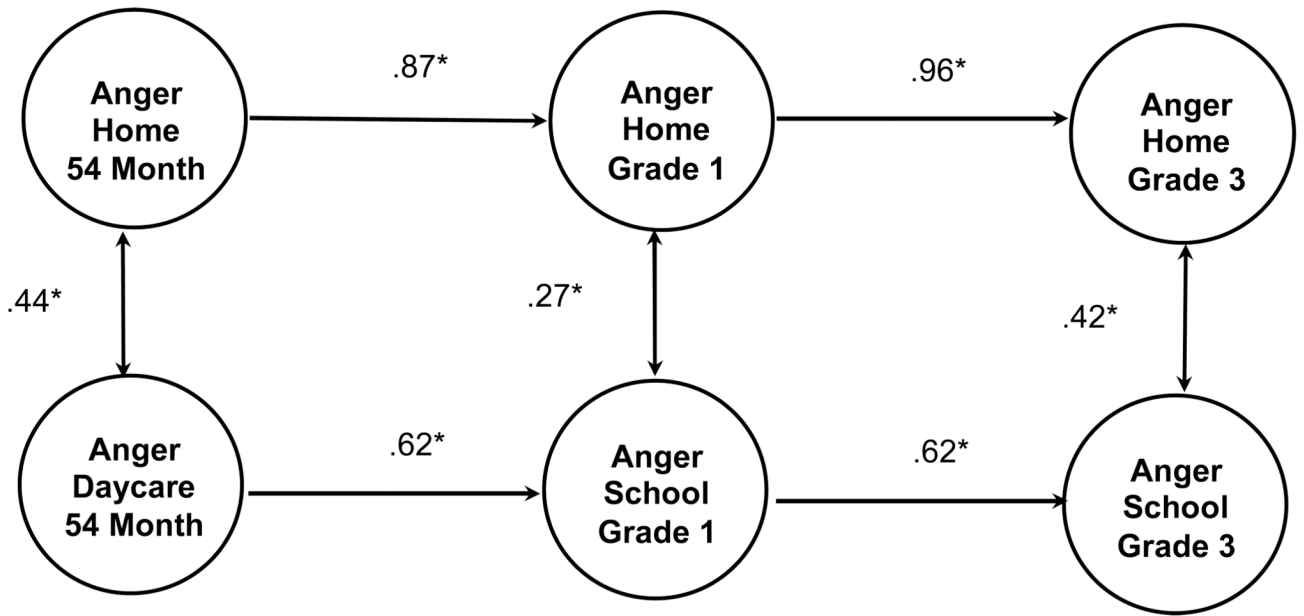
Figure 3.
Structural equation model showing standardized estimates of stability coefficients and cross-context correlations for Anger. *p < .05.
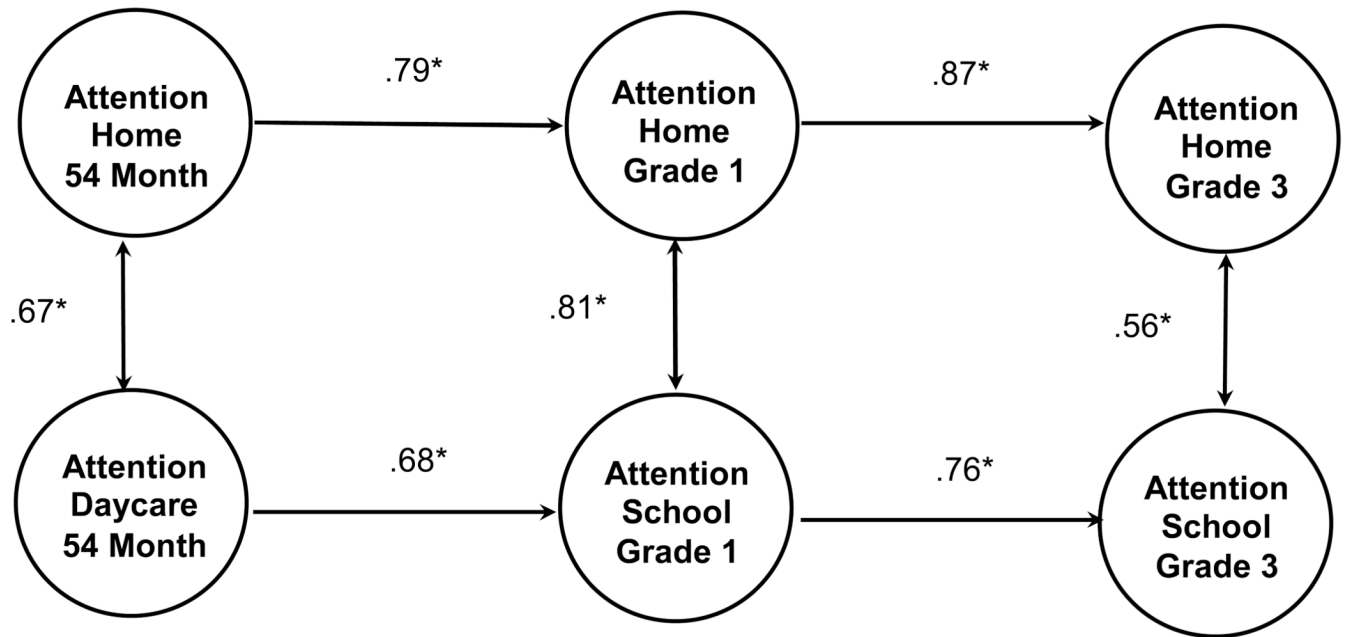
Figure 4.
Structural equation model showing standardized estimates of stability coefficients and cross-context correlations for Attention. *p < .05.

Table 1

Measures Used to Construct Anger and Attention Factors by Time Point, and *N* with Valid Data for Mothers' (M), Fathers' (F), Caregivers' (C), Teachers' (T), and Observers' (O) Reports

	54 months				First Grade				Third Grade			
	M	F	C/T	O	M	F	C/T	O	M	F	C/T	O
Child Behavior Checklist	1,052	802	—	—	1,008	668	—	—	1,006	636	—	—
Child Behavior Questionnaire	1,023	—	729	—	—	—	—	—	—	—	—	—
Classroom Observation System	—	—	—	—	—	—	—	966	—	—	—	971
Disruptive Behaviors Disorders	—	—	—	—	—	—	—	—	1,024	749	977	—
Friendship Interaction Coding	—	—	—	746	—	—	—	—	—	—	—	—
Observational Rating of the Caregiving Environment	—	—	—	854	—	—	—	—	—	—	—	—
Social Skills Rating System	—	—	—	—	1,029	775	1,006	—	1,028	752	980	—
Teacher Report Form	—	—	770	—	—	—	1,006	—	—	—	981	—
Unstructured Peer Observation	—	—	—	—	—	—	—	966	—	—	—	—