Published in final edited form as:

Psychophysiology. 2018 May ; 55(5): e13027. doi:10.1111/psyp.13027.

Approaches to Modeling the Development of Physiological Stress Responsivity

J. Benjamin Hinnant, Lauren E. Philbrook, Stephen A. Erath, and Mona El-Sheikh

Auburn University

Abstract

Influential biopsychosocial theories have proposed that some developmental periods in the lifespan are potential pivot points or opportunities for recalibration of stress response systems. To date however, there have been few longitudinal studies of physiological stress responsivity and no studies comparing change in physiological stress responsivity across developmental periods. Our goals were to: (1) address conceptual and methodological issues in studying the development of physiological stress responsivity within- and between-individuals; and (2) provide an exemplar for evaluating development of responsivity to stress in the parasympathetic nervous system, comparing respiratory sinus arrhythmia (RSA) responsivity from middle to late childhood with middle to late adolescence. We propose the use of latent growth modeling of stress responsivity that includes time-varying covariates to account for conceptual and methodological issues in the measurement of physiological stress responsivity. Such models allow researchers to address key aspects of developmental sensitivity including within-individual variability, mean level change over time, and between-individual variability over time. In an empirical example, we found significant between-individual variability over time in RSA responsivity to stress during middle to late childhood but not during middle to late adolescence, suggesting that childhood may be a period of greater developmental sensitivity at the between-individual level.

Keywords

stress responsivity; physiological reactivity; developmental sensitivity; longitudinal analysis; autonomic nervous system; respiratory sinus arrhythmia; childhood; adolescence

Psychophysiological research has provided important insights about the underpinnings of behavior (Porges, 2007; Raine, 2002), markers of psychopathology (Beauchaine, 2001; Beauchaine & Thayer, 2015; Zisner & Beauchaine, 2016), variability in environmental effects (El-Sheikh & Erath, 2011), and physiological consequences of long-term stress exposure (El-Sheikh & Hinnant, 2011; Lupien, McEwen, Gunnar, & Heim, 2009; McEwen & Stellar, 1993). Physiological stress responsivity features heavily in several influential biopsychosocial theories, such as diathesis-stress (Sameroff, 1983) and biological sensitivity to context (Boyce & Ellis, 2005; Ellis & Boyce, 2008; similar to differential susceptibility, Belsky & Pluess, 2009). In general, these biopsychosocial theories posit that some

developmental periods are marked by heightened sensitivity to environmental influences and stand out as points for potential recalibration of stress response systems. During these more malleable developmental periods, physiological responsivity to stress or challenge may be more open to modification or adaptation in response to environmental demands. Despite significant theoretical and empirical progress, the vast majority of relevant studies have been limited to cross-sectional or cross-lagged assessments, neither of which are amenable to assessing the developmental dynamics of physiological stress responsivity. We suspect that the lack of studies on the development of physiological responsivity to stress, indexed by either level of arousal during challenge or change in level from resting state, is due to both statistical and conceptual complexities, key issues we address in this report.

Our two main purposes are to: (1) provide a clear and accessible guide to analyzing repeated measures of physiological responsivity to challenge in order to address questions of within- and between-individual development; and (2) provide an empirical example by evaluating development of stress responsivity in one branch of the autonomic nervous system (ANS) in middle to late childhood and middle to late adolescence. Our first aim addresses an important statistical and methodological issue in biopsychological research – how best to analyze change over time in physiological responsivity to challenge when responsivity often covaries with resting or baseline levels of physiological arousal. Failure to account for resting physiological arousal may be problematic in deriving accurate estimates of change over time in stress responsivity, a key point we expand on below. We propose the use of resting physiological arousal as a time-varying covariate (TVC) in latent growth models of level of physiological responsivity during challenge and physiological reactivity to challenge (a difference, or score) to account for resting state physiological arousal as a potentially significant confound. Throughout we use the term responsivity broadly to include level of arousal during challenge as well as reactivity to challenge (a difference, or score, which is the more common conceptualization in the literature), and the latter two terms more specifically where appropriate. The second aim addresses a key developmental question of physiological responsivity to challenge and its within-individual variability over time (i.e., reliability), mean level change over time (i.e., continuity), and between-individual variability over time (i.e., individual differences in trajectories). We address this question by focusing on and comparing change over time in physiological responsivity at two developmental transitionary periods – middle to late childhood at the transition to adolescence and middle to late adolescence at the transition to early adulthood. In particular, we are interested in whether either time period is marked by significant between-individual variability in change over time, which would suggest heightened phenotypic (i.e., between-individual) plasticity in the development of stress responsivity.

Respiratory Sinus Arrhythmia Responsivity

The ANS is a major psychophysiological component of the stress response system, and the parasympathetic nervous system (PNS) is the regulatory branch of the ANS. The PNS serves as a “brake” (via the vagus nerve) that decelerates heart rate and facilitates calmness, attentional focus, and social engagement under normal circumstances. The deceleration in heart rate produced by higher vagal output to the heart is reflected in heart rate variability across the breathing cycle (e.g., slower heart rate during exhalation than inhalation), which is

referred to as respiratory sinus arrhythmia (RSA; see Berntson, Cacioppo, & Quigley, 1993; Grossman & Taylor, 2007; Porges & Byrne, 1992). Under challenging or threatening conditions, the vagal brake can be withdrawn, yielding an incremental and efficient increase in heart rate and metabolic output that may allow individuals to engage with environmental demands and employ active coping strategies in a regulated manner. Thus, greater reductions in RSA in challenging situations (often termed RSA withdrawal or suppression) should reflect flexible adaptation to environmental demands (Porges, 2007).

A growing body of child and adolescent research has examined RSA withdrawal as a predictor of psychological adjustment or as a moderator or outcome of social-environmental influences (see Beauchaine & Thayer 2015; Zisner & Beauchaine, 2016). These studies have reliably linked RSA withdrawal with fewer internalizing, externalizing, cognitive-academic, and social problems in community samples (Graziano & Derefinko, 2013), although children in clinical samples often exhibit excessive RSA withdrawal (Beauchaine, Gatzke-Kopp, & Mead, 2007). Furthermore, multiple studies have shown that RSA withdrawal moderates associations between environmental stress and behavioral or psychological outcomes, such that RSA withdrawal serves a protective function (reviewed in El-Sheikh & Erath, 2011; Hinnant, Erath, & El-Sheikh, 2015). However, the vast majority of the relevant studies include single-time assessments of RSA, which do not allow for the assessment of developmental sensitivity to stress.

A small number of studies have examined change over time in RSA responsivity to stress in early childhood using growth modeling techniques (Alkon, Boyce, Davis, & Eskenazi, 2011; Conradt et al., 2014; Perry et al., 2013). To our knowledge, no studies have examined RSA responsivity to stress across multiple time points within adolescence, nor have any studies evaluated developmental sensitivity of RSA responsivity to challenge across different periods of the lifespan. Furthermore, longitudinal work to-date has not taken the dependency between resting RSA and RSA responsivity into account by controlling potential time-varying effects of resting activity.

Conceptual Challenges in Stress Responsivity

Sometimes confusion arises from the terminology used to characterize developmental processes that are relevant to the study of physiological stress response systems, both at rest and in response to stress or challenge. Towards enhanced clarity, we will define the key ideas for developmental processes used here in statistical terms and then link these terms to broader conceptual interpretations.

Continuity refers to mean level change over time (Bornstein & Suess, 2000). Continuous development would be indicated by, on average, no significant change over time in stress responsivity for a sample or population. Discontinuous development would be indicated by, on average, significant increases or decreases in stress responsivity over time for a sample or population. Continuous or discontinuous development of stress responsivity addresses a specific question: “As development unfolds during a particular period, on average, do individuals become more or less physiologically responsive to stress?” This and similar questions such as, “Are individuals in developmental period X more or less responsive to

stress than individuals in developmental period Y?” address mean level change or mean differences and are important because they have the potential to answer questions about developmental sensitivity to stress. For example, Alkon and colleagues (2011) showed that, on average, RSA during challenge as well as RSA increased from age six months to five years indicating discontinuous development. Both methods of assessing RSA responsivity indicated that over time the stress response became less robust. These RSA findings were supported by Conradt and colleagues (2013) with a longitudinal sample from age 3 to 6, but contradicted by Perry et al. (2013) who found no mean level change in change in RSA from age 3 to 5. In this report we provide recommendations for answering these questions about the continuous or discontinuous nature of the development of stress responsivity, but in a statistical context that also allows assessment of other aspects of the development of physiological stress responsivity.

However, questions only about mean level change (continuity) would ignore critical aspects of development: the variability within- and between-individuals that allows us to derive estimates of mean level change in the first place. It is critical both statistically and conceptually to distinguish the “types” of variability that may occur in developmental processes, as these distinctions may have significant implications for theory development and testing. Variability over time may be observed *within* individuals (e.g., stress responsivity fluctuates within individuals over time, perhaps partially as a function of changes in environmental contexts). Variability over time may also be observed *between* individuals (e.g., individual differences in trajectories of stress responsivity). The correlation of repeated measures over time between individuals (i.e., maintenance of rank order) has been termed *stability* (Bornstein & Suess, 2000). Notably, in early and middle childhood RSA level during challenge (without subtracting or accounting for baseline levels) does appear to be stable at the between-individual level while RSA appears to be less stable, especially at younger ages (Alkon et al., 2011; Conradt et al., 2013; El-Sheikh, 2005; Perry et al., 2013). “True” *between*-individual variability (i.e., between-individual variability not due to measurement error or random fluctuation) in stress responsivity could be accounted for by between-individual predictors, such as individual differences in exposure to environmental stressors. In the context of latent growth models, described below, analyses focused on accounting for between-individual variability in developmental trajectories are by far the most common.

In addition to conceptualizing *stability* in terms of maintenance of rank order, stability can also be applied to within-individual variability (i.e., the variation between repeated measures within individuals from one time point to another). Within-individual variability is closely linked with the familiar statistical concept of reliability (for a refresher see Shrout & Fleiss, 1979). As with items within a questionnaire scale, adequate reliability is needed for repeated measures within individuals to be valid. Thus, low reliability in repeated measures of physiological responsivity is often a cause for concern. In principle, “true” *within*-individual variability in stress responsivity (i.e., within-individual variability not due to measurement error or random fluctuation) could be accounted for by within-individual predictors. In longitudinal analyses, within-individual predictors are usually called time-varying covariates. Applied to the development of stress responsivity, one would probably hope for a good balance of reliability in repeated measures within individuals while also having some

within-individual variability, particularly if the research question is focused on accounting for within-individual change. We are aware of only one study that has reported on the within-individual stability of RSA level during challenge or RSA, a point we address under Methodological Challenges. Alkon and colleagues (2011) found evidence for moderate stability of RSA during challenge and very low stability of RSA in children assessed from six months to five years of age.

The variability in how physiological systems change or adapt to environmental conditions over time within or between individuals is often referred to as *plasticity* (Debat & David, 2001). By contrast, canalization is indicated by lack of malleability or flexibility in stress responsivity; the system in question has already been “locked in” or set in its patterns of stress responsivity. Canalization would be inferred by lack of within- or between-individual variability, indicating that there is not much variability to predict in statistical models. From both a conceptual and practical statistical standpoint, this means that plasticity could be indicated by either significant variability within individuals (i.e., how stress responsivity changes in a given individual from time point to time point) or significant variability between individuals (i.e., how change over time in stress responsivity differs from individual to individual). Clearly defining and differentiating the focus on within- or between-individual variability may thus help to resolve confusion and lead to more clearly articulated theoretical propositions regarding developmental sensitivity to stress. For example, Belsky and Pluess (2013) make a clear distinction for within- and between-individual variability in their theoretical model of differential susceptibility to stress, labeling between-individual variability *phenotypic plasticity*. Belsky and Pluess emphasize the differences between individuals in developmental trajectories, and thus also point to a key avenue for exploration: comparisons of between-individual variability in stress responsivity over time at different developmental periods. By extension, lack of between-individual variability could be taken as evidence of phenotypic canalization. Of the three studies to evaluate change in RSA responsivity to challenge, two found evidence for significant between-individual variability in the development of RSA in early childhood (Conradt et al., 2013; Perry et al., 2013) and one did not report between-individual variability (Alkon et al., 2011). In our empirical example, we emphasize between-individual variability over time and its implications for developmental sensitivity to stress at different developmental periods.

To summarize and give examples of these important developmental concepts outside the context of physiological responsivity, we could consider within-individual variability, mean level change, and between-individual variability as applied to change over time in planning ability as measured by the Tower of London task in different developmental periods. As compared to adolescence, we would expect that repeated measures of planning ability would have lower reliability in childhood, perhaps reflecting changes in planning strategies that parallel cognitive development. We would probably expect to see greater mean level increases in children’s planning ability over time (i.e., discontinuity), as planning abilities are fairly mature by late adolescence (Albert & Steinberg, 2011). Finally, we might expect to see greater between-individual variability over time in children as compared to adolescents due to individual differences in brain maturation and myelination of neuronal connections needed for effective planning that are largely complete by late adolescence (Casey, Tottenham, Liston, & Durston, 2005).

Methodological Challenges in Stress Responsivity

Law of initial values

Why have there not been more studies assessing change over time in physiological stress responsivity? Aside from the practical challenges of collecting multiple waves of physiological data, we think that the answer to this question might be attributed to the Law of Initial Values (LIV). The LIV states that physiological resting states often impose limits on physiological responsivity to challenge (Wilder, 1967). These biological limitations may be thought of as “soft” floor or ceiling effects. By soft we mean floor or ceiling effects that may vary both within individuals (e.g., maximum heart rate may change within individuals over time as a function of age and biological maturation) and between individuals (e.g., some individuals have a higher maximum heart rate due to physical conditioning and genetic factors), but exist for the population. The practical effect of the LIV is that resting level activity of physiological systems often covary with challenge levels or challenge reactivity (Wilder, 1967). This may create a potential confound in the interpretation of stress responsivity level or reactivity scores: Where you start (i.e., level at rest) may limit where you can go (i.e., level of responsivity to stress), particularly for those near the biological floor or ceiling. Thus, it is worth considering the statistical options for accounting for the LIV phenomenon. Burt and Obradovi (2013) provide an excellent overview of the LIV, including how the LIV may be confounded with measurement error and regression to the mean. Treatment of these particular types of potential confounds are outside of the aims of this paper, but the concept of the LIV is central to accurately modeling change over time in physiological stress responsivity. Specific to our example, we show that high resting RSA (i.e., RSA levels nearer the ceiling) is associated with greater RSA withdrawal (i.e., a decrease in RSA in response to challenge), a finding that is common in the literature (Graziano & Derefinko, 2013). In the three exemplar studies reviewed (Alkon et al., 2011; Conratt et al., 2013; Perry et al., 2013) none controlled for time-varying effects of resting RSA even though resting levels were frequently associated with RSA during challenge and RSA. In the case that resting and responsivity measures are associated, whether it is necessary and how to take these dependencies into account in repeated measures data when both vary across time points become important questions.

Classic approaches to the LIV

There are two classic approaches to operationalizing change in physiological arousal in response to stress: the difference score and the residual score. Burt and Obradovi (2013) also provide a wonderful treatment of these approaches, so we will only briefly review the procedures, advantages, and disadvantages of each. The difference score (also called the change score or Δ score) is intuitively appealing and provides a clear interpretation of stress responsivity. It is usually calculated as resting state or baseline measurement of physiological arousal subtracted from physiological arousal during a task or challenge. In the measurement of RSA for example, a negative difference score indicates RSA withdrawal (decreases) in response to challenge and a positive difference indicates RSA augmentation (increases) in response to challenge.

The residual score approach and its interpretation are less intuitive. The residual score is the individual level deviation of challenge physiological arousal from expected challenge physiological arousal, accounting for individual deviation of resting level from expected resting level and the regression of physiological arousal during challenge on physiological arousal during rest. Simply put, the residual score is used to measure individual deviation in challenge scores from expected scores, accounting for resting level. However, the definition of expected scores is specific to the sample in question, making cross study comparisons or even cross time point comparisons difficult (see Burt & Obradovi, 2013 for additional details). There is a longstanding debate over the relative utility and reliability of difference scores and residual scores (e.g., Edwards, 2001; Rogosa & Willett, 1983; Rogosa, Brandt, & Zimowski, 1982; Stemmler, 1987), though the resolution may be sample specific and dependent upon the correlation between resting and challenge physiological arousal (Williams, Zimmerman, & Mazzagatti, 1987). In practice, difference scores and residual scores tend to be highly correlated (r 's of around .9 in this study). Our purpose is not to advocate for one method over another, but we do emphasize that, regardless of the approach, the repeated measures should have adequate reliability (i.e., stability; Singer & Willett, 2003). In the context of repeated measures, this means that scores within individuals are “nested” or dependent upon one another. Reliability of repeated measures can be empirically tested using the intraclass correlation coefficient (ICC), a statistic that is interpreted as the expected correlation between any two randomly chosen scores within an individual (Heck & Thomas, 2015). Although there are no strict rules for adequate stability of repeated measures, Heck and Thomas give the very liberal guideline of .05, though much higher intraclass correlation coefficients are common with repeated measures data. In either case of difference scores or residual scores, the purpose is usually to characterize mean levels of stress responsivity to challenge at the group level to make between-group comparisons or to use between-individual variability in scores as a predictor or outcome. Notably, if researchers use *standardized* residual scores that can be estimated and saved as new variables, which is an option in most statistical programs, these scores will have the same mean and variance at each time point which makes them a highly inadvisable choice for longitudinal analysis (Singer & Willett, 2003). In the repeated measures of RSA responsivity examples that we provide, we emphasize the TVC of resting RSA to account for potential time-varying associations between resting RSA and RSA responsivity (discussed below).

Newer ways to address the LIV

Latent growth modeling (LGM) is a type of structural equation modeling that is capable of addressing questions of within-individual variability via covariances among repeated measures, mean level change via intercept and slope estimates, and between-individual variability via intercept and slope variances. LGM can also account for time-varying associations between physiological arousal at rest and during stress responsivity measures (for an example applied to change in stress responsivity across an experimental task see Kristjansson, Kircher, & Webb, 2007). Because there are numerous excellent resources for latent growth modeling (e.g., Curran, Obeidat, & Losardo, 2010; McArdle, 2009; Singer & Willett, 2003), we provide only a brief overview of this statistical method, but do provide example syntax files in Appendix A. LGMs are a special case of structural equation models in which latent (i.e., unobserved) variables are used to account for variance in observed

repeated measures. User-defined regression coefficients (i.e., factor loadings) of the observed repeated measures on (usually) two latent variables define the latent intercept (e.g., starting point or levels at Time 1) and (usually) a latent linear slope (i.e., linear change over time). Importantly, the user's choice of values to assign to factor loadings is important for the interpretation of the effect of time on the observed repeated measures (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). As an example, in the current study we have three equally spaced time points of measurement for each developmental period. We used factor loadings of 1, 1, 1, for the latent intercept and 0, 1, 2 for the latent slope, which functions to define the latent intercept as physiological stress responsivity at Time 1 and the latent slope as linear change over time in physiological stress responsivity. LGMs are highly flexible in that they take into account variability within individuals and between individuals in defining any given model. The general formula for the within-individual component of a LGM is:

$$y_{it} = \alpha_i + \beta_i \lambda_t + \varepsilon_{it}$$

y_{it} is the score of individual i at time t as a function of individual i 's intercept (α), linear slope (β) as a function of time (λ), and the residual (ε) of individual i at time t . Thus, each individual has his or her own intercept and developmental trajectory. A between-individual component of a LGM is represented by the latent intercept and slope variables that each have a mean and variance. The general formulas for the latent intercept and slope are:

$$\begin{aligned}\alpha_i &= \mu_\alpha + \zeta_{\alpha i} \\ \beta_i &= \mu_\beta + \zeta_{\beta i}\end{aligned}$$

Expanding this equation yields one in which individual i 's score at time t is a function of the sample intercept (μ_α), individual deviation from the intercept ($\zeta_{\alpha i}$), sample slope (μ_β) multiplied by the effect of time (λ_t), individual slope deviation ($\zeta_{\beta i}$) multiplied by the effect of time, and residual:

$$y_{it} = \mu_\alpha + \zeta_{\alpha i} + \mu_\beta \lambda_t + \zeta_{\beta i} \lambda_t + \varepsilon_{it}$$

Adding a TVC to account for resting levels that may vary over time, defined as individual i 's effect of resting level at time t ($\gamma_t baseline_{it}$) is done at the within-individual level (Curran & Willoughby, 2003), but is included in the full equation:

$$y_{it} = \mu_\alpha + \zeta_{\alpha i} + \mu_\beta \lambda_t + \zeta_{\beta i} \lambda_t + \gamma_t baseline_{it} + \varepsilon_{it}$$

From a functional point of view, an LGM with a TVC can be used to derive estimates of initial levels (intercept) and growth (slope) in an outcome, controlling for a covariate that may have different effects at each time point of measurement. An example of one of our analytical models for assessing change over time in RSA responsivity to challenge, controlling for resting RSA at each time point, is presented in Figure 1.

Importantly, multilevel models (MLMs) are also capable of addressing these same questions of within-individual variability, mean level change, and between-individual variability and are similar enough to LGMs that one should derive the same results from either model assuming they are specified in the same manner (Curran, 2003). We elected to use LGMs in the current paper because they provide model fit indices that are useful for comparing non-nested models that are not available in MLMs. In Appendix A, we provide Mplus syntax for both LGMs and MLMs to assess change over time in physiological responsivity to stress, accounting for time-varying resting physiological arousal. One difference worth noting is that in LGM users will see effects of time-varying effects of resting level on responsivity for each time point in their output, whereas in MLM users will see a single estimate of the time-varying effect of resting level on responsivity.

Our study's first aim is to illustrate two ways of analyzing change over time in physiological responsivity to stress, operationalized through repeated measures of RSA, using: (1) repeated measures of RSA level during challenge, and (2) repeated measures of RSA in response to challenge. In both examples, we evaluate models that omit or control for the TVC of resting RSA to determine if modeling this potential confound is needed in general or may be sample-specific. Thus, we present four sets of growth models for each developmental period: RSA level during challenge, RSA level during challenge corrected for resting state, RSA in response to challenge, and RSA in response to challenge corrected for resting state. Our second aim is to investigate the development of physiological responsivity to stress in two developmental periods, middle childhood and middle adolescence. More specifically, within and across these developmental periods, we address within-individual variability, mean level change, and between-individual variability in RSA responsivity to challenge.

Method

Participants

Participants were part of the Family Stress and Youth Development: Bioregulatory Effects project, a longitudinal study of family stress and child development. Children and their families were recruited from area public elementary schools in 2005. To be eligible for the study children needed to be from two-parent homes and have no diagnosis of a learning disability, chronic illness, or sleep disorder. The study consisted of six waves of data collection, three of which took place in middle childhood and three in middle adolescence. Two hundred fifty-one children (51% female) and families participated in the study at the first time point (T1). Of these children, 86% (217) participated at T2 and 78% (194) at T3. On average, children were 8.22 ($SD = .71$), 9.31 ($SD = .78$), and 10.26 ($SD = .78$) years old across the three waves, respectively. Sixty-four percent of the sample was European American and 36% African American, representative of the community from which the sample was drawn. Hollingshead Index (Hollingshead, 1975) indicated that families represented a wide of socioeconomic backgrounds ($M = 3.21$; $SD = 0.91$; range: 1–5). Median family income was between \$35000–50000. Families who dropped from the study did not differ from the original sample on any demographics.

At T4, an additional 53 adolescents were recruited into the study from the same school districts using identical criteria. There were no demographic differences between those families recruited at T4 vs. T1. Information regarding family size was also available at T4 and was used to calculate income-to-needs ratio (annual family income divided by the poverty threshold with respect to family size; U.S. Department of Commerce, 2013). Family income-to-needs ratio indicated that 14% of participant families were living below the poverty line (ratio < 1), 28% were living near the poverty line (ratio > 1 and < 2), 22% were lower middle class (ratio > 2 and < 3), and 36% were middle class (ratio > 3). In total, a sample of 246 adolescents participated in the study at T4, 93% (229) of whom remained at T5 and 89% (219) who participated at T6. Youth were 15.78 ($SD = .80$), 16.78 ($SD = .77$), and 17.69 ($SD = .75$) years old on average at the three waves. Families who dropped from the study at waves 5 or 6 did not differ from the T4 sample, with the exception that families with higher incomes were more likely to remain in the study at T6. Youth were scheduled to participate in the study at approximately the same time during the school year across waves to minimize the effects of variability in assessment intervals on the physiological data. The mean interval between waves was 12.0 months ($SD = 2.4$ months).

In total, useable RSA data were available for 232, 187, 173, 222, 206, and 187 youth across the six waves, respectively. Full information maximum likelihood estimation (FIML) was used in the analyses to account for missing RSA data. The range of missingness (8–15%) is well-within the amount acceptable for FIML, which has been shown to produce the least biased estimates and fewest Type I errors in comparison to other methods (Enders & Bandalos, 2001).

Procedure

All study procedures were approved by the university's institutional review board and were identical across waves unless otherwise noted. Following obtainment of parental consent and youth assent in our lab, a research assistant placed three disposable pediatric snap electrocardiogram (ECG) electrodes on the participant's chest in a lead-II configuration to collect cardiac information. Data were collected using hardware made by the James Long Company (Caroga Lake, NY) for the first three waves of the study, and using the Mindware BioNex 8-slot chassis acquisition system (Mindware Technologies, Inc., Gahanna, OH) for the second three waves. The James Long hardware system also includes a pneumatic bellows that was placed around the chest to measure respiration. The Mindware hardware system collects thoracic impedance (Z_0) data to derive respiration (Ernst, Litvack, Lozano, Cacioppo, & Berntson, 1999). To collect impedance data, voltage electrodes were placed at the apex and base of the thorax, and dual electrodes were placed on the back 1.5 inches above and below the voltage electrodes in a four spot impedance electrode configuration (Berntson & Cacioppo, 2004).

Once the equipment was placed, youth spent 6–9 minutes sitting quietly. The first 3–6 minutes of this period served as acclimation time to the laboratory (6 minutes at T1, 3 minutes at all other waves); the last 3 minutes were used to obtain a measure of arousal at rest. This length of acclimation is appropriate for obtaining a valid assessment of resting physiology (Zisner & Beauchaine, 2016). Following the resting period youth completed a

series of laboratory tasks. RSA reactivity to a cognitive challenge task was of interest in the current investigation because this task was administered at all waves of the study. Parents completed questionnaires regarding participants' demographic information.

Measures

Challenge task—To facilitate cross-time comparisons, youth completed the same cognitive challenge task at each wave. The task consisted of tracing a star on a sheet of paper using only the star's mirrored reflection as a guide (Mirror Tracer, Lafayette Instrument Company, Lafayette, IN) in 3 minutes. This places attentional demands on the participant, elicits frustration, and has been shown to evoke RSA reactivity in youth (e.g., Allen & Matthews, 1997; Staton, Hinnant, Buckhalt, & El-Sheikh, 2014). Additionally, there is likely some influence of motor activity on the challenge RSA values (Bush, Alkon, Obradovi, Stamperdahl, & Boyce, 2011) because tracing the star requires minor movements of the arm.

RSA—RSA indicates the degree of variability in heart rate during the respiration cycle. We followed standard guidelines for deriving RSA from the cardiac and respiratory data (Berntson et al., 1997). For the first three waves, the James Long Interbeat Interval (IBI) Analysis System software was used to identify R-waves for calculating RSA. Misidentified waves were manually corrected. RSA was calculated using the peak-to-valley method and the unit of measurement was seconds. However, because the very small numbers led to problems with model convergence during data analysis, all values were multiplied by 100. For the second three waves, Mindware's analysis system (Heart Rate Variability Version 3.0.21; Demaree, Pu, Robinson, Schmeichel, & Everhart, 2006) was used to calculate RSA. The Mindware software package filtered the data, after which project staff visually inspected it. Artifacts and missing R peaks were manually corrected using the most probable interbeat intervals as a guide. RSA was quantified using spectral analysis as the natural log of the variance of heart period within the high-frequency band of respiration (.15–.40 Hz), following the procedures outlined by Berntson et al. (1997), and the unit of measurement was milliseconds squared. Even with the differences in scaling, the peak-to-valley and spectral methods of calculating RSA have been shown to be nearly equivalent with within-individual correlations approaching 1.0 (Grossman, van Beek, & Wientjes, 1990; Lewis, Furman, McCool, & Porges, 2012). Notably, however, because of the differences in analysis and scaling, the calculated means are not comparable across the two systems. At all waves, RSA was averaged across the baseline period and star-tracing task to create mean individual scores for these two conditions.

Results

Plan of Analysis

Following preliminary analyses, two nested LGMs were evaluated for each repeated measures index of RSA responsivity (RSA level during challenge and RSA), for each developmental period. In the first model, repeated measures of resting RSA were included but time-varying effects on RSA responsivity were constrained to be zero. In the second model, time-varying effects of resting RSA on RSA responsivity were freely estimated. Relative fit indices for model comparisons included AIC, BIC, and sample size adjusted BIC

(SSA BIC). Lower values for these indices relative to a comparison model indicate better fit. Formal model comparisons were also conducted through χ^2 difference tests. Overall model fit indices included RMSEA, SRMR, and CFI. Standard practices were used in determining whether each model provided an adequate fit to the data (Schreiber, Nora, Stage, Barlow, & King, 2006). Resting RSA was grand mean centered at each time point, which yields estimates of RSA responsivity intercepts and slopes when resting RSA is at the average value at each time point.

Preliminary Analyses

Means and standard deviations as well as bivariate correlations for resting RSA, challenge RSA, and RSA at each wave in childhood and adolescence are provided in Tables 1 and 2, respectively. All variables were normally distributed; the most extreme estimate of skew was 2.1. Because only one variable (RSA at T1) exhibited this level of skew, and transforming this measure would have required transforming all other repeated measures of RSA and therefore changing the metric of the model, we retained the original values. Of note, FIML has been found to be fairly robust to moderate departures from normality of indicator variables (Enders, 2001). On average, participants tended to suppress RSA in response to the star-tracing challenge at all time points. Generally, variables were moderately to highly correlated across waves, although there were fewer correlations between the RSA reactivity variables in middle childhood than during middle adolescence, suggesting lower between-individual stability at the earlier developmental period.

Development of RSA Level during Challenge

Table 3 displays estimates of model fit, within-individual variability (based on the ICC estimate), mean level change (intercept and slope estimates), and between-individual variability (intercept and slope variances) of RSA level during challenge in middle childhood and middle adolescence.

Model fit—Fit indices for model 1 in childhood in which regression coefficients between repeated measures of RSA level and resting RSA were constrained to be zero (i.e., a model positing no relationship between repeated measures of resting RSA and RSA level during challenge) were either poor or borderline acceptable. Fit indices were much better in model 2 where the regression coefficients were freely estimated, and model 2 was a significant improvement over model 1 as indicated by the χ^2 difference test. In testing the LGM for middle adolescence, we estimated a negative variance for the latent slope (an impossible and inadmissible solution). There are two common approaches to resolving this type of error in a LGM; one is to constrain the variance of the slope to zero and the other is to constrain the residuals of the observed repeated measures indicator variables to be equal. Given our focus on between-individual variability in developmental trajectories, we decided to constrain the residuals of the indicators of RSA responsivity to be equal, which is the cause of the two additional degrees of freedom in the models of middle adolescence relative to middle childhood. Importantly, this constraint was needed for all models of RSA responsivity in middle adolescence. Fit indices for models 1 and 2 in middle adolescence were generally acceptable. Relative comparisons of AIC, BIC, and SSA BIC indicated that model 2 fit a little more poorly, but the χ^2 difference test showed that the two models were not

significantly different. This indicated that controlling for the TVC of resting RSA did not significantly improve the fit in modeling development of RSA level during challenge in adolescence. Thus, model 2 controlling for time-varying effects of resting RSA fit best for the development of RSA level during challenge in middle childhood and in middle adolescence the more parsimonious model 1 (i.e., not controlling for the TVC of resting RSA) was selected.

Within-individual variability—Intraclass correlation coefficients (ICC), interpreted as the expected within-individual correlation between RSA level during challenge at any two random time points, were used to evaluate reliability. ICC estimates were high in both childhood and adolescence (.44 and .68, respectively), indicating that during childhood 56% of the total variance in RSA level during challenge was within individuals and 44% was between individuals while during adolescence 32% was within individuals and 68% was between individuals. There was considerable reliability in RSA level during challenge at both developmental periods.

Mean level change—Mean level change was assessed by the sample level estimates of the latent intercepts and slopes in childhood and adolescence. In both childhood and adolescence the intercepts for RSA level during challenge, interpreted as RSA level during challenge at the first time points of measurement during the respective developmental periods, were significantly different from zero. In childhood, the latent slope was not significantly different from zero. In adolescence, the latent slope was significant and negative, indicating that on average RSA levels during challenge declined over time. Thus, RSA level during challenge in middle childhood could be described as continuous in that there was no significant group level change over time. However, in middle adolescence, RSA level during challenge was discontinuous as indicated by a mean level decrease over time.

Between-individual variability—Between-individual differences in initial levels and rates of change over time in RSA during challenge were evaluated through variance estimates of the intercepts and slopes. In middle childhood, there was significant between-individual variability in both the latent intercept and slope. This indicates that individuals varied in their initial levels of RSA during challenge and in their change over time in RSA during challenge, controlling for time-varying effects of resting RSA. By contrast, in middle adolescence there was significant between-individual variability in the latent intercept but not the latent slope, indicating that individuals varied in their initial levels of RSA during challenge but not in how RSA during challenge changed over time.

In summary, we found that after controlling for resting RSA, RSA during challenge in middle childhood exhibited a great deal of both within- and between-individual variability but no mean level change. In middle adolescence, RSA during challenge also exhibited a good deal of within-individual variability, but there was no significant between-individual variability in linear change over time, and, on average, decreased over time. This comparison of between-individual variability in developmental periods is emphasized in the Discussion. Figures 2A and 2B depict change over time in RSA level during challenge along with 95% confidence intervals for RSA during challenge at each time point. Resting RSA is overlaid in

these figures as a reference to challenge levels. To aid in this reference, the latent linear slope of RSA level during challenge in middle childhood was positive but not significant ($B = .37$, $SE = .25$, $p = .12$) while a post-hoc supplementary model of growth of resting RSA shows that the latent slope was marginally significant and positive ($B = .50$, $SE = .28$, $p = .07$). The latent linear slope of RSA level during challenge in middle adolescence was significant and negative ($B = -.09$, $SE = .04$, $p = .02$) while the post-hoc model of resting RSA had a very flat and nonsignificant latent slope ($B = -.01$, $SE = .04$, $p = .93$).

Development of RSA

Estimates of model fit, within-individual variability, mean level change, and between-individual variability of RSA in middle childhood and middle adolescence are shown in Table 4.

Model fit—Fit for model 1 in middle childhood was unacceptable as assessed by any measure. Fit in model 2 was much improved as measured by all nested model comparisons and the χ^2 difference test and was selected as the best fitting. Model 2 controlling for time-varying effects of resting RSA was also selected over model 1 in middle adolescence for the same reasons. Notably, in middle adolescence absolute fit indices of SRMR and CFI in model 2 were acceptable, but RMSEA was poor.

Within-individual variability— RSA scores in middle childhood had low reliability ($ICC = .07$), though this estimate does surpass the general rule of thumb that an ICC of .05 or greater is needed to predict meaningful individual differences (Heck & Thomas, 2015). However, this does indicate that there is limited nesting or dependence of repeated measures of RSA within individuals during this developmental period. During middle adolescence there was much higher reliability in RSA ($ICC = .42$). As noted above, the ICC can also be interpreted as the proportion of variance that is between individuals. Thus, in middle childhood 93% of the total variance in RSA was within individuals as compared to 58% of the total variance being within individuals in adolescence.

Mean level change—Controlling for the TVC of resting RSA in each model, the latent intercepts of RSA were significantly different from zero in both middle childhood and middle adolescence. Additionally, in both developmental periods the slopes were not significantly different from zero, indicating that RSA did not show significant change over time.

Between-individual variability—Despite the fact that most of the total variance in RSA middle childhood was within individuals, there was significant between-individual variability in both the latent intercept and slope. This indicates there are important individual differences in initial levels of RSA and change over time in RSA. By contrast, in middle adolescence there was significant variability in the latent intercept but not the latent slope.

To summarize, in childhood RSA exhibited a high amount of within-individual variability, significant between-individual variability over time, but no significant mean level change over time. In adolescence, RSA also exhibited a high degree of within-individual variability, but there was not significant between-individual variability in its change over

time, nor was there significant mean level change over time. Figures 3A and 3B depict change over time in RSA in childhood and adolescence with 95% confidence intervals shown at each time point.

One caveat worth noting for the growth model of RSA in adolescence is that a selected model controlling for the TVC of resting RSA still had poor absolute fit indices. We present the model “as is” for consistency with the parameterization of the other models but note that poor model fit is something that should typically be addressed, as it indicates a mismatch between the covariance matrix specified by the researcher and the true covariance matrix of the data. In Appendix B we present a model that constrains the estimated covariances between the observed repeated measures of resting RSA and the latent intercept and slope of RSA to zero. Although the absolute fit index of RMSEA was improved, the χ^2 difference test showed that this respecified model with constraints was not a significantly better fit. It is also worth noting that the parameters of interest did not differ substantially between these two models.

Discussion

Our goals in this study were to (1) contribute to methodology in research on physiological stress responsivity by illustrating options for analysis of change over time in these outcomes; and (2) investigate change over time in physiological responsivity to challenge in two developmental periods, middle childhood and middle adolescence. Our specific measure of physiological responsivity to stress was RSA, assessed with both RSA level during challenge and RSA in response to challenge. We addressed translation of important concepts of developmental sensitivity to analytic models by comparing within-individual variability, mean level change, and between-individual variability in physiological responsivity to challenge across two developmental periods while evaluating the need to control for time-varying effects of resting physiological arousal.

Latent growth models evaluating the need to include time-varying covariates of resting RSA were used to evaluate change over time in RSA responsivity and to illustrate how such models can be applied to other developmental time periods or indices of physiological responsivity. We found that intraclass correlations, an indicator of within-individual stability and reliability, for RSA during challenge were relatively high in both childhood and adolescence. Consistent with previous work (Alkon et al., 2011), RSA was much less reliable in childhood but reliable in adolescence. In other words, during middle childhood the repeated measures of RSA were not highly dependent (i.e., nested) within individuals, and there was very high within-individual variability. This instability runs contrary to a basic premise of longitudinal data analysis that there is a dependence or nesting of the repeated measures within individuals. Thus, some caution is warranted in interpreting findings for change over time in RSA in childhood.

In middle childhood there was not significant mean level change in RSA during challenge over time (continuous development), whereas in middle adolescence RSA during challenge decreased over time (discontinuous development). Further, RSA over time showed no mean level change in either childhood or adolescence. This point addresses an important

question in the study of stress reactivity: Are children in different developmental periods becoming more or less reactive to challenge over time? Looking across the analyses, three of four models indicated no mean level change in RSA responsivity. This pattern of results is similar to those found in one study of early childhood (Perry et al., 2013), but contradicts findings of two other studies indicating discontinuous development (Alkon et al., 2011, Conradt et al. 2014). This suggests that the most important changes may be at the individual rather than aggregate level, discussed below.

We think a very interesting question concerns between-individual variability over time, which seems to be most clearly related to theories of developmental sensitivity to stress (Belsky & Pluess, 2013; Ellis & Del Giudice, 2014; Hostinar & Gunnar, 2013; Juster, McEwen, & Lupien, 2010). Between-individual variability indicates that a trait or process is theoretically malleable at the phenotypic level, which describes variability in the population. While it is possible that individuals change at different rates for no detectable reason, often the goal of developmental researchers is to account for *why* individuals change in different ways and at different rates by using predictors of between-individual variability of change over time. Looking across the results for the two developmental periods, a clear pattern of developmental sensitivity emerges: regardless of the index of RSA responsivity, in middle childhood there is significant between-individual variability in RSA responsivity over time, replicating other work in early childhood (Conradt et al., 2014; Perry et al., 2013). In middle adolescence, RSA responsivity to a problem-solving, frustrating task over time appears to be less variable between individuals and may be thought of as canalized or developmentally insensitive at the phenotypic level. However, these findings are in relation to responsivity to one of many types of laboratory challenges used by researchers, a point addressed under our study limitations.

Notably, it is important to distinguish between individual differences in RSA responsivity (or any measure of physiological responsivity) within a given time point and individual differences in change over time in RSA responsivity. Regardless of the developmental period, at any given time point there may be significant between-individual variability in RSA responsivity, but only in middle childhood did we find that individuals vary in how RSA responsivity changes over time. This finding has implications for biopsychosocial theories of the development of physiological responses to stress, as it may indicate that by adolescence stress responsivity is less variable over time and perhaps less affected by environmental influences at the phenotypic level. As suggested by El-Sheikh and Erath (2011), early experiences may directly influence child physiology, but in adolescence as physiological patterns become more stable, physiology may be more likely to be a moderator of contextual influences on behavior.

In addition to addressing key developmental concepts and demonstrating analytic methods for assessing developmental changes in physiological responsivity (also see Appendix A for Mplus syntax for analytic models), we propose a few recommendations that may help researchers to address developmental questions about physiological responsivity with their own data:

1. Should I use level of physiological responsivity during challenge or reactivity scores (scores) in response to challenge as repeated measures?

Conceptually, these are slightly different questions and either can be appropriate. If evaluating change over time in level of physiological responsivity, it may be useful to show resting levels as a comparison. Practically, assessing reliability of the repeated measures as well as model fit indices is important. Unreliable data or poor fitting models may end up influencing the choice of physiological responsivity measurement. Researchers should include discussion of these issues where appropriate. As noted by Burt and Obradovi (2013) and others, scores may be especially likely to exhibit low reliability (i.e., low within-individual correlations of repeated measures). Based on the findings here and on theoretical rationale, it is possible that this will be more of an issue at earlier developmental time periods before physiological stress response systems have stabilized. In our example, we found that in middle childhood reliability of RSA level during challenge was high but reliability of RSA was low. In general, model fit for RSA growth models were also poorer than models of RSA level to challenge.

2. Should I control for resting state levels of physiological arousal?

The implication of the Law of Initial Values is that biological floor and ceiling effects create covariances between resting state levels of arousal and levels of physiological responsivity. Examination of the data can test whether this is the case in specific samples. Bivariate between-individual correlations are a good place to start. In more complex growth models, comparing fit of nested models that estimate or constrain regression paths from repeated measures of resting levels to repeated measures of physiological responsivity will be helpful. In our example, we found that in three out of four cases the best fitting model included controlling for the TVC of resting RSA.

Limitations

One limitation to the current study is that equipment for physiological data collection and the software for processing that data differed in childhood and adolescence. Although the algorithms used to calculate RSA are similar and the resulting estimates are highly correlated (Grossman et al., 1990; Lewis et al., 2012), the scaling is different. While we can say, for example, that a latent slope is significant at one developmental period but not another, this measurement issue does make direct comparisons between developmental periods less than ideal and formal tests of equivalence impossible. Another limitation is that data from the two developmental periods were, for the most part, drawn from the same sample. In other words, the majority of the repeated measures in childhood and adolescence were nested within the same individuals. The lack of equivalent measurement in the two developmental periods prevented us from treating the data as six repeated measures rather than two separate sets of three repeated measures or from constraining parameters to be equal between developmental periods.

We evaluated physiological responsivity to a specific challenge, a star-tracer task that, while validated for eliciting stress response system activity and feelings of frustration, is still only

one task of many possible selections. Change over time in physiological responsivity may be dependent upon the type of task. For example, while we did not find evidence for between-individual variability in physiological responsivity over time to a problem-solving, frustrating task in adolescence, during this time period ANS reactivity to social stressors such as peer rejection may be highly plastic at the phenotypic level. Furthermore, it is important to note that our study did not include a baseline period that accounted for motor activity, which would ideally be used to remove any influence of motor activity from the star-tracing task on the challenge RSA values (Bush et al., 2011). However, we believe that the use of a consistent task across time reduces the possibility that some time points were more affected by motor activity than others.

Future Directions and Conclusion

Based on the limitations of the study, important future steps are replication and extension of these findings. Replication is a keystone of scientific progress and we hope that another research group will evaluate the development of stress responsivity, such as within- and between-individual variability and mean level change in RSA over time with other samples. Important extensions of this work might include evaluating change over time in other developmental periods, in response to other types of challenge tasks, or in other physiological systems critical for adaptive stress responsivity.

Another important application of these modeling approaches for evaluating developmental of stress responsivity is in theory testing. As noted above, while we only included within-individual time-varying effects of resting arousal to account for the LIV, the inclusion of other within- and between-individual predictors could be a very potent tool for theory refinement. Although we paint the relevant theories in broad strokes here, consider this example of theory testing. A diathesis-stress model would likely propose a linear relationship between environmental stress and change over time in stress responsivity wherein high levels of stress “burn out” functionality of stress response systems (Evans & Kim, 2007; Juster, et al., 2010). The competing biological sensitivity to context model proposes a nonlinear quadratic (u-shaped) relationship between stress and stress responsivity in which both low *and* high levels of stress promote heightened responsivity and moderate levels of stress result in lower levels of responsivity (Boyce & Ellis, 2005; Ellis & Boyce 2008). Finally, the adaptive calibration model of stress responsivity posits a nonlinear cubic (wave-shaped) relationship between stress and stress responsivity in which low stress promotes heightened “sensitive” responsivity, moderate stress promotes moderate “buffered” responsivity, high stress promotes heightened “vigilant” responsivity, and severe stress promotes blunted “unemotional” responsivity (Del Giudice, Ellis, & Shirtcliff, 2011). These competing hypotheses could be tested longitudinally, controlling for time-varying effects of resting arousal if needed, with the methods presented here by including key stress variables as linear, quadratic, and cubic effects on between-individual intercepts and slopes of stress responsivity. Notably, all of these theories are implicitly or explicitly phrased to address questions of phenotypic plasticity (i.e., between-individual differences in stress responsivity over time). To the best of our knowledge, there are no published papers testing theories of within-individual plasticity in stress response systems which would, of necessity, utilize within-individual predictors. We think that this work will help researchers to think about and

answer questions about when physiological responsivity to challenge stabilizes and transitions from being malleable (at either the within- or between-individual level) to being more canalized and resistant to influence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by Grant R01-HD046795 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development awarded to Mona El-Sheikh. We wish to thank our research laboratory staff, particularly Bridget Wingo, for data collection and preparation, as well as the children and parents who participated. The authors have no conflicts of interest to declare.

References

- Alkon A, Boyce WT, Davis NV, Eskenazi B. Developmental changes in autonomic nervous system resting and reactivity measures in Latino children from 6 to 60 months of age. *Journal of Developmental & Behavioral Pediatrics*. 2011; 32(9):668–677. <https://doi.org/10.1097/DBP.0b013e3182331fa6>. [PubMed: 22008788]
- Allen MT, Matthews KA. Hemodynamic responses to laboratory stressors in children and adolescents: The influences of age, race, and gender. *Psychophysiology*. 1997; 34(3):329–339. <https://doi.org/10.1111/j.1469-8986.1997.tb02403.x>. [PubMed: 9175447]
- Albert D, Steinberg L. Age differences in strategic planning as indexed by the Tower of London. *Child Development*. 2011; 82(5):1501–1517. <https://doi.org/10.1111/j.1467-8624.2011.01613.x>. [PubMed: 21679178]
- Beauchaine TP. Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Developmental Psychopathology*. 2001; 13:183–214. <https://doi.org/10.1017/S0954579401002012>.
- Beauchaine TP, Gatzke-Kopp L, Mead HK. Polyvagal Theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology*. 2007; 74(2):174–184. <https://doi.org/10.1016/j.biopsycho.2005.08.008>. [PubMed: 17045726]
- Beauchaine TP, Thayer JF. Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology*. 2015; 98(2):338–350. DOI: 10.1016/j.ijpsycho.2015.08.004 [PubMed: 26272488]
- Belsky J, Pluess M. Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*. 2009; 135(6):885–908. <https://doi.org/10.1037/a0017376>. [PubMed: 19883141]
- Belsky J, Pluess M. Beyond risk, resilience, and dysregulation: Phenotypic plasticity and human development. *Development and Psychopathology*. 2013; 25:1243–1261. <https://doi.org/10.1017/S095457941300059X>. [PubMed: 24342838]
- Berntson GG, Bigger JT Jr, Eckberg DL, Grossman P, Kaufmann PG, Malik M, ... van der Molen MW. Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*. 1997; 34:623–648. <https://doi.org/10.1111/j.1469-8986.1997.tb02140.x>. [PubMed: 9401419]
- Berntson, GG., Cacioppo, JT. Heart rate variability: Stress and psychiatric conditions. In: Malik, M., Camm, AJ., editors. *Dynamic Electrocardiography*. New York, NY: Blackwell/Futura; 2004. p. 57-64.
- Berntson GG, Cacioppo JT, Quigley KS. Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*. 1993; 30(2): 183–196. <https://doi.org/10.1111/j.1469-8986.1993.tb01731.x>. [PubMed: 8434081]

- Biesanz JC, Deeb-Sossa N, Papadakis AA, Bollen KA, Curran PJ. The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*. 2004; 9:30–52. <https://doi.org/10.1037/1082-989X.9.1.30>. [PubMed: 15053718]
- Bornstein MH, Suess PE. Child and mother cardiac vagal tone: Continuity, stability, and concordance across the first 5 years. *Developmental Psychology*. 2000; 36(1):54–65. <https://doi.org/10.1037/0012-1649.36.1.54>. [PubMed: 10645744]
- Boyce WT, Ellis BJ. Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Developmental Psychopathology*. 2005; 17(2):271–301. <https://doi.org/10.1017/S0954579405050145>.
- Burt KB, Obradovi J. The construct of psychophysiological reactivity: Statistical and psychometric issues. *Developmental Review*. 2013; 33(1):29–57. <https://doi.org/10.1016/j.dr.2012.10.002>.
- Bush NR, Alkon A, Obradovi J, Stamperdahl J, Boyce WT. Differentiating challenge reactivity from psychomotor activity in studies of children's psychophysiology: Considerations for theory and measurement. *Journal of Experimental Child Psychology*. 2011; 110(1):62–79. DOI: 10.1016/j.jecp.2011.03.004 [PubMed: 21524757]
- Casey BJ, Tottenham N, Liston C, Durston S. Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences*. 2005; 9(3):104–110. DOI: 10.1016/j.tics.2005.01.011 [PubMed: 15737818]
- Conradt E, Degarmo D, Fisher P, Abar B, Lester BM, Lagasse LL, ... Hammond JA. The contributions of early adverse experiences and trajectories of respiratory sinus arrhythmia on the development of neurobehavioral disinhibition among children with prenatal substance exposure. *Development and Psychopathology*. 2014; 26:901–916. <https://doi.org/10.1017/s095457941400056x>. [PubMed: 24909973]
- Curran PJ. Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*. 2003; 38(4):529–569. https://doi.org/10.1207/s15327906mbr3804_5. [PubMed: 26777445]
- Curran PJ, Obeidat K, Losardo D. Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*. 2010; 11(2):121–136. <https://doi.org/10.1080/15248371003699969>. [PubMed: 21743795]
- Curran PJ, Willoughby MT. Implications of latent trajectory models for the study of developmental psychopathology. *Development and Psychopathology*. 2003; 15:581–612. <https://doi.org/10.1017/s0954579403000300>. [PubMed: 14582933]
- Debat V, David P. Mapping phenotypes: Canalization, plasticity and developmental stability. *Trends in Ecology & Evolution*. 2001; 16(10):555–561. DOI: 10.1016/S0169-5347(01)02266-2
- Del Giudice M, Ellis BJ, Shirtcliff EA. The Adaptive Calibration Model of stress responsivity. *Neuroscience and Biobehavioral Reviews*. 2011; 35(7):1562–1592. <https://doi.org/10.1016/j.neubiorev.2010.11.007>. [PubMed: 21145350]
- Demaree HA, Pu J, Robinson JL, Schmeichel BJ, Everhart E. Predicting facial valence to negative stimuli from resting RSA: Not a function of active emotion regulation. *Cognition and Emotion*. 2006; 20(2):161–176. <https://doi.org/10.1080/02699930500260427>.
- Edwards JR. Ten difference score myths. *Organizational Research Methods*. 2001; 4(3):265–287. <https://doi.org/10.1177/109442810143005>.
- Ellis BJ, Boyce WT. Biological sensitivity to context. *Current Directions in Psychological Science*. 2008; 17(3):183–187. DOI: 10.1111/j.1467-8721.2008.00571.x
- El-Sheikh M. Stability of respiratory sinus arrhythmia in children and young adolescents: A longitudinal examination. *Developmental Psychobiology*. 2005; 46:66–74. <https://doi.org/10.1002/dev.20036>. [PubMed: 15690389]
- El-Sheikh M, Erath SA. Family conflict, autonomic nervous system functioning, and child adaptation: State of the science and future directions. *Development and Psychopathology*. 2011; 23:703–721. <https://doi.org/10.1017/S0954579411000034>. [PubMed: 23786705]
- El-Sheikh M, Hinnant JB. Marital conflict, respiratory sinus arrhythmia, and allostatic load: Interrelations and associations with the development of children's externalizing behavior. *Development and Psychopathology*. 2011; 23:815–829. <https://doi.org/10.1017/S0954579411000320>. [PubMed: 21756434]

- Ellis BJ, Del Giudice M. Beyond allostatic load: Rethinking the role of stress in regulating human development. *Development and Psychopathology*. 2014; 26(01):1–20. <https://doi.org/10.1017/S0954579413000849>. [PubMed: 24280315]
- Enders CK. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*. 2001; 6(4):352–370. DOI: 10.1037//1082-989X.6.4.352 [PubMed: 11778677]
- Enders CK, Bandalos DL. The relative performance of full Information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*. 2001; 8(3):430–457. https://doi.org/10.1207/s15328007sem0803_5.
- Ernst JM, Litvack DA, Lozano DL, Cacioppo JT, Berntson GG. Impedance pneumography: Noise as signal in impedance cardiography. *Psychophysiology*. 1999; 36:333–338. <https://doi.org/10.1017/S0048577299981003>. [PubMed: 10352556]
- Evans GW, Kim P. Childhood poverty and health cumulative risk exposure and stress dysregulation. *Psychological Science*. 2007; 18(11):953–957. <https://doi.org/10.1111/j.1467-9280.2007.02008.x>. [PubMed: 17958708]
- Graziano P, Derefinko K. Cardiac vagal control and children's adaptive functioning: A meta-analysis. *Biological Psychology*. 2013; 94(1):22–37. <https://doi.org/10.1016/j.biopsycho.2013.04.011>. [PubMed: 23648264]
- Grossman P, Taylor EW. Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological Psychology*. 2007; 74(2):263–285. DOI: 10.1016/j.biopsycho.2005.11.014 [PubMed: 17081672]
- Grossman P, van Beek J, Wientjes C. A comparison of three quantification methods for estimation of respiratory sinus arrhythmia. *Psychophysiology*. 1990; 27:702–714. <https://doi.org/10.1111/j.1469-8986.1990.tb03198.x>. [PubMed: 2100356]
- Heck, RH., Thomas, SL. An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus. Routledge; 2015.
- Hinnant JB, Erath SA, El-Sheikh M. Harsh parenting, parasympathetic activity, and development of delinquency and substance use. *Journal of Abnormal Psychology*. 2015; 124(1):137– 151. <https://doi.org/10.1037/abn0000026>. [PubMed: 25688440]
- Hollingshead, AB. Four factor index of social status. 1975. unpublished manuscript
- Hostinar CE, Gunnar MR. The developmental effects of early life stress an overview of current theoretical frameworks. *Current Directions in Psychological Science*. 2013; 22(5):400–406. <https://doi.org/10.1177/0963721413488889>. [PubMed: 25419054]
- Juster RP, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*. 2010; 35(1):2–16. <https://doi.org/10.1016/j.neubiorev.2009.10.002>. [PubMed: 19822172]
- Kristjansson SD, Kircher JC, Webb AK. Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. *Psychophysiology*. 2007; 44(5): 728–736. <https://doi.org/10.1111/j.1469-8986.2007.00544.x>. [PubMed: 17596179]
- Lewis GF, Furman SA, McCool MF, Porges SW. Statistical strategies to quantify respiratory sinus arrhythmia. *Biological Psychology*. 2012; 89:349–364. DOI: 10.1016/j.biopsycho.2011.11.009 [PubMed: 22138367]
- Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*. 2009; 10:434–445. <https://doi.org/10.1038/nrn2639>. [PubMed: 19401723]
- McArdle JJ. Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*. 2009; 60:577–605. <https://doi.org/10.1146/annurev.psych.60.110707.163612>.
- McEwen BS, Stellar E. Stress and the individual: Mechanisms leading to disease. *Archives of Internal Medicine*. 1993; 153:2093–2101. <https://doi.org/10.1001/archinte.1993.00410180039004>. [PubMed: 8379800]
- Perry NB, Nelson JA, Swingler MM, Leerkes EM, Calkins SD, Marcovitch S, O'Brien M. The relation between maternal emotional support and child physiological regulation across the preschool years.

- Developmental Psychobiology. 2013; 55(4):382–394. <https://doi.org/10.1002/dev.21042>. [PubMed: 22573287]
- Porges SW. The polyvagal perspective. Biological Psychology. 2007; 74(2):116–143. <https://doi.org/10.1016/j.biopsycho.2006.06.009>. [PubMed: 17049418]
- Porges SW, Byrne EA. Research methods for measurement of heart rate and respiration. Biological Psychology. 1992; 34(2):93–130. DOI: 10.1016/0301-0511(92)90012-J [PubMed: 1467397]
- Raine A. Annotation: The role of prefrontal deficits, low autonomic arousal, and early health factors in the development of antisocial and aggressive behavior in children. Journal of Child Psychology and Psychiatry. 2002; 43:417–434. <https://doi.org/10.1111/1469-7610.00034>. [PubMed: 12030589]
- Rogosa, D., Brandt, D., Zimowski, M. A growth curve approach to the measurement of change. Quantitative Methods in Psychology. 1982. <https://doi.org/10.1037/0033-2909.92.3.726>
- Rogosa DR, Willett JB. Demonstrating the reliability the difference score in the measurement of change. Journal of Educational Measurement. 1983; 20(4):335–343. <https://doi.org/10.1111/j.1745-3984.1983.tb00211.x>.
- Sameroff, AJ. Development systems: Contexts and evolution. In: Kessen, W., editor. Handbook of child psychology: Vol I. History, theory, and methods. New York: Wiley; 1983. p. 237-294.
- Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: A review. The Journal of Educational Research. 2006; 99(6): 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86(2):420–428. <https://doi.org/10.1037/0033-2909.86.2.420>. [PubMed: 18839484]
- Singer, JD., Willett, JB. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.
- Staton L, Hinnant JB, Buckhalt J, El-Sheikh M. Sleep and cognitive performance: The role of income and respiratory sinus arrhythmia reactivity. Developmental Psychobiology. 2014; 56(7):1528–1540. <https://doi.org/10.1002/dev.21247>. [PubMed: 25130163]
- Stemmler G. Implicit measurement models in methods for scoring physiological reactivity. Journal of Psychophysiology. 1987; 1(2):113–125.
- U.S. Department of Commerce. How the Census Bureau measures poverty. 2013. Retrieved from <http://www.census.gov/hhes/www/poverty/about/overview/measure.html>
- Wilder, JF. Stimulus and response: The law of initial value. Bristol; Wright: 1967.
- Williams RH, Zimmerman DW, Mazzagatti RD. Large sample estimates of the reliability of simple, residualized, and base-free gain scores. The Journal of Experimental Education. 1987; 55(2):116–118. <https://doi.org/10.1080/00220973.1987.10806443>.
- Zisner, AR., Beauchaine, TP. Psychophysiological methods and developmental psychopathology. In: Cicchetti, D., editor. Developmental Psychopathology. 3. Hoboken, NJ: Wiley; 2016.

Figure 1.

Analytic model for assessing change over time in RSA responsiveness to challenge, controlling for time-varying effects of resting RSA at each time point.

Figure 2.

Figures 2A and 2B. Model estimates of change over time in RSA level during challenge with 95% confidence intervals for RSA during challenge at each time point; resting RSA is overlaid in these figures as a reference to challenge levels. At middle childhood, RSA is measured in seconds (multiplied by 100 to facilitate model convergence). At middle adolescence, RSA is measured in milliseconds squared.

Figure 3.

Figures 3A and 3B. Model estimates of change over time in RSA in childhood and adolescence with 95% confidence intervals shown at each time point. At middle childhood, RSA is measured in seconds (multiplied by 100 to facilitate model convergence). At middle adolescence, RSA is measured in milliseconds squared.

Table 1
Bivariate correlations, means, and standard deviations for main study variables in middle childhood

	1. T1 Resting RSA	2. T2 Resting RSA	3. T3 Resting RSA	4. T1 Challenge RSA	5. T2 Challenge RSA	6. T3 Challenge RSA	7. T1 RSA	8. T2 RSA	9. T3 RSA	RSA
1										
2	.63**									
3	.56**	.65**								
4	.66**	.55**	.49**							
5	.50**	.74*	.54**	.52**						
6	.63**	.69**	.82**	.50**	.61**					
7	-.41**	-.03	-.05	.42**	.05	-.12				
8	-.21**	-.42**	-.25**	-.10	.29**	-.19*	.14			
9	-.02	-.06	-.40**	-.09	.03	-.19*	-.11	.15		
<i>M</i>	15.39	16.51	16.34	13.00	13.56	13.86	-2.35	-2.77	-2.57	
<i>SD</i>	7.96	8.19	8.69	7.88	7.63	8.08	6.47	5.67	4.95	

Note. RSA is calculated as challenge – resting scores, such that more negative scores represent greater RSA withdrawal in response to the challenge. All values were multiplied by 100 for estimation purposes in the statistical software and can be returned by the original metric (seconds) by dividing by 100.

**
 $p < .01$

*
 $p < .05$

Table 2
Bivariate correlations, means, and standard deviations for main study variables in middle adolescence

	1. T4 Resting RSA	2. T5 Resting RSA	3. T6 Resting RSA	4. T4 Challenge RSA	5. T5 Challenge RSA	6. T6 Challenge RSA	7. T4 RSA	8. T5 RSA	9. T6 RSA
1									
2	.65**								
3	.61**	.71**							
4	.23**	.12	.18*						
5	.13	.19*	.06	.66**					
6	.16	.13	.16	.67**	.69**				
7	-.56**	-.34**	-.29**	.68**	.41**	.45**			
8	-.33**	-.66**	-.47**	.39**	.61**	.44*	.58**		
9	-.23*	-.31**	-.58**	.52**	.50**	.71**	.65**	.61**	
<i>M</i>	6.91	6.94	6.84	6.70	6.64	6.44	-.21	-.25	-.45
<i>SD</i>	1.15	1.10	1.23	1.21	1.11	1.30	1.46	1.33	1.58

Note. RSA is calculated as challenge – resting scores, such that more negative scores represent greater RSA withdrawal in response to the challenge.

**
 $p < .01$

*
 $p < .05$

Table 3

Growth model of RSA level during challenge in childhood and adolescence

	Middle Childhood		Middle Adolescence	
	Model 1: Challenge RSA			
	Intercept (SE)	Linear slope (SE)	Intercept (SE)	Linear slope (SE)
Estimate	12.95*** (.48)	.37 (.28)	6.68*** (.08)	-.09* (.04)
Variance	34.80*** (7.11)	6.04** (2.35)	1.02*** (.14)	.03 (.04)
ICC	.51		.68	
AIC	7867.75		3345.20	
BIC	7948.56		3421.60	
SSA BIC	7875.65		3355.01	
RMSEA	.21		.06	
SRMR	.05		.13	
CFI	.92		.97	
χ^2 (df)	48.85*** (4)		12.80* (6)	
Model 2: Challenge RSA controlling for resting RSA				
	Intercept (SE)	Linear slope (SE)	Intercept (SE)	Linear slope (SE)
Estimate	13.02*** (.37)	.39 (.25)	6.68*** (.08)	-.09* (.04)
Variance	16.22** (5.80)	5.43** (1.97)	1.00*** (.17)	.05 (.06)
ICC	.44		.66	
AIC	7825.32		3348.66	
BIC	7916.67		3435.98	
SSA BIC	7834.25		3359.87	
RMSEA	< .01		.09	
SRMR	< .01		.12	
CFI	1.00		.97	
χ^2 (df)	.42 (1)		10.23* (3)	
Model comparison: χ^2 difference test				
	48.43*** (3)		2.57 (3)	

Note.

* denotes $p < .05$.

.100' > *p* < .001.

denotes *p* < .01.
**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Growth model of RSA in childhood and adolescence

		Middle Childhood		Middle Adolescence	
Model 1: RSA		Intercept (SE)	Linear slope (SE)	Intercept (SE)	Linear slope (SE)
Estimate		-2.46 *** (.40)	-.07 (.29)	-.18 (.10)	-.07 (.06)
Variance		16.31 *** (5.74)	7.96 *** (2.42)	1.30 *** (.23)	.05 (.08)
ICC		.06		.61	
AIC		7799.30		2914.06	
BIC		7880.02		2986.80	
SSA BIC		7807.11		2920.24	
RMSEA		.21		.23	
SRMR		.06		.06	
CFI		.72		.78	
χ^2 (df)		45.42 *** (4)		81.18 *** (6)	
Model 2: RSA controlling for resting RSA		Intercept (SE)	Linear slope (SE)	Intercept (SE)	Linear slope (SE)
Estimate		-2.51 *** (.37)	-.08 (.25)	-.23 * (.09)	-.04 (.04)
Variance		15.76 *** (5.73)	5.57 ** (1.96)	.98 *** (.20)	.02 (.07)
ICC		.07		.42	
AIC		7760.60		2854.53	
BIC		7851.85		2937.67	
SSA BIC		7769.43		2861.59	
RMSEA		< .01		.13	
SRMR		.01		.05	
CFI		1.00		.96	
χ^2 (df)		.72 (1)		15.65 ** (3)	
Model comparison: χ^2 difference test		44.70 *** (3)		65.53 *** (3)	

Note.

* denotes $p < .05$.

.100' > *p* < .001.

denotes *p* < .01.
**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript