Associations Between Respiratory Sinus Arrhythmia (RSA) Reactivity and Effortful Control in Preschool-Age Children

ABSTRACT: We tested whether respiratory sinus arrhythmia (RSA) reactivity in response to each of three self-regulation tasks (bird and dragon; knock-tap; and gift wrap) would predict self-regulation performance in a sample of 101 preschool-age children (M age = 4.49, SD = .64). While controlling for baseline RSA, decreases in RSA from bird and dragon to knock-tap (but not from baseline to bird and dragon) predicted a latent variable measuring self-regulation. Furthermore, increases in RSA from the knock-tap to gift wrap—the only task involving delay of gratification—were related to concurrent task performance while controlling for the relation between RSA reactivity and the latent self-regulation variable. Results suggest that the relations between RSA reactivity and self-regulatory ability are influenced by task-specific demands and possibly by task order. Furthermore, RSA reactivity appears to relate differently to performance on motivationally salient self-regulation tasks such as delay of gratification relative to cool executive function tasks.

Keywords: effortful control; executive function; respiratory sinus arrhythmia (RSA); vagal tone; self-regulation

INTRODUCTION

Effortful control (EC)—the self-regulatory aspect of temperament that supports volitional control of attention, emotion, and behavior—is associated with a number of important socioemotional outcomes, including social competence, internalizing and externalizing symptoms, and academic achievement (Eisenberg, Spinrad, & Eggum, 2010; Liew, 2012). Temperament is believed to have an underlying biological basis, and respiratory sinus arrhythmia (RSA) has been proposed as a physiological marker of EC. RSA is a cardiac measure of parasympathetic nervous system activation that is commonly referred to as vagal tone because this effect is mediated by the vagus nerve (Porges et al., 1981). Our study was motivated by a desire to better understand how biological and behavioral measures of self-regulation are related to one another (Porges, 2007).

Previous research has indicated that resting RSA is positively related to executive function (Hansen, Johnsen, & Thayer, 2003; Marcovitch et al., 2010; Staton, El-Sheikh, & Buckhalt, 2009), a construct that overlaps considerably with EC (Zhou, Chen, & Main, 2012). Investigators have also studied RSA reactivity, the increase or decrease in RSA relative to baseline or resting levels in response to changing environmental demands, as a correlate of self-regulatory processes. Given that resting measures of RSA have been found to predict self-regulation (Hansen et al., 2003; Marcovitch et al., 2010; Staton et al., 2009) and that RSA reactivity...
is often found to overlap with resting measures of RSA (Burt & Obrodovic, 2013), one unresolved question is whether measures of RSA reactivity predict the capacity for self-regulation over and above the relation between resting RSA and self-regulation.

Porges (2007) theorized that RSA suppression (i.e., decreases in RSA relative to baseline RSA), even more so than baseline RSA, is an index of social and emotional regulation. However, empirical evidence for relations between RSA reactivity in response to self-regulation tasks and performance on those tasks is limited in early and middle childhood. In one study of 42 low-income preschool children, RSA reactivity in response to two self-regulation tasks was unrelated to a composite measure of performance on those tasks (Blair & Peters, 2003). Similarly, no relation between RSA reactivity in response to a reaction time task and performance on an executive function task was found in a sample of 41 school-age children (Staton et al., 2009). However, the small sample size in these studies translates into low statistical power, which limits the conclusions that can be drawn from these studies. In contrast to these null results, Marcovitch et al. (2010) reported a quadratic relation between RSA reactivity and executive functioning (measured using two laboratory tasks) in a larger sample (N = 220) of 3.5-year olds. In that study, moderate RSA reactivity during these tasks was associated with high executive function, whereas both low and high RSA reactivity were associated with low executive function. In these studies, measures of RSA reactivity have typically been based on a single task (e.g., Staton et al., 2009) or averaged across tasks to compute a single RSA reactivity score (e.g., Blair & Peters, 2003; Marcovitch et al., 2010). It is unknown whether multiple consecutive measures of RSA reactivity, when considered independently, would each relate to self-regulation in a similar way.

Miyake et al. (2000) have differentiated between three distinct but interrelated aspects of executive function: working memory, attention shifting, and inhibitory control. Many of the studies relating RSA to executive function include tasks that assess working memory—the ability to keep information in short-term memory and manipulate or update this information—or attentional control (e.g., Hansen et al., 2003; Marcovitch et al., 2010). This is particularly true for studies of adolescents and adults, perhaps because inhibitory control—the ability to suppress a dominant behavioral response—improves rapidly in early childhood, and some laboratory measures assessing inhibitory control show ceiling effects by age 6–7 (Lagattuta, Sayfan, & Monsour, 2011). However, working memory is not typically included in the construct of EC (Eisenberg & Zhou, in press). It is, therefore, not clear whether RSA reactivity to cognitively demanding tasks predicts performance on self-regulatory or executive function tasks that measure primarily inhibitory control and attention and have low working memory demands.

A second important theoretical distinction made in research on self-regulation is between “cool” tasks that have primarily cognitive demands and “hot” tasks such as delay of gratification that involve a strong motivational component (Metcalfe & Mischel, 1999). Cool self-regulation is believed to rely more on the dorsolateral prefrontal cortex and the anterior cingulate cortex, whereas hot self-regulation is believed to rely more on the ventromedial prefrontal cortex and orbitofrontal cortex (Happaney, Zelazo, & Stuss, 2004). In some studies, cool self-regulation tasks have been found to differentially predict children’s academic outcomes, whereas hot self-regulation tasks have been found to differentially predict children’s socioemotional outcomes (Kim, Nordling, Yoon, Boldt, & Kochanska, 2012; Willoughby, Kupersmidt, Voegler-Lee, & Bryant, 2011). However, little is known about whether RSA reactivity relates differently to performance on hot and cool self-regulation tasks.

The Present Investigation

Relations between baseline RSA and EC have already been examined in this sample (Sulik, Eisenberg, Silva, Spinrad, & Kupfer, 2013). Consequently, in this study, we focus on the relations between RSA reactivity during a series of self-regulation tasks and performance on the self-regulation tasks. We address three research questions.

One goal of this study was to describe the pattern of changes in RSA over the course of each task and between tasks. Previous studies have frequently aggregated RSA reactivity scores across multiple tasks without examining whether patterns of change differ across tasks, or how change scores across multiple tasks are related.

A second goal was to test whether RSA reactivity across self-regulation tasks predicted children’s latent self-regulation ability (based on task performance). We hypothesized that children with greater RSA suppression (i.e., larger decreases from baseline RSA) would have greater EC because task RSA has been found to negatively predict performance on cognitively demanding tasks in older children, adolescents, and young adults (Chapman, Woltering, & Lewis, 2010; Duschek, Muckenthaler, Werner, & Reyes del Paso, 2009). However, based on the findings reported by Marcovitch et al. (2010) using a preschool-age sample, we also tested for quadratic relations between RSA reactivity and self-regulation.
A third goal was to evaluate whether there were unexplained relations between RSA reactivity in response to each task and performance on that task that was independent of the relation between RSA reactivity and self-regulation performance across all tasks. We tested this by estimating the residual covariance between concurrently measured RSA reactivity and performance for each of the three tasks that included physiological recording (i.e., bird and dragon, knock-tap, and gift wrap). A significant residual covariance would indicate that the relation between RSA reactivity and a child’s self-regulation ability does not completely explain the relation between RSA reactivity and concurrently measured task performance. Because RSA has been theorized to index responsivity to changing environmental demands (Porges, 1995), the presence of substantial residual covariances when controlling for latent self-regulation could potentially indicate that RSA is related to an individual’s effort or task engagement (Duschek et al., 2009). However, the presence of residual covariances could also indicate that RSA is differentially related to specific types of self-regulation tasks. Specifically, we predicted that RSA might relate differently to performance on hot and cool self-regulation tasks.

**METHOD**

In this study, preschool-age children participated in two laboratory visits. At the first visit, heart rate and respiration were recorded during a baseline film and during three self-regulation tasks. A computerized continuous performance (CPT) task, another laboratory measure of self-regulation, was completed in a second laboratory session (without physiological monitoring).

**Participants**

Parental consent and child assent for physiological recording were obtained for 106 children (42 girls) attending any of three preschools at a large university campus in the southwestern United States. One child was dropped from analysis due to concerns about that child having a developmental delay (see Sulik et al., 2013). An additional four children were missing all physiological data due to experimenter error (e.g., incorrect placement of respiration bellows or electrodes). All analyses include the remaining 101 children (40 girls) with some usable physiological data. Age ranged from 3.31 to 5.88 years ($M = 4.49$, $SD = .64$).

A parent from 83 families (69 mothers, 17 fathers) returned questionnaires that included demographic information. Median parental education (averaged across the educational attainment of both parents) reported on a 7-point scale ranging from “did not graduate high school” to “Ph.D. or professional degree” was “4-year college graduate.” Median family income was $75,000–100,000, and 6% of children were from single-parent families. Racial composition was as follows: 73% Caucasian; 2% African American; 9% Asian; 4% Native American; and 12% other/multi-racial. Eighteen percent of children were of Mexican American/Hispanic ethnicity.

**Procedure**

Laboratory visits were video recorded and were conducted by an experimenter and a camera person. Upon entering the laboratory, the experimenter obtained verbal assent and attached three electrodes to the child’s chest and abdomen in an inverted triangle configuration and placed a respiration bellows around the child’s torso. Children were seated in front of a laptop computer. The experimenter explained that the child would first watch a short movie, and that he or she (i.e., the experimenter) had computer work to do and would not be able to talk while the movie was playing. Baseline physiological data were collected while children watched a 2 min and 38 s meditation video showing dolphins swimming while relaxing music played. If children stopped attending to the video, fidgeted excessively, or attempted to engage the experimenter or camera person in conversation, the experimenter instructed the child to continue watching the film. Following the film, children participated in three self-regulation tasks, with no breaks between tasks. For each task, the experimenter first explained the rules and, for the first two tasks, practiced with the child. The experimenter started physiological recording immediately before starting the assessment period of each task, which was designed to last approximately 1 min. Physiological recording for each task was terminated when the task ended. In a separate laboratory session without physiological recording, children completed a CPT, another measure of self-regulation.

**Laboratory Measures of Effortful Control**

Laboratory measures that are commonly used to study self-regulation in preschool-age children were used in this study. The video recordings of each task were coded by two undergraduate research assistants who were not involved in data collection. All tapes were scored by a primary coder and at least 25% of the tapes were independently scored by a reliability coder. Intraclass correlations (ICCs), used to assess reliability between the two coders, are reported below for each task.
Bird and Dragon. For the bird and dragon task (Kochanska, Murray, Jacques, Koenig, & Vandengeest, 1996), the experimenter presented the child with two puppets, a “nice” bird and a “mean” dragon. The experimenter instructed the child to “Do what the nice bird says” but “Don’t do what the mean dragon says.” After completing practice trials to verify that the child understood the rules of the game, these puppets were used to give the child commands (e.g., “touch your nose”). The actions used in the instructions for this task were adapted to minimize the amount of movement required, although some movement was required (e.g., “wiggle your fingers,” “touch your nose,” “touch your hair”). Each of the 6 bird trials and 10 dragon trials was scored as correct (2), partially correct (1), or incorrect (0). Within the bird trials and the dragon trials, scores were averaged. A composite measure of performance on this task was computed as the product of the bird (ICC = 1.00) and the dragon (ICC = .99) scores to ensure that children had to perform well on both types of trials to receive a high score. Although we are primarily interested in responses to the inhibitory control trials, using the product score prevents a child who failed to respond correctly to either the inhibition or activation trials (e.g., due to impulsivity or behavioral inhibition) from receiving a high score for this task.

Knock-Tap. For the knock-tap task (Luria, 1966), children were first instructed to imitate the experimenter when he or she knocked (with a closed fist) or tapped (with an open palm) on a table for eight trials. Then the experimenter instructed the child to knock when the experimenter tapped, and to tap when the experimenter knocked. Following practice trials to ensure that children understood the rules, the experimenter performed 24 test trials that were scored as correct or incorrect. Performance on this task was computed as the proportion of correct trials (ICC = .98).

Gift Wrap. In the gift wrap task (Kochanska et al., 1996), the experimenter explained that he or she had a surprise for the child, and that the child should look straight ahead at the wall while the experimenter wrapped a gift behind the child. The experimenter reminded the child not to peek and noisily “wrapped” the gift for 1 min, after which the experimenter gave the gift (an animal finger puppet) to the child. Higher scores indicated better performance: 5 = child does not peek; 4 = Child peeks, but does not turn body and does not turn head over shoulder; 3 = Child peeks, but does not turn body; 2 = Child turns body while peeking in last 10 s, or child turns body while peeking for 3 s or less; 1 = Child turns body while peeking for more than 3 s (ICC = .81).

Continuous Performance Task. In a separate session (without physiological recording), children completed a CPT adapted from the NICHD Study of Early Child Care and Youth Development (NICHD Early Child Care Research Network, 2003). In this task, a series of cartoon pictures were displayed on a laptop computer with all keys covered except for the space bar. Children were asked to press the space bar each time they saw a fish, but not to press it when they saw other pictures (e.g., a beach ball). Signal detection theory (Wickens, 2002) was used to score the CPT: the proportion of errors of commission (i.e., pressing the space bar when a fish was not presented) and errors of omission (i.e., failing to press the space bar when a fish was presented) were used to generate a single detectability (d′) score that indicates how well a child performed across both types of trials.

Physiological Data

Heart rate and respiration data were recorded using James Long Company (Caroga Lake, New York) equipment and the data acquisition program SnapMaster for Windows. Interbeat interval (IBI) data were scored using James Long Company (2008) software and visually inspected for errors. Missing IBI data, although rare, were prorated based on the surrounding IBIs. RSA (in ms²) was calculated with James Long Company (2008) software using the peak-to-trough method (Grossman, Karemaker, & Wieling, 1991) and was subsequently natural log transformed to reduce skewness (Rinioło & Porges, 2000).

RESULTS

Missing Data and Preliminary Analyses

Age and sex were not missing for any participants. Missing data percentages for other study variables were as follows: bird and dragon = 3%, knock-tap = 5%, gift wrap = 1%, CPT = 12%, baseline RSA = 1%, bird and dragon RSA = 10%, knock-tap RSA = 13%, gift wrap RSA = 17%. Child refusal resulted in a small amount of missing physiological data for each task: n = 2 for bird and dragon; n = 1 for knock-tap; and n = 1 for gift wrap. Artifact in the EKG also resulted in some missing physiological data: n = 3 for bird and dragon, n = 2 for knock-tap, and n = 6 for gift wrap. The remaining missing physiological data were attributable to experimenter error or equipment failure (n = 1 for baseline RSA; n = 6 for bird and dragon; n = 4 for knock-tap; and n = 12 for gift wrap).
Variables were inspected for high skewness and kurtosis (West, Finch, & Curran, 1995) and the presence of outliers. To improve the distributional properties of the EC measures, a square-root transformation was applied to the knock-tap scores and an inverse transformation was applied to the bird and dragon and gift wrap scores (Tabachnik & Fidell, 2006). All analyses were conducted using Mplus 7.11 (Muthén & Muthén, 2012) using the robust maximum likelihood estimator and full information maximum likelihood as a missing data treatment. Descriptive statistics and correlations among study variables (following transformation, as described above) are presented in Table 1.

The minimum and maximum log transformed RSA values for each task were as follows: baseline = 3.71 and 8.97; bird and dragon = 3.76 and 9.62; knock-tap = 3.13 and 8.49; and gift wrap = 3.42 and 8.96. We computed the difference between each successive task and calculated the percentage of individuals who exhibited RSA suppression (i.e., a decrease in RSA from the previous task). These percentages were as follows: from baseline to bird and dragon = 40%; from bird and dragon to knock-tap = 86%; and from knock-tap to gift wrap = 24%.

Analysis Plan

We began by presenting measurement models for EC and for RSA and then examined the relations among these constructs in a combined model that also included children’s age and sex as predictors of latent EC. For structural equation models, we report the model $\chi^2$ and the following fit indices: root mean square error of approximation (RMSEA) with 90% confidence interval, comparative fit index (CFI), and standardized root mean square residual (SRMR).

Measurement Model for Effortful Control

Latent EC was defined by three tasks (bird and dragon, knock-tap, and gift wrap) measured concurrently with RSA and one additional task (CPT) completed in a separate laboratory session. A single factor model for EC fit the data well, $\chi^2(2) = .456$, $p = .796$, RMSEA = .000 (90% CI [.000, .125]), CFI = 1.000, SRMR = .012. Standardized factor loadings were as follows: bird and dragon $\lambda = .67$ ($SE = .10$), knock-tap $\lambda = .49$ ($SE = .10$), gift wrap $\lambda = .61$ ($SE = .10$), and CPT $\lambda = .65$ ($SE = .11$). All standardized factor loadings were significant at $p < .001$. The reliability of the EC factor was estimated to be .66, 95% CI [.49, .83] (Raykov, 2009).

Measurement Model for Baseline RSA and RSA Reactivity

We used latent difference score modeling (McArdle, 2009) to examine mean changes in RSA across the four episodes (i.e., baseline, bird and dragon, knock-tap, and gift wrap) and latent growth curve modeling (Bollen & Curran, 2006) to examine mean changes within each of the four episodes. The simultaneous use of these two approaches in a single measurement model allows us to describe (1) mean changes in RSA from each task to the next; and (2) the mean changes in RSA over the course of each individual task.

We created three RSA indicators for each episode, with each indicator corresponding to one third of the recording time for that task: total recording time was 2 min and 38 s for the baseline film and 60 s for each of the EC tasks. As a consequence, RSA scores during each of the EC tasks were based on 20 s of data, which is comparable to prior research (Blair & Peters, 2003; Marcovitch et al., 2010). The measurement model for RSA is depicted in Figure 1. We estimated a linear

### Table 1. Correlations and Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>–.14</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>4.49</td>
<td>.64</td>
</tr>
<tr>
<td>Baseline RSA</td>
<td>–.09</td>
<td>.13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6.39</td>
<td>1.11</td>
</tr>
<tr>
<td>Bird and Dragon RSA</td>
<td>–.11</td>
<td>.22</td>
<td>.81</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6.42</td>
<td>1.27</td>
</tr>
<tr>
<td>Knock-tap RSA</td>
<td>–.21</td>
<td>.08</td>
<td>.76</td>
<td>.79</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>5.52</td>
<td>1.19</td>
</tr>
<tr>
<td>Gift wrap RSA</td>
<td>–.04</td>
<td>.14</td>
<td>.68</td>
<td>.71</td>
<td>.69</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>5.89</td>
<td>1.17</td>
</tr>
<tr>
<td>Bird and Dragon Performance</td>
<td>.06</td>
<td>.31</td>
<td>.11</td>
<td>.20</td>
<td>–.04</td>
<td>.05</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.66</td>
<td>0.36</td>
</tr>
<tr>
<td>Knock-tap Performance</td>
<td>–.09</td>
<td>.49</td>
<td>.15</td>
<td>.26</td>
<td>.08</td>
<td>.12</td>
<td>.34</td>
<td>–</td>
<td>–</td>
<td>–0.53</td>
<td>0.25</td>
</tr>
<tr>
<td>Gift wrap performance</td>
<td>.06</td>
<td>.35</td>
<td>.24</td>
<td>.23</td>
<td>.11</td>
<td>.32</td>
<td>.40</td>
<td>.31</td>
<td>–</td>
<td>.84</td>
<td>0.27</td>
</tr>
<tr>
<td>CPT performance</td>
<td>.16</td>
<td>.42</td>
<td>.14</td>
<td>.11</td>
<td>–.03</td>
<td>.09</td>
<td>.46</td>
<td>.30</td>
<td>.42</td>
<td>2.86</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Note. $p < .05$ is bold.
growth model with slope loadings fixed to $-1, 0,$ and $1$ so that the latent intercept for each task could be interpreted as the average RSA score across that task. In preliminary models, the slope variances were not significantly different from zero, so we fixed all four slope variances to zero. The intercepts for each episode were used as the basis for latent difference score modeling. This model fit well, $\chi^2(60) = 54.005, p = .693, \text{RMSEA} = .000 \ (90\% \ CI [.000, .049]), \text{CFI} = 1.000, \text{SRMR} = .051.$ The variance explained in each of the indicators ranged from $.74$ to $.97.$

**Mean Changes in RSA Within Each Task.** There were no changes in RSA over time within the baseline period, slope $M = -0.03, SE = .04, p = .526.$ However, RSA declined over time (indicating RSA suppression) within the bird and dragon task and within the knock-tap task, slope $Ms = -0.09$ and $-0.18, SEs = .03$ and $.04, p = .005$ and $<.001.$ There was no change in RSA over time within the gift wrap task, slope $M = -0.01, SE = .05, p = .813.$

**Mean Changes in RSA Across Tasks.** With respect to changes across tasks, there was no mean change from baseline to bird and dragon, $M = .05, SE = .08, p = .528.$ This was followed by a substantial decrease (i.e., RSA suppression) from bird and dragon to knock-tap, $M = -0.93, SE = .09, p < .001,$ and a recovery (i.e., RSA augmentation) from knock-tap to gift wrap, $M = .40, SE = .11, p < .001.$ The cumulative change in RSA from baseline to gift wrap was negative and significantly different from zero, $M = -0.48, SE = .09, p < .001,$ indicating that RSA levels during the gift wrap task had not fully recovered to baseline levels. The absence of a complete recovery is consistent with prior research, which suggests that watching a film is associated with elevated RSA (Bush, Alkon, Obradović, Stamperdahl, & Boyce, 2011).

All correlations among baseline RSA and the latent difference scores were freely estimated in this model. The intercept for baseline RSA was unrelated to the first (bird and dragon), second (knock-tap), and third (gift wrap) latent difference scores: $\phi_s = .06, -.16,$ and $-.14,$ $SEs = .11, .12,$ and $.13,$ and $ps = .602, .173,$ and $.281.$ The lack of significant correlations between the baseline RSA scores and the RSA reactivity scores indicates that subsequent changes in

---

1 We used an alternative null model in which the repeated measures were modeled using a random intercept model (i.e., no growth over time) when calculating the CFI in all models that include latent growth curves because the standard null model is not nested with a growth curve model (Widaman & Thompson, 2003).
RSA did not depend on baseline values. In contrast, the first latent difference score was negatively associated with the second latent difference score, $\phi = -.54$, $SE = .09$, $p < .001$, and the second latent difference score was, in turn, negatively associated with the third latent difference score, $\phi = -.44$, $SE = .12$, $p < .001$. The first latent difference score was not significantly related to the third latent difference score, $\phi = -.08$, $SE = .19$, $p = .676$. These correlations among adjacent latent difference scores (i.e., from the first to second task, and from the second to third task) indicate that changes in RSA were negatively related to the immediately preceding change score. This means that individuals with more positive than average $\Delta$RSA scores in response to one task were likely to have more negative than average $\Delta$RSA change scores in response to the next task. The nonsignificant correlation between the first and third RSA change score suggests that this dependency on prior changes does not extend beyond a single task.

**Combined Structural Models**

To examine the relations between baseline RSA, RSA reactivity across the three tasks, and EC, we switched to a latent residualized change score model for RSA because this model is ideally suited to test hypotheses about RSA reactivity independent of prior RSA scores (Burt & Obradović, 2013). This change was accomplished by replacing all (nondirectional) correlations among the baseline RSA intercept and the latent difference scores with directional paths in which later scores are predicted by earlier scores. In this parameterization, the residualized latent difference ($\Delta$RSA) scores can be interpreted as the change in RSA that is orthogonal to baseline RSA and any prior $\Delta$RSA scores. In this model, latent EC was predicted by age and sex of the child, baseline RSA intercept, and all three $\Delta$RSA scores. In addition, we included residual covariances between concurrently measured $\Delta$RSA scores and task performance. This model fit the data well, $\chi^2(133) = 135.281$, $p = .429$, RMSEA = .013 (90% CI [.000, .050], CFI = .986, SRMR = .079.

To test whether each subsequent measure of RSA uniquely contributed to prediction, we adopted a hierarchical regression approach. We first constrained all the relations between the RSA scores and EC to zero, and successively released these constraints to estimate the unique contributions of each of these predictors over and above all prior RSA scores. These results are presented in Table 2. Although baseline RSA and the knock-tap $\Delta$RSA score both added to prediction, the only physiological variable that uniquely predicted EC in the final model was the knock-tap $\Delta$RSA score. In the final model, there was a significant standardized residual covariance between performance on the gift-wrap task and gift-wrap $\Delta$RSA, $\phi = .34$, $SE = .14$, $p = .015$. This indicates that performance on the gift-wrap task that was independent of the overall EC factor was related to concurrently measured $\Delta$RSA. This residual covariance was positive, which means that individuals who exhibited more positive changes in RSA from the knock-tap task to the gift-wrap task performed better than expected on the gift-wrap task relative to their latent self-regulation ability, as measured by the EC factor. The corresponding standardized residual covariances for bird and dragon and for knock-tap were nonsignificant, $\phi s = .21$ and .18, $SE s = .12$ and .25, $p s = .091$ and .487.

Given the complexity of testing quadratic terms as predictors in a latent difference score model, we adopted a simpler approach for testing quadratic relations between RSA reactivity and the EC factor. Estimating a separate model for each task, we tested task RSA and the square of task RSA as predictors of latent EC while controlling for age and sex of the child and baseline RSA. None of the quadratic terms significantly predicted latent EC, indicating the absence of a quadratic relation between RSA reactivity and self-regulation.

**DISCUSSION**

To our knowledge, this is the first manuscript that has attempted to differentiate within-task RSA reactivity and between-task RSA reactivity. Separating these two types of change yielded an interesting pattern of descriptive and predictive results that would have been obscured by aggregating the RSA reactivity scores across tasks.

**Describing Patterns of RSA Reactivity**

One goal of this investigation was to describe the pattern of changes in RSA during and across a baseline period and a series of self-regulation tasks. Mean within-task changes in RSA were not consistent for

2 When each $\Delta$RSA score was examined as a predictor while controlling for baseline RSA but not the other $\Delta$RSA scores, the second $\Delta$RSA score (from bird and dragon to knock-tap) was the only $\Delta$RSA score to significantly predict the EC factor.

3 For comparison, we also estimated a model combining RSA reactivity across all three self-regulation tasks. In this model, the combined RSA reactivity latent variable was unrelated to the EC factor while controlling for age, sex, and baseline RSA.
different tasks: RSA was stable over the course of the baseline period, declined over the course of the bird and dragon and knock-tap tasks, and remained stable over the course of the gift wrap task. It is notable that the pattern of within-task changes differed for the more cognitive self-regulation tasks (bird and dragon; knock-tap) and the delay of gratification task (gift wrap). RSA declined over the course of the more cognitive tasks, but was stable during the delay of gratification task. However, we cannot rule out the possibility that the fixed ordering of the tasks or different task demands (e.g., the need for motor activity during the bird and dragon and knock-tap tasks), rather than the difference in motivational salience between the “cool” and “hot” self-regulation tasks, influenced the within-task pattern of changes in RSA.

Unlike Brooker and Buss (2010), who estimated growth in RSA across a fear-eliciting task in a sample of 2-year olds, we did not find significant variability for the within-task slopes. It is unclear whether this indicates that RSA reactivity over the course of cognitively demanding self-regulation tasks is less variable than responses to emotion-eliciting tasks, or whether the selected nature of Brooker and Buss’ (2010) sample (which included a mix of typically developing and high-fear toddlers) contributed to variability in physiological reactivity to the fear-eliciting episode. The shorter length of time in which RSA was recorded could also have contributed to differences between our results and the results reported by Brooker and Buss (2010). Additional research is needed to understand how task demands and individual differences in emotional reactivity and self-regulation affect variability in the slope of RSA over the course of a single task. With respect to changes in mean RSA across tasks, we also found a differentiated pattern of changes. RSA did not change from baseline to bird and dragon, decreased substantially from bird and dragon to knock-tap, and then showed a partial recovery toward baseline values from knock-tap to gift wrap. This pattern of mean-level change was somewhat surprising because bird and dragon and knock-tap were characterized by similar cognitive and motor demands, so differences in mean RSA were not hypothesized for these tasks (Bush et al., 2011). One possibility is that children were not fully relaxed during the baseline film, which could potentially limit initial RSA reactivity. Ordering effects and differences in task difficulty, the amount of gross motor movement required by bird and dragon and knock-tap, the slower pace of the bird and dragon trials relative to the knock-tap trials, or other task differences such as greater language processing during bird and dragon could also have contributed to differences in average RSA reactivity for bird and dragon and knock-tap. In future research, the use of matched control tasks (Bush et al., 2011), counterbalanced task order, and experimentally varying task characteristics are suggested to better understand how task characteristics affect RSA reactivity. Because investigators have drawn somewhat conflicting conclusions about the effect of small amounts of movement on RSA reactivity (Bush et al., 2011; Porges et al., 2007), investigators might consider using actigraphs to quantify movement in future studies of RSA reactivity.

Correlations among the latent difference scores in the RSA measurement model indicated that changes in RSA were negatively associated across tasks, such that individuals with more positive than average RSA change scores in response to a given task tended to

### Table 2. Hierarchical Regression Results

<table>
<thead>
<tr>
<th>Predictor Added</th>
<th>Baseline RSA</th>
<th>ΔRSA BD</th>
<th>ΔRSA KT</th>
<th>ΔRSA GW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>SE</td>
<td>β</td>
<td>b</td>
</tr>
<tr>
<td>Age</td>
<td>.61</td>
<td>.13***</td>
<td>.61</td>
<td>.62</td>
</tr>
<tr>
<td>Female</td>
<td>.26</td>
<td>.15†</td>
<td>.26</td>
<td>.15†</td>
</tr>
<tr>
<td>Baseline RSA</td>
<td>.13</td>
<td>.05*</td>
<td>.21</td>
<td>.13</td>
</tr>
<tr>
<td>ΔRSA BD</td>
<td>−.07</td>
<td>.12</td>
<td>−.06</td>
<td>−.29</td>
</tr>
<tr>
<td>ΔRSA KT</td>
<td>−.41</td>
<td>.19*</td>
<td>−.49</td>
<td>−.52</td>
</tr>
<tr>
<td>ΔRSA GW</td>
<td>4.744</td>
<td>3.05</td>
<td>3.05</td>
<td>4.286†</td>
</tr>
</tbody>
</table>

Note. Satorra-Bentler adjustment was applied to χ² difference tests (Satorra & Bentler, 2001). BD, Bird and dragon; KT, Knock-tap; GW, Gift wrap.

***p < .001; **p < .01; *p < .05; †p < .10.
have more negative than average RSA change scores in response to the subsequent task. This serial dependence suggests that earlier RSA reactivity can potentially constrain subsequent RSA reactivity (Thayer & Lane, 2000). In future investigations in which multiple tasks are used, we recommend that researchers allow a sufficient amount of time between each task to prevent serial dependency in RSA reactivity scores. For young children, who have a limited attention span, this suggestion should be balanced against the time required to administer multiple tasks. At present, it is difficult to judge how much of a problem serial dependency is in the published literature because investigators using multiple RSA reactivity tasks do not typically report the correlations among change scores. We suggest that even when multiple baselines are used (e.g., Bush et al., 2011), investigators should report whether serial dependence of RSA scores extends beyond a single task and adjust appropriately for such effects (Burt & Obradović, 2013).

Relations Between Task Performance and RSA Reactivity

One goal of this study was to determine whether RSA reactivity in response to a battery of self-regulation tasks would predict children's self-regulatory performance over and above baseline RSA. The inclusion of RSA reactivity scores across three different self-regulation tasks allowed us to examine the unique contributions of baseline RSA and RSA reactivity to each task in the prediction of children's self-regulation. RSA during the the knock-tap task was the only RSA score that uniquely predicted self-regulation when baseline RSA and all three ARSA scores were included as predictors of latent EC. Consistent with other studies in older populations (Chapman et al., 2010; Duscheck et al., 2009), RSA reactivity (in this case, only for the knock-tap task) was negatively related to EC: greater than expected decreases in RSA predicted better performance on the self-regulation tasks. This differential pattern of results is noteworthy because the first two tasks (bird and dragon, knock-tap) had similar cognitive and motor demands (Bush et al., 2011). Our results suggest that initial reactivity to self-regulation tasks could be less informative (relative to baseline values) than subsequent reactivity in response to sustained demands on self-regulation. Furthermore, this differential pattern of relations between measures of RSA reactivity and EC would have been obscured if RSA reactivity had been aggregated across tasks rather than examined separately.

Unlike Marcovitch et al. (2010), we did not find evidence of a quadratic relation between RSA reactivity and self-regulation. Differences in the measures could potentially explain this difference in findings. Marcovitch et al.'s (2010) measure of self-regulation included working memory, which is considered an aspect of executive function but not an aspect of effortful control (Eisenberg & Zhou, in press); however, baseline RSA was positively related to self-regulation in both studies. In the present investigation, children generally performed at a high level on the EC tasks. The relations between RSA reactivity and self-regulation could potentially be different for children who, on average, find the tasks to be more challenging (Marcovitch et al., 2010).

Another goal of this study was to determine whether, after accounting for the relations between the latent EC factor and RSA reactivity, there would be residual associations between RSA reactivity in response to each task and concurrent performance on each self-regulation task. Although RSA reactivity to the third task did not predict the latent factor, which measured EC performance across all tasks, it was positively related to the residual of performance on the third task after accounting for the EC factor. The third task, gift wrap, was unique in that it involved delay of gratification and may have measured temperamental surgency (impulsivity or approach behavior) as well as EC. One possibility is that RSA reactivity to the third task was related to surgency over and above the self-regulatory demands of this task, although this finding will need to be replicated to verify that it was not idiosyncratic to the ordering of the tasks or to other characteristics of the tasks such as movement, which was required by the more cognitive tasks (bird and dragon, knock-tap) but not the delay task. Some investigators have reported that tasks involving “hot” and “cool” self-regulation are differentially related to children’s academic and socio-emotional outcomes (Kim et al., 2012; Willoughby et al., 2011). The possibility that measures of autonomic function differentially relate to children’s EC and more reactive aspects of temperament such as surgency is an important avenue for future research.

Limitations and Strengths

One limitation was that our sample was composed of primarily middle-class, nonminority children, all of whom were attending preschool, and it is unclear whether these results would generalize to other populations. This was balanced by considerable strengths, including the high-quality measurement of EC (which included four structured laboratory tasks) and the use of latent variable modeling, which provides more accurate estimates of relations among constructs by accounting for measurement error (Bollen, 1989).
Conclusion

This study draws attention to the need to carefully consider the measurement of RSA reactivity when multiple tasks are used. Our results suggest that simply aggregating across multiple tasks when calculating RSA reactivity is inappropriate, and we provide evidence that more sophisticated methods such as modeling RSA reactivity over time within a single task (Brooker & Buss, 2010) and using latent variable methods for analyzing change (Burt & Obradović, 2013) can provide a more nuanced understanding of RSA reactivity and that these methods can be applied even in relatively small samples. Finally, we find that more positive change in RSA in response to some self-regulation tasks predicted preschoolers’ worse self-regulation performance (even while controlling for baseline RSA), although more needs to be done to understand how this relation is influenced by specific task characteristics.

NOTES

This research was supported by a grant from the Arizona State University Graduate and Professional Student Association to Michael J. Sulik and a grant from the National Institute of Mental Health to Nancy Eisenberg and Tracy L. Spinrad. We express our appreciation to the parents and children who participated in the study and to the many research assistants who contributed to this project. We also gratefully acknowledge Anne Kupfer and Snjezana Huerta for their assistance with data collection.

REFERENCES


