

# Automated respiratory sinus arrhythmia measurement: Demonstration using executive function assessment

Meghan Hegarty-Craver<sup>1</sup> · Kristin H. Gilchrist<sup>1</sup> · Cathi B. Propper<sup>2</sup> · Gregory F. Lewis<sup>2</sup> · Samuel J. DeFilipp<sup>1</sup> · Jennifer L. Coffman<sup>2</sup> · Michael T. Willoughby<sup>1</sup>

Published online: 8 August 2017  
© Psychonomic Society, Inc. 2017

**Abstract** Respiratory sinus arrhythmia (RSA) is a quantitative metric that reflects autonomic nervous system regulation and provides a physiological marker of attentional engagement that supports cognitive and affective regulatory processes. RSA can be added to executive function (EF) assessments with minimal participant burden because of the commercial availability of lightweight, wearable electrocardiogram (ECG) sensors. However, the inclusion of RSA data in large data collection efforts has been hindered by the time-intensive processing of RSA. In this study we evaluated the performance of an automated RSA-scoring method in the context of an EF study in preschool-aged children. The absolute differences in RSA across both scoring methods were small (mean RSA differences =  $-0.02$ – $0.10$ ), with little to no evidence of bias for the automated relative to the hand-scoring approach. Moreover, the relative rank-ordering of RSA across both scoring methods was strong ( $r_s = .96$ – $.99$ ). Reliable changes in RSA from baseline to the EF task were highly similar across both scoring methods (96%–100% absolute agreement; Kappa =  $.83$ – $1.0$ ). On the basis of these findings, the automated RSA algorithm appears to be a suitable substitute for hand-scoring in the context of EF assessment.

**Keywords** Respiratory sinus arrhythmia · Executive functions · Heart rate · Error correction · Automated processing

Respiratory sinus arrhythmia (RSA) reflects autonomic nervous system regulation and provides significant physiological and behavioral information across a range of disciplines (Porges, 2001). In cognitive assessments, RSA provides a physiological marker of attentional engagement that supports cognitive and affective regulatory processes. However, the inclusion of RSA data in large data collection efforts has been hindered by the time-intensive processing of RSA. RSA data are sensitive to errors in heartbeat interval streams, and it has been reported that a single artifact over a 2-min segment can alter the RSA estimate more than a typical effect size seen in psychophysiological studies (Berntson & Stowell, 1998). Therefore, artifacts are normally removed through hand correction by a trained researcher. Eliminating the requirement for manual processing of the data would reduce the resource requirements for RSA measurement to enable its inclusion in a larger number of studies and facilitate timelier reporting of results.

In this study, we focused on the inclusion of RSA data in an assessment of executive functions in children. Executive functions (EFs) are a set of cognitive abilities—inhibitory control, working memory, attention shifting—that support self-regulation and undergo rapid change across early childhood (Diamond, 2013). Autonomic flexibility in response to changing environmental demands supports performance on EFs (Porges, 2007). Individual differences in EFs during the preschool period contribute to social and academic aspects of children’s school readiness because EFs facilitate children’s ability to learn how to learn (Ursache, Blair, & Raver, 2012). EFs represent one mechanism through which poverty

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13428-017-0950-2) contains supplementary material, which is available to authorized users.

---

✉ Kristin H. Gilchrist  
kgilchrist@rti.org

<sup>1</sup> RTI International, 3040 Cornwallis Rd., Research Triangle Park, NC 27709, USA

<sup>2</sup> University of North Carolina, Chapel Hill, NC, USA

negatively impacts children's performance in school (Buckner, Mezzacappa, & Beardslee, 2009; Crook & Evans, 2014; Nesbitt, Baker-Ward, & Willoughby, 2013). Moreover, deficits in EFs are implicated in multiple developmental disabilities (Sinzig, Vinzelberg, Evers, & Lehmkuhl, 2014; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005), as well as in numerous pediatric diseases and physical health conditions (Mendley et al., 2015; Morales et al., 2013; Winter et al., 2014). As a result, there is widespread interest among educators, clinicians, and policymakers in developing and deploying individualized and classroom-based programs that enhance EF abilities, and improved measurement of EF is essential for these program evaluation activities.

RSA is a quantitative metric that can be added to EF assessments with minimal participant burden because of the commercial availability of lightweight, wearable electrocardiogram (ECG) sensors. Changes in RSA that occur when individuals transition from a (resting) baseline state to engaging in cognitively or emotionally challenging tasks represent a parasympathetic response to external challenge. Individual differences in both resting baseline level and the degree of change in RSA (from baseline to task engagement) correlate with EF task performance, including in preschool-aged children (Blair & Peters, 2003; Hinnant, El-Sheikh, Keiley, & Buckhalt, 2013; Hovland et al., 2012; Marcovitch et al., 2010). The current requirement of manual editing of individual files for artifacts is a barrier to the more widespread inclusion of RSA in EF assessments, particularly in large-scale studies (where manual editing of individual files is too burdensome to be feasibly executed). The goal of this work was to evaluate the performance of an automated RSA-scoring method in the context of an EF study in preschool-aged children. Raw ECG data collected in the study were processed using an automated RSA algorithm and compared to the RSA values obtained by traditional hand scoring.

## Method

### Participants

The participants included in the current analyses were selected from the Learning, Emotion and Play in School (LEAPS) study ( $N = 102$ ), which was designed to examine the interplay between child self-regulation, parenting, and preschool classroom quality in the prediction of kindergarten readiness. Child physiological and behavioral responses were measured during a battery of computerized EF tasks administered outside of the classroom. Continuously throughout all tasks, participants wore cardiac monitors that recorded ECG signals for the calculation of RSA.

The participants were recruited from 12 preschools and day care centers in central North Carolina. The subsample of

children included in the current analyses ( $N = 40$ ) had cardiac data available for four EF tasks. Of this subsample, 11 participants were designated as the "training group," whose data were used to develop and refine the new automated RSA-scoring algorithm. The first four participants in this group were chosen on the basis of experimenter ratings of their behavior during the EF assessment (Preschool Self-Regulation Assessment: Assessor Report; Smith-Donald, Raver, Hayes, & Richardson, 2007). Two high-scoring (i.e., sat quietly, followed directions) and two low-scoring (i.e., trouble sitting still or listening, difficulty completing tasks) participants were chosen to evaluate the new automated algorithm across different physiological and behavioral profiles. The different behavior ratings were not associated with significant differences in algorithm performance; thus, the next seven participants were chosen at random. This "training group" had a mean age of 4.87 years, and 72% were female. As reported by their mothers, this group was 81% Caucasian and 19% Hispanic, with family incomes ranging from \$93,000 to \$166,000 ( $M = \$128,000$ ). The data from an additional 29 participants were used to test the new automated algorithm. This group had a mean age of 4.95 years and was 49% female. Almost half of this group were Caucasian (48%), with 21% of the remainder identifying as African American, 17% as Hispanic, and 14% as other, with family incomes ranging from \$59,000 to \$210,000 ( $M = \$132,084$ ).

### Executive function assessment

*EF Touch* is a computerized battery of EF tasks that were initially created, administered, and extensively studied in paper-and-pencil format and that now run in a Windows OS environment (see Willoughby & Blair, 2016, for an overview). The battery is modular in nature (i.e., any number of tasks can be administered in any desired order). Two "warm-up" tasks (1–2 min each) are typically administered first, in order to acclimate children to using the touchscreen. Four tasks from the EF Touch were used in this study. Only abbreviated task descriptions are presented below, since the primary focus of this study was a comparison of methods of measuring RSA acquired during EF task administration.

**Silly sounds Stroop (SSS)** This 17-item, Stroop-like task measured inhibitory control. Each item displayed pictures of a dog and cat (the left–right placement on the screen varied across trials) and presented the sound of either a dog barking or a cat meowing. Children were instructed to touch the picture of the animal that did not make the sound (e.g., touch the cat when hearing a dog bark). Each item was presented for 3,000 ms, and the accuracy and reaction times of responses were recorded. The mean accuracy across all items was used to index performance.

**Animal go/no-go (AGNG)** This 40-item go/no-go task measured inhibitory control. Individual pictures of animals were presented, and children were instructed to touch a centrally located “button” on their screen every time they saw an animal (the “go” response), except when that animal was a pig (the “no-go” response). Each item was presented for 3,000 ms, and the accuracy and reaction times of responses were recorded. The mean accuracy across all no-go responses was used to index task performance.

**Working memory span (WMS)** This 18-item span task measured working memory. Each item depicted a picture of one or more houses, each of which contained a picture of an animal, a colored dot, or a colored animal. Children verbally labeled the contents of each house. After a brief delay, the house(s) were displayed again without their contents. Children were asked to recall either the animal or color (of the animal) that was in each house (i.e., the nonrecalled contents served as a distraction). Items were organized into arrays of two-, three-, four-, and six-house trails. The mean accuracy of responses was used to index task performance.

**Something’s the same (STS)** This 30-item task is intended to measure attention shifting and flexible thinking. In the first 20 items, children are presented with two pictures (animals, flowers, etc.) that are described as being similar with respect to their color, shape, or size. A third picture is then presented alongside the original two pictured, and the child is asked to select which of the original pictures is similar to the new picture along some other dimension (e.g., color, shape, or size). In last ten items, the child is presented with three pictures and asked to identify two of the pictures that are similar, and then a second pair of the same three pictures that are similar in some other way. The mean accuracy of responses was used to index task performance.

### Respiratory sinus arrhythmia (RSA) assessment

Cardiac data were collected during a baseline period and all EF tasks. The Actiwave Cardio monitor (Camntech, Cambridge, UK) was used to collect ECG signals via two disposable electrodes (Conmed Huggables) that were attached to the left side of the child’s chest. The ECG signals were sampled continuously at 1024 Hz with 10-bit resolution. RSA was extracted from these ECG signals using both an established method with manual correction and an automated method. Both methods require the identification of R-wave peaks in the ECG signal in order to obtain an accurate R–R interval. (The R–R interval is the difference between successive R peaks and is the reciprocal of the heart rate.) RSA values are calculated from this R–R interval stream.

**Manual method** In the established method, the raw ECG data were converted for input into software for manual editing using an accepted method and software package (CardioEdit, 2007). The detection and correction of artifacts was done by persons who had completed a CardioEdit training course. The corrected R–R intervals were analyzed for variations using the Porges–Bohrer method of calculating RSA (Lewis, Furman, McCool, & Porges, 2012; Porges & Bohrer, 1990). A moving polynomial filter (with band-pass filter set to 0.24–1.04 Hz, the frequency of spontaneous respiration in children) removed frequencies lying outside the normal physiological range, and RSA was computed as the natural log of the variance on 30-s, nonoverlapping windows of filtered data.

**Automated method** In the fully automated method, raw ECG data were processed with a custom algorithm developed in MATLAB (The MathWorks) using the data from 11 participants from the larger study. Details about the algorithm are provided in a [supplemental](#) section. In the first step of the algorithm, the R–R intervals were extracted from the ECG signal using a novel method for R-wave peak identification in a continuously streamed ECG signal that maintains high accuracy in moving participants. The method was based on the well-known Pan–Tompkins approach for detecting R peaks in an ECG signal (Hamilton & Tompkins, 1986; Pan & Tompkins, 1985), but it was modified to rely more on signal timing and less on signal amplitude for determining the R-peak validity.

An R–R interval correction routine was implemented to remove erroneous intervals that met the acceptance criteria of the R-peak detection algorithm, as well as to adjust actual long or short intervals that did not preserve the overall pattern of the R–R interval stream. This was accomplished by comparing the current R–R interval to a running estimate based on the previous six correct R–R intervals. For this study, intervals were flagged for correction that fell outside of 80%–140% of the running estimate. The need for correction results from either a missed R peak (interval too long) or something other than an R peak being detected (interval too short) (Berntson, Quigley, Jang, & Boysen, 1990). Arrhythmias can also produce true skipped or extra beats, which—despite representing contractions of the heart—are independent of the pattern of neural regulation that is the target of RSA estimates. Regardless of the source of the error, a correction was attempted using the splitting, summing, or averaging of intervals. For the correction to be accepted, it was required to fall within 85%–130% of the running estimate. Otherwise, the original interval was preserved.

After corrections were applied, the R–R interval stream was processed as it is in CardioEdit, by resampling at 5 Hz using cubic spline interpolation and computing RSA using the Porges–Bohrer method (Lewis et al., 2012). The corrected,

resampled R–R interval stream was high-pass filtered using a 21-point, third-order polynomial filter to remove low frequencies. The signal was further filtered to the RSA band for children (0.24–1.04 Hz) using a 26-point FIR filter designed using a Kaiser window. RSA was computed as the log of the variance on 30-s, nonoverlapping windows of the filtered data.

A final step in the automated algorithm was to flag any of the 30-s windows with a high likelihood of reporting an erroneous RSA value. Beat detection errors that make it through the correction procedure can result in a large increase in signal amplitude of the R–R interval stream filtered to the RSA band. For each 30-s window, a point-by-point  $z$  score of the RSA band signal amplitude was computed using the average and standard deviation from the entire testing session. If there were two or more values in the window with an absolute  $z$  score greater than 4, the segment was flagged.

**Method comparison** The overarching objective of this study was to formally contrast RSA values that were obtained from the automated RSA algorithm to the manually derived scores. Consistent with convention in the literature (see, e.g., Propper & Holochwost, 2013), comparisons focused on the mean RSA across all segments of each EF task. Three strategies were used to inform a comparison of the RSA methods. First, Pearson correlations were used to test whether the rank-ordering of individual RSA values was preserved across the automated and manual methods for each task, as well as whether the intertask correlations of the RSA values were comparable across the automated and manual methods. Second, Bland–Altman plots were constructed to visualize agreement (Bland & Altman, 1986, 1995). The Bland–Altman procedure for comparing agreement between two measurement approaches involves (1) obtaining the difference in values across measurement approaches for each individual, (2) computing the mean and standard deviation of these differences, and (3) plotting the mean value against the difference with a superimposed 95% limit-of-agreement interval (95% LoA; i.e., the mean difference  $\pm$  1.96 standard deviations of the difference). The 95% LoA demarcates the range of values in which 95% of future measurements of similar individuals are intended to lie. Third, we considered changes in RSA from baseline to EF task performance. Specifically, we computed the simple difference scores that represented the degree of change in RSA from baseline to each EF task performance. These difference scores (one per task, per child) represent RSA-related task engagement. Pearson correlations were used to determine the extent to which RSA change was preserved across the algorithmic and manually edited methods. In addition, we computed reliable-change indices separately for each task and method (Maassen, 2004, 2010). Reliable-change indices were computed to determine whether each child exhibited a significant increase (augmentation of RSA), decrease (RSA withdrawal), or no change in RSA from baseline to task

engagement. Following convention, the threshold for change was defined as changes that exceeded two standard errors of measurement from the mean RSA at baseline. Cross tabulations of change status (augmentation, no change, withdrawal) across methods were evaluated using percentage agreement and Kappa coefficients.

## Results

The automated RSA algorithm was implemented in MATLAB and was able to convert the raw ECG data to RSA scores at a rate of approximately 34 s for 1 h of ECG data. A batch-processing utility enables processing of the ECG data from all participants with a single process initiation. The automated RSA algorithm was applied to up to four EF tasks, as well as to the baseline task, for 29 participants. Windows that were flagged as having potentially erroneous RSA values were excluded from the analysis. The mean and median percentage of windows flagged per subject were 7.4% and 5.5%, respectively, with a range of 0%–16.4%. The flagged windows were well distributed across tasks (7.6% for SSS, 9.9% for AGNG, 7.0% for WMS, and 7.6% for STS), but only 1.5% of windows from the baseline periods were flagged.

After the flagged windows were removed, each task was further required to have a minimum of three 30-s windows in order to be included in comparisons between the algorithm and hand-scored approaches for computing RSA. The mean RSA scores across all available windows for each task were computed separately for the algorithmic and hand-derived RSA values, which yielded a single score for each participant for each task. In total, 22–28 participants had RSA data from both scoring methods (algorithm and hand-scoring) for each task (see Table 1).

**Table 1** Descriptive statistics for respiratory sinus arrhythmia values, by executive function task and scoring method

Task	Method	<i>N</i>	<i>M</i>	<i>SD</i>	Difference ( <i>SE</i> )	<i>t</i> ( <i>df</i> )	Prob
Baseline	Manual	28	5.1	1.4			
	Automated	28	5.2	1.3	0.10 (0.04)	2.7 (27)	.01
SSS	Manual	22	5.8	1.0			
	Automated	22	5.9	0.9	0.08 (0.06)	1.4 (21)	.19
AGNG	Manual	25	5.8	1.0			
	Automated	25	5.8	1.0	0.06 (0.03)	2.1 (24)	.04
WMS	Manual	23	5.3	1.0			
	Automated	23	5.3	1.0	−0.02 (0.04)	−0.5 (22)	.62
STS	Manual	26	5.6	1.1			
	Automated	26	5.6	1.0	0.06 (0.04)	1.4 (25)	.17

SSS = Silly sounds Stroop; AGNG = Animal go/no-go; WMS = working memory span; STS = Something the same; *M* = mean; *SD* = standard deviation; *SE* = standard error; *df* = degrees of freedom; prob = probability

## Bivariate associations

Children's RSA values across EF tasks were strongly correlated with each other for both scoring methods ( $r_s = .81-.95$ ; see the off-diagonal elements of Table 2). In general, children with relatively higher RSA levels on one task also exhibited higher RSA on all of the remaining tasks. The patterns of intertask correlations for RSA values across tasks were highly similar across scoring methods. Indeed, the absolute differences in any task-by-task correlation across scoring methods were small ( $|r| = .00-.03$ ; cf. the corresponding elements above and below the diagonal of Table 2). Moreover, the mean RSA scores for each EF task were strongly associated across scoring methods ( $r_s = .96-.99$ ,  $p_s < .001$ ; see the diagonal of Table 2). In terms of the relative rank-ordering of participants, both the algorithmic and hand-scored values showed a very high degree of correspondence.

## Bland–Altman plots

Separate Bland–Altman plots were created for each of the four EF tasks, as well as for the baseline condition (see Fig. 1). The solid black line in each panel of Fig. 1 represents the mean difference between methods and is a measure of bias. There was evidence for a slightly positive bias in algorithm relative to baseline scores, though only two of the five comparisons were statistically significant and the degree of bias was small (point estimates for the degree of bias and the associated statistical significance are presented in the rightmost column of Table 1). Across all tasks in Fig. 1, the magnitude of differences across methods is small and symmetric around the estimate of bias (i.e., values hover around 0 for most cases and most tasks). Moreover, the magnitude of differences between scoring methods was not systematically related to the overall mean RSA (i.e., differences were of comparable magnitudes along the full spectrum of RSA values in the data set).

**Table 2** Correlations for hand and algorithm scored RSA values across executive function tasks

	Base– Manual	SSS– Manual	AGNG– Manual	WMS– Manual	STS– Manual
Base–Auto	<b>.99</b>	.82	.84	.85	.89
SSS–Auto	.85	<b>.96</b>	.93	.85	.92
AGNG–Auto	.81	.91	<b>.99</b>	.92	.92
WMS–Auto	.82	.84	.93	<b>.99</b>	.91
STS–Auto	.88	.92	.95	.93	<b>.98</b>

Values above and below the diagonal are manual-scored and automated-scored RSA values, respectively. Bold values along the diagonal represent agreement in RSA values across scoring methods

## RSA reactivity

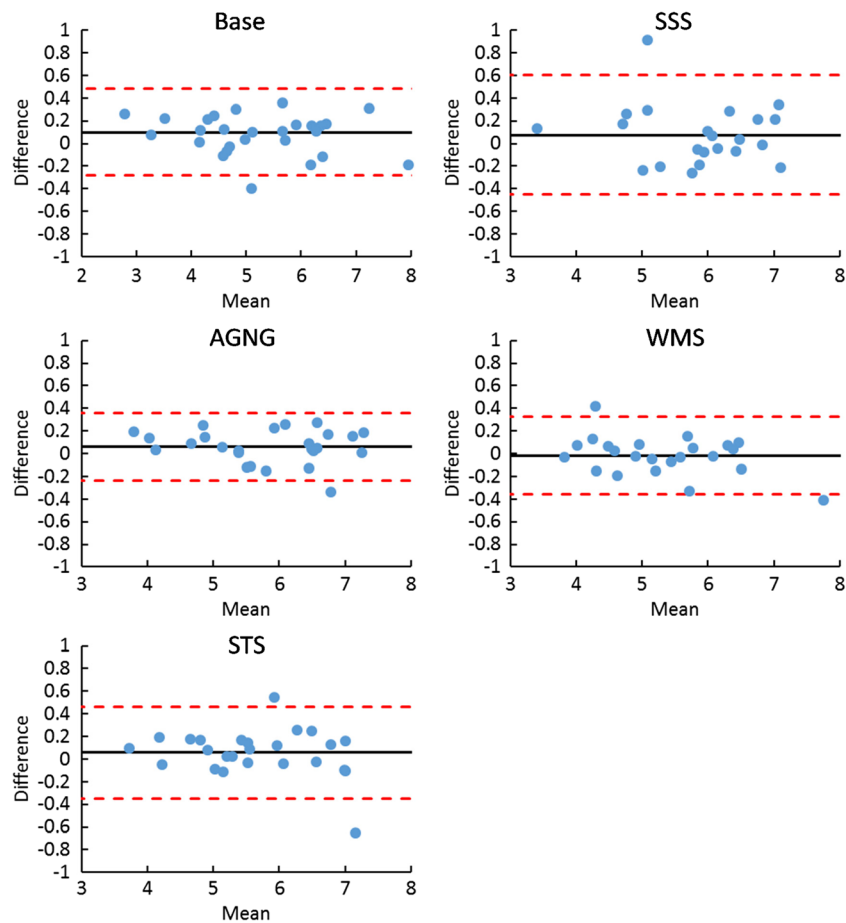
In light of the high degree of similarity across methods with respect to the estimated task-specific RSA for the baseline and EF tasks (see Table 1), we anticipated that the differences in RSA from baseline to EF task engagement would also be highly similar across methods, and this was indeed the case ( $r_{SSS} = .94$ ,  $r_{AGNG} = .98$ ,  $r_{WMS} = .97$ ,  $r_{STS} = .94$ ; all  $p_s < .0001$ ). In terms of reliable change, only 4%–14% of children demonstrated significant change in RSA from baseline to EF task engagement, and in each case RSA exclusively increased (augmentation). The manual and algorithm scoring approaches were in high agreement with respect to the presence and direction of change. As is summarized in Table 3, for three of the four EF tasks the two methods demonstrated perfect agreement ( $Kappas = 1.0$ ), and the fourth EF task exhibited nearly perfect agreement (96%;  $Kappa = .83$ ).

## Discussion

RSA analysis is a powerful tool for quantitatively assessing attentional engagement that supports cognitive and affective regulatory processes. Although RSA analysis can be incorporated into EF assessments with minimal participant burden, the large overhead for staff training and editing time associated with the requirement for hand-correction of the heartbeat interval streams has hindered the inclusion of RSA data in large collection efforts. Moreover, the additional time needed to hand-edit files after collection often delays analyses and dissemination by months or longer. In this study we explored the accuracy of an automated approach for calculating RSA from raw ECG data that included multiple stages of error identification and correction. The average RSA for each EF task was compared between the automated and traditional hand-correcting and scoring approaches.

The automated algorithm compared each heartbeat interval to several preceding intervals in order to determine whether it was outside of the expected range of interbeat variability. This technique is similar to that applied by a trained RSA editor, except that the algorithm used a fixed set of rules, whereas editors can choose to correct intervals differently. For example, Rand and colleagues (2007) found significant differences in the correction types applied by two editors, especially as the complexity of the correction grew. In addition to introducing inconsistency, manual editing requires a significant time investment for both training and actual editing. In this study, the editors were required to participate in a full day of training and complete numerous practice and validation cases before editing any study data.

The automated method does not require any manual inspection of the ECG data. In this particular study, all of the data were previously analyzed with the established method



**Fig. 1** Bland–Altman plots of respiratory sinus arrhythmia measurements during baseline and four executive function tasks. The solid black lines represent the mean differences between the manual and automated methods, and the dashed red lines represent the 95% limits of agreement

and therefore known to contain valid ECG data. For application of the algorithm to signals of unknown quality, we have developed an automated ECG prescreening routine that can be used to determine whether the data are suitable for automated processing.

A key feature of the automated algorithm is accurate R-peak detection, because correction does not preserve the

original RR interval pattern. We developed a custom R-peak detection method with high accuracy in moving participants so that the required number of corrections would be minimized. After application of a fixed set of correction rules, the automated algorithm enabled RSA computation with high agreement to the manual method. The most common sources of differences between the two methods follow: (1) The

**Table 3** Comparison of reliable change across methods

Task	Method	Reliable Change			Agreement	
		Withdraw ( <i>N</i> )	None ( <i>N</i> )	Augment ( <i>N</i> )	% Agree	Kappa
SSS	Manual	0	18	3	100	1.00
	Automated	0	18	3		
AGNG	Manual	0	19	5	100	1.00
	Automated	0	19	5		
WMS	Manual	0	22	1	100	1.00
	Automated	0	22	1		
STS	Manual	0	22	3	96	.83
	Automated	0	21	4		

SSS = Silly sounds Stroop; AGNG = Animal go/no-go; WMS = Working memory span; STS = Something the same

automated R-peak detection correctly identified a peak missed by the manual method, so correction was not required in the automated case. (2) A subjective correction of a true interval to preserve the RR interval pattern was made inconsistently between the two methods. (3) The automated method failed to correct an erroneous peak because the resulting intervals did not fall outside of the acceptable range.

RSA assessment during EF tasks provides a particular challenge for automated scoring because short task durations mean that a single erroneous RSA value can significantly impact the task mean. Therefore, the algorithm flagged windows of data with a high likelihood of artifacts that would potentially introduce error into the RSA calculation. The flagging was not based on poor agreement to hand-scored values, because this information would not be available in a fully automated study, but rather was based on abrupt changes in the filtered RR interval signal amplitude characteristic of uncorrected errors. In this data set, only 7% of the windows were flagged, indicating that the majority of errors were corrected in the initial stage. Many of the flagged windows corresponded with sections of poor ECG signal quality where heartbeats were not readily identifiable. A high percentage of flagged intervals may suggest that a segment of ECG data is not of sufficient quality for automated processing, but we note that poor ECG quality also presents challenges for manual editing. The overall percentage of flagged windows was higher during the EF tasks than during the baseline period, perhaps due to the increased motion in completing tasks, in comparison to baseline.

The results of the automated approach were highly correlated with the traditional approach across all EF tasks. Bland–Altman analysis revealed that the differences in RSA means were small and symmetric around the estimate of bias (there was evidence for a slightly positive bias in algorithm relative to baseline scores). Additionally, the magnitude of the difference was not dependent on the mean RSA value. Both methods agreed on the presence and direction of change between an individual's baseline and EF task RSA scores. In terms of the relative rank-ordering of participants, both approaches showed a very high degree of correspondence. Consistent with the standard use of RSA in the context of EF testing, this study focused on aggregate agreement between the hand-scoring and automated methods. In other applications that involve longer periods of RSA monitoring, it will be important to test the equivalence of methods at finer-grained units of time (i.e., epoch-by-epoch agreement across each task). Additional studies will be needed to cross-validate these results in larger and more diverse samples.

This study validates the use of an automated RSA-scoring method in the context of EF testing in children. The approach is amenable to use in adults, with appropriate modification to account for the different range of normal heart intervals, but this would require additional validation studies. Many other

applications could benefit from an automated RSA-scoring approach. Examples include studies of stress or behavior that require monitoring large numbers of participants and/or monitoring them over extended periods of time. Additionally, the ability to provide results in near real time enables applications in which RSA responses can be used to trigger timely intervention. The addition of RSA scoring to psychological and educational assessments and monitoring of chronic conditions will enrich knowledge concerning the underlying causes of differences between individuals, with minimal burden to the user.

**Author note** The project described was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number UL1TR001111. The R-detection algorithm was developed under National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health Award Number R01EB014742. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## References

- Berntson, G. G., & Stowell, J. R. (1998). ECG artifacts and heart period variability: Don't miss a beat! *Psychophysiology*, *35*, 127–132.
- Berntson, G. G., Quigley, K. S., Jang, J. F., & Boysen, S. T. (1990). An approach to artifact identification: Application to heart period data. *Psychophysiology*, *27*, 586–598.
- Blair, C., & Peters, R. (2003). Physiological and neurocognitive correlates of adaptive behavior in preschool among children in Head Start. *Developmental Neuropsychology*, *24*, 479–497.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*, 307–310.
- Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: Why plotting difference against standard method is misleading. *Lancet*, *346*, 1085–1087.
- Buckner, J. C., Mezzacappa, E., & Beardslee, W. R. (2009). Self-regulation and its relations to adaptive functioning in low income youths. *American Journal of Orthopsychiatry*, *79*, 19–30.
- CardioEdit. (2007). *CardioEdit 1.5—Inter-beat interval editing for heart period variability analysis: An integrated training program with standards for student reliability assessment*. Brain–Body Center, University of Illinois at Chicago.
- Crook, S. R., & Evans, G. W. (2014). The role of planning skills in the income-achievement gap. *Child Development*, *85*, 405–411. doi:10.1111/Cdev.12129
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *63*(64), 135–168. doi:10.1146/annurev-psych-113011-143750
- Hamilton, P. S., & Tompkins, W. J. (1986). Quantitative investigation of QRS detection rules using the MIT/BIH Arrhythmia Database. *IEEE Transactions on Biomedical Engineering*, *33*, 1157–1165. doi:10.1109/Tbme.1986.325695
- Hinnant, J. B., El-Sheikh, M., Keiley, M., & Buckhalt, J. A. (2013). Marital conflict, allostatic load, and the development of children's fluid cognitive performance. *Child Development*, *84*, 2003–2014. doi:10.1111/Cdev.12103
- Hovland, A., Pallesen, S., Hammar, A., Hansen, A. L., Thayer, J. F., Tarvainen, M. P., & Nordhus, I. H. (2012). The relationships among heart rate variability, executive functions, and clinical variables in

- patients with panic disorder. *International Journal of Psychophysiology*, 86, 269–275. doi:10.1016/j.ijpsycho.2012.10.004
- Lewis, G. F., Furman, S. A., McCool, M. F., & Porges, S. W. (2012). Statistical strategies to quantify respiratory sinus arrhythmia: Are commonly used metrics equivalent? *Biological Psychology*, 89, 349–364.
- Maassen, G. H. (2004). The standard error in the Jacobson and Truax Reliable Change Index: The classical approach to the assessment of reliable change. *Journal of the International Neuropsychology Society*, 10, 888–893.
- Maassen, G. H. (2010). The two errors of using the within-subject standard deviation (WSD) as the standard error of a reliable change index. *Archives of Clinical Neuropsychology*, 25, 451–456. doi:10.1093/arclin/acq036
- Marcovitch, S., Leigh, J., Calkins, S. D., Leerks, E. M., O'Brien, M., & Blankson, A. N. (2010). Moderate vagal withdrawal in 3.5-year-old children is associated with optimal performance on executive function tasks. *Developmental Psychobiology*, 52, 603–608. doi:10.1002/Dev.20462
- Mendley, S. R., Matheson, M. B., Shinnar, S., Lande, M. B., Gerson, A. C., Butler, R. W., ... Hooper, S. R. (2015). Duration of chronic kidney disease reduces attention and executive function in pediatric patients. *Kidney International*, 87, 800–806. doi:10.1038/ki.2014.323
- Morales, G., Matute, E., Murray, J., Hardy, D. J., O'Callaghan, E. T., & Tlacuilo-Parra, A. (2013). Is executive function intact after pediatric intracranial hemorrhage? A sample of Mexican children with hemophilia. *Clinical Pediatrics*, 52, 950–959. doi:10.1177/0009922813495311
- Nesbitt, K. T., Baker-Ward, L., & Willoughby, M. T. (2013). Executive function mediates socio-economic and racial differences in early academic achievement. *Early Childhood Research Quarterly*, 28, 774–783. doi:10.1016/j.ecresq.2013.07.005
- Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32, 230–236. doi:10.1109/Tbme.1985.325532
- Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, 42, 123–146.
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, 74, 116–143. doi:10.1016/j.biopsycho.2006.06.009
- Porges, S. W., & Bohrer, R. E. (1990). Analyses of periodic processes in psychophysiological research. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (1st ed., pp. 708–753). New York, NY: Cambridge University Press.
- Propper, C. B., & Holochwost, S. J. (2013). The influence of proximal risk on the early development of the autonomic nervous system. *Developmental Review*, 33, 151–167. doi:10.1016/j.dr.2013.05.001
- Rand, J., Hoover, A., Fishel, S., Moss, J., Pappas, J., & Muth, E. (2007). Real-time correction of heart interbeat intervals. *IEEE Transactions on Biomedical Engineering*, 54, 946–950. doi:10.1109/Tbme.2007.893491
- Sinzig, J., Vinzelberg, I., Evers, D., & Lehmkuhl, G. (2014). Executive function and attention profiles in preschool and elementary school children with autism spectrum disorders or ADHD. *International Journal of Developmental Disabilities*, 60, 144–154. doi:10.1179/2047387714y.0000000040
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, 22, 173–187.
- Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, 6, 122–128. doi:10.1111/j.1750-8606.2011.00209.x
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, 57, 1336–1346.
- Willoughby, M. T., & Blair, C. B. (2016). Longitudinal measurement of executive function in preschoolers. In J. Griffin, L. Freund, & P. McCardle (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopment and translational research* (pp. 91–113). Washington DC: American Psychological Association.
- Winter, A. L., Conklin, H. M., Tyc, V. L., Stancel, H., Hinds, P. S., Hudson, M. M., & Kahalley, L. S. (2014). Executive function late effects in survivors of pediatric brain tumors and acute lymphoblastic leukemia. *Journal of Clinical and Experimental Neuropsychology*, 36, 818–830. doi:10.1080/13803395.2014.943695