Neuropsychology

Measures of Attention in Rett Syndrome: Internal Consistency Reliability

Susan A. Rose, Sam V. Wass, Jeffery J. Jankowski, and Aleksandra Djukic

Online First Publication, June 7, 2021. http://dx.doi.org/10.1037/neu0000744

AMERICAN PSYCHOLOGICAL ASSOCIATION

Neuropsychology

ISSN: 0894-4105

https://doi.org/10.1037/neu0000744

Measures of Attention in Rett Syndrome: Internal Consistency Reliability

Susan A. Rose[1], Sam V. Wass[2], Jeffery J. Jankowski[1, 3], and Aleksandra Djukic[1, 4]

[1] Department of Pediatrics, Montefiore Medical Center, Albert Einstein College of Medicine/Children's Hospital at Montefiore

[2] Department of Psychology, University of East London

[3] Department of Social Sciences, Queensborough Community College/CUNY

[4] Rett Syndrome Center, Department of Neurology, Montefiore Medical Center, Albert Einstein College of Medicine/Children's Hospital at Montefiore

Objective: Rett syndrome (RTT), an x-linked neurodevelopmental disorder caused by spontaneous mutations in the MECP2 gene, is characterized by profound impairments in expressive language and purposeful hand use. We have pioneered the use of gaze-based tasks to by-pass these limitations and developed measures suitable for clinical trials with RTT. Here we estimated internal consistency reliability for three aspects of attention that are key to cognitive growth and that we previously identified as impaired in RTT. ***Method:*** Using a sample of 66 children with RTT (2–19 years), we assessed *Sustained Attention* (butterfly task: Butterfly traverses the screen only when fixated and distractors are ignored); *Disengaging/Shifting Attention* ("gap/overlap" task: Shifts of gaze from central to peripheral targets are compared in conditions where the central stimulus remains or disappears at the onset of the peripheral target); *Selective Attention* (search task: the target is embedded in arrays differing in size and distractor type). ***Results:*** Reliability was acceptable to excellent on almost all key measures from tasks assessing Sustained Attention and Disengaging/Shifting Attention, with split-half coefficients and Cronbach alphas ranging from .70 to .93. Reliability increased as more trials were aggregated, with acceptable levels often reached with as few as six to nine trials. Measures from Selective Attention showed only limited reliability. ***Conclusion:*** Finding that critical aspects of attention can be reliably assessed in RTT with gaze-based tasks constitutes a major advance in the development of cognitive measures appropriate for clinical and translational work.

Key Points

Question: We sought to determine whether measures of three aspects of attention that are key to cognitive growth, and that we previously identified as impaired in children in Rett syndrome (RTT), would show internal consistency reliability. ***Findings:*** Reliability was acceptable to excellent on almost all key measures from tasks assessing both sustained attention and disengaging/shifting attention, but limited for measures assessing selective attention. ***Importance:*** These data provide the first evidence that measures of critical aspects of cognition, assessed in RTT with gaze-based measures, show good psychometric reliability; this represents an important step in establishing objective behavioral outcome measures for use in assessing the effects of therapeutic interventions. ***Next Steps:*** These findings point the way to improving the reliability for measures of Sustain Attention and, more generally, for developing objective, performance-based assessments for other cognitive domains and children and with other disorders.

Keywords: Rett syndrome, attention, reliability, eye-tracking

Background

Rett syndrome (RTT), an x-linked neurodevelopmental disorder affecting approximately 1/10,000 females, is caused by spontaneous mutations in the MECP2 gene (Amir et al., 1999; Kaufmann et al., 2005). RTT is characterized by apparently normal development until 6–18 months of age and then a severe regression that includes profound impairments in expressive language and purposeful hand use, along with the appearance of gait abnormalities and hand stereotypies. Other symptoms include the development of apraxia, spasticity and scoliosis, breathing irregularities (hyperventilation, breath holding, apnea), and a slowing of brain and head growth (Chahrour, 2007; Neul et al., 2010).

Because the profound impairments in speech and motor control in RTT preclude standard neuropsychological testing, the cognitive phenotype of this disorder remains largely unknown. As a result, there has been an absence of objective behavioral outcome measures for use in assessing individual differences and developmental change. This absence has also hampered attempts to assess the effects of the novel therapeutic interventions that are beginning to appear to treat this disorder. Indeed, the absence of such measures

This research was funded by Grant # 3516 from the International Rett Syndrome Foundation (IRSF). We wish to thank Nathaniel Paley for assistance with testing and data transfer/management and to gratefully acknowledge all the research participants and their families.

The authors declare no conflict of interest.

Correspondence concerning this article should be addressed to Susan A. Rose, Department of Pediatrics, Montefiore Medical Center, Albert Einstein College of Medicine/Children's Hospital at Montefiore, Van Etten Building, 1300 Morris Park Avenue, Bronx, NY 10461, United States. Email: susan.rose@einsteinmed.org

has been a major obstacle to demonstrating behavioral changes in human trials with RTT.

However, the situation shows signs of changing. A few studies have now reported the effectiveness of evoked response potential measures for assessing some aspects of cognition (Foxe et al., 2016; LeBlanc et al., 2015) and others have pointed to the usefulness of gaze-based measures to assess stimulus discrimination and learning (Baptista et al., 2006; Fabio et al., 2009, 2019). And indeed, an eye-tracking version of the Peabody Picture Vocabulary Test has recently been developed that revealed marked individual differences in language comprehension among females with RTT (Ahonniska-Assa et al., 2018). In our own lab, we have been using this technology to develop visually based cognitive measures that would make suitable endpoints for assessing the success of clinical interventions. These measures have revealed quantifiable atypicalities in three areas that are key to cognitive growth: memory, anticipation, and attention (Djukic et al., 2012, 2014; Djukic & McDermott, 2012; Rose et al., 2013, 2016, 2017, 2019a, 2019b). Thus, studies based on preferential looking are showing that eye gaze can be used in this population as an indication of interest, in addition to its use as a way to make requests and as a mode of communication. These new measures show considerable promise for use in clinical trials because (a) the methods have proven feasible with RTT and (b) some of the measures themselves have already been shown to be sensitive to change after treatment with Glatiramer Acetate (Copaxone) (Djukic et al., 2016) and Lovastatin (Mevacor) (in preparation). Thus, they satisfy two important characteristics of an optimal outcome measure: feasibility and sensitivity. Here we examine another important characteristic such measures must satisfy to be acceptable endpoints in translational research—namely, reliability.

The present study examines the reliability of measures from tasks that we have previously used to assess attention in children with Rett Syndrome. These tasks focus on three different aspects of attention that have both theoretical and practical significance: (a) sustained attention, (b) disengagement and shifting attention, and (c) selective attention. These aspects of attention are considered to be a driving force of cognitive development (Amso & Scerif, 2015; Hendry et al., 2016, 2019). A number of studies have found that measures of these three aspects of attention show good internal consistency reliability (assessed with alphas or split-half coefficients (de Jong et al., 2016; Rose et al., 2001, 2012) and good test–retest reliability (van Baar et al., 2020) in typically developing infants and young children. Other studies have shown that these aspects of attention relate to key areas of cognitive growth, and are predictive of later IQ (Cornish et al., 2005; Rose et al., 2005, 2008; Rose, Feldman, & Jankowski, 2011; Rose, Feldman, Jankowski, et al., 2011; Scerif et al., 2012), executive function (Cuevas & Bell, 2014; Rose et al., 2012), and academic achievement (Blankenship et al., 2019; Bornstein et al., 2013). Using structural equation modeling, we were successful in modeling pathways from infant attention to various aspects of later cognition, including executive functioning, that extended through toddlerhood and on into adolescence. Indeed, attentional control has been characterized as a "hub" cognitive domain that gates the subsequent acquisition of skills in other areas, and has become a prominent domain for intervention in cognitive training studies (Forssman & Wass, 2018; Rueda et al., 2005; Wass, 2015; Wass et al., 2011).

Overall then, there is evidence that measures of attention are reliable in typically developing children and that they are predictive of later cognition—a further indication of their reliability (since reliability serves as a constraint on predictive validity). However, little is known about the reliability (or predictability) of such measures in non-typical populations. And, the reliability of measures could be much less in RTT children because they often complete fewer trials of an assessment than their typically developing peers, and reliability is dependent on the number of trials or items aggregated. Overall, given the usefulness of measures of attention in characterizing the cognitive phenotype of Rett Syndrome, and their rich potential for use as clinical biomarkers and outcome measures, knowledge about their reliability has become increasingly important.

Aspects of Attention Considered

The three aspects of attention that we singled out for examination here are characterized and described below. This study builds on earlier work where we developed gaze-based measures for assessing each aspect in children with RTT, demonstrated their feasibility for use in this population, and compared the performance of children with RTT to that of their age- and gender-matched typically developing (TD) peers. A brief overview of the findings of that work is provided here to lay the foundation for the present study.

Sustained Attention

Sustained attention refers to the ability to focus or concentrate attention and maintain vigilance in the face of distractors. It is thought to involve top-down connectivity extending from the anterior attention system, particularly prefrontal and parietal regions in the right hemisphere (Grahn & Manly, 2012; Sarter et al., 2001; Silver et al., 2007). To examine sustained attention in children with RTT we used an innovative, gaze-based task modeled after Wass and colleagues (Wass et al., 2011). In this task, the target (a butterfly) appears on the left of a computer screen and remains stationary until the child fixates it; as long as it is fixated it flutters and moves from left to right, while distractors appear in the periphery and scroll in the opposite direction. If the child looks at any of the distractors, the target freezes and the distractors disappear, leaving only the static target on the screen. In effect then, the child is "rewarded" for staying focused and ignoring peripheral stimuli. Children with RTT looked to the target only about 25% of the time (vs. 60% for the TD group). The difficulty experienced by children with RTT in sustaining attention on the target was due largely to three factors—they were *slower* to engage, *more* distractible, and *slower* to re-engage (Rose et al., 2017).

Disengaging and Shifting Attention

Disengagement and shifting are core components of attention orienting (Fan et al., 2005; Posner & Petersen, 1990), with visual exploration of the world requiring the ability to flexibly break attention, or disengage, in order to be able to shift attention elsewhere. While attentional shifting is thought to involve largely low-level oculomotor networks, especially pathways from the retina to the superior colliculus (Karatekin et al., 2007), attentional disengagement is thought to rely more heavily on prefrontal and parietal areas of the cortex (Csibra et al., 1997, 1998). To examine these aspects of attention in children with RTT, we used a computerized version of the classic "gap/overlap" task (Elsabbagh et al., 2009), where the appearance of a central fixation stimulus is followed, after a brief delay, by the appearance of a peripheral one. On some trials (gap or baseline), the central stimulus disappears at the onset of the peripheral target; on other trials (overlap), the central stimulus remains present, competing for attention. Looks to the peripheral target are rewarded by the brief presentation of an animated cartoon. In effect then (and unlike sustained attention), on this task the child is "rewarded" for flexibly switching attention. Children with RTT had considerable difficulty shifting attention, making fewer saccades to the peripheral target than their TD counterparts in both conditions, but especially on overlap trials (63% vs. 90%); in addition, the younger patients showed slower saccadic reaction times (SRTs) even when they did disengage (Rose et al., 2019a).

Selective Attention

Selective attention, the process of preferentially directing awareness to specific information or stimuli while simultaneously ignoring distracting stimuli, is often tested with visual search tasks (Treisman & Gelade, 1980). In such tasks. the distractors generally differ from targets either by (a) a single feature (e.g., color *or* shape), which creates an effect where the target tends to "pop-out" or (b) a conjunction of features (e.g., color *and* shape), where a more time consuming serial search is required to locate the target. Performance on search tasks relies heavily on the parietal cortex (Corbetta et al., 2008; Nobre et al., 2003). To examine visual search in children with RTT we adapted a gaze-based task developed by Kaldy (Kaldy et al., 2011), where children search for a target (red apple) in different sized arrays containing distractors (blue apples and red cylinders); the red apple rotates and emits a pleasant sound when the child fixates on it. Children with RTT were less successful than their TD counterparts in locating the target in both types of trials (40% vs. 80%) and even when successful, often took significantly longer to find it (Rose et al., 2019b).

Assessing the Reliability of Attention Measures

While the findings from these three gaze-based tasks of attention have helped to quantify the cognitive phenotype of RTT, to serve as important endpoints in assessing the cognitive efficacy of clinical interventions, such measures need to be shown to be reliable. Given the paucity of measures to evaluate cognition in this population, such evidence of reliability would constitute a major advance for the area. Of particular importance is their internal consistency reliability, which is necessary for both their predictive validity and their usefulness as endpoints in clinical intervention studies. The internal consistency of task scores impacts effect size, the power of hypothesis tests, predictive validity, and the replicability of results across studies (Green & Yang, 2009; Humphreys & Drasgow, 1989; Lebel & Paunonen, 2011; Schmidt & Humter, 1999). For these reasons, we have focused here on gauging the internal consistency of the measures used to assess sustained attention, disengaging/shifting attention, and selective attention.

Internal consistency reliability refers to the overall consistency of responses across trials or items. When composite scores are made up of items that all measure the same ability or trait, their reliability is generally estimated with an index such as the Spearman–Brown split-half coefficient or Cronbach's alpha coefficient (α), which is equivalent to the average of all possible split-half correlations. In assessing internal consistency, researchers generally use the individual items but, when there are a number of related constructs, as in the search task, the condition-averaged trials can serve as the items for Cronbach's α (Thigpen et al., 2017; Van de Weijer-Bergsma et al., 2015), so for this task we will compute alphas both ways (i.e, using individual trials as items and using condition-averaged trials as items). Alpha coefficients, computed either way, take into account the number of test items and their average inter-correlation. The magnitude of these coefficients indicates the extent to which the trials or items do actually share common variance (Cronbach & Warrington, 1951). While it has recently been argued that internal consistency reliability should be estimated even in experimental studies (Green et al., 2016), the need to establish such reliability has long been recognized as critical in the context of clinical and translational work (Thigpen et al., 2017; Towers & Allen, 2009).

Although Cronbach's alpha is the most widely used method for estimating internal consistency reliability, it has known limitations (Bowden & Finch, 2017; Trizano-Hermosilla & Alvarado, 2016). These limitations stem from the underlying assumptions of classic test theory, namely, that each observed score has two parts, a true score and an error component, where the true scores are the same for all items (tau-equivalence) and the error components are uncorrelated. To the extent these assumptions are violated, reliability may be underestimated by Cronbach's α, by an amount varying from .6% to 11.1%, depending on the extent of the violation (Green & Yang, 2009). To address this issue, a number of alternative lower bound estimates of reliability have been proposed (Bendermacher, 2010; Guttman, 1945; Trizano-Hermosilla & Alvarado, 2016). The potential impact of such lower bound estimates here will be examined in supplementary analyses.

Objective

This study, a critical next step in our efforts to develop outcome measures for use in clinical trials with RTT, focuses on establishing the reliability (internal consistency across items within a test) of measures from tasks used to assess three key aspects of attention—sustained attention, disengaging and shifting attention, and selective attention. For measures that are dichotomous in nature (e.g., correct/incorrect; presence/absence of shift to peripheral target), reliability will be assessed with split-half coefficients; for those that are continuous (e.g., latency to respond), reliability will be assessed with Cronbach's alpha. Because it is often difficult to get a high number of usable trials from children with RTT, and alpha is dependent on this factor, we specifically looked at how alpha varies as a function of the number of trials included in the analysis.

This study is part of our effort to address the critical need that has been noted for measures that are sensitive (distinguishing children with RTT from their typically developing peers), reliable, and relevant for assessing the cognitive benefits of therapeutic interventions (Katz et al., 2012).

Method

Participants

The final sample of participants included 66 females with clinically diagnosed classical Rett syndrome (Neul et al., 2010), consecutively recruited from the Rett Center at the Children's Hospital of Montefiore ($M = 9.05$ years; $SD = 4.67$, range = 2–19 years.) RTT was genetically confirmed in all participants. Testing was attempted, but discontinued, for an additional 15 children either because they could not successfully complete the calibration procedure ($N = 6$), were too overactive/restless to complete any of the three tasks ($N = 7$), or because of technical difficulties ($N = 2$). We recruited over a wide age range, which tends to be common in RTT studies (Fabio et al., 2009; Sysoeva et al., 2020), to better establish the cognitive phenotype of RTT and examine the impact of age on performance. The three gaze-based tasks used in the present study have been shown to be appropriate over this age range, even for the youngest children included in the present sample (Elsabbagh et al., 2013; Wass et al., 2012). All children had personal eye gaze-based Augmentative and Alternative Communication (AAC) devices, and thus were familiar with this mode of communication.

Most of the 66 participants contributed data on each task: sustained attention ($N = 57$), disengaging and shifting attention ($N = 57$), selective attention ($N = 54$). Children were excluded from the sustained attention task if they failed to look at the target on at least 6 trials ($N = 9$), from the disengagement task if they failed to look at the central target on at least 12 trials ($N = 9$), and from the selective attention task if they failed to look at the computer screen on at least 4 trials ($N = 12$). A series of t-tests carried out for each task indicated that the children excluded did not differ from the rest of the sample either in age or on any of the clinical/background factors (described below).

Data for about half the sample (51%, 56%, and 46%) were previously reported in studies comparing children with RTT to their neuro-typical peers on, respectively, sustained attention (Rose et al., 2017), disengaging and shifting attention (Rose et al., 2019a), and selective attention (Rose et al., 2019b).

Clinical Characteristics

Of the 66 participants, 76% had one of the most frequently identified mutations on the MECP2 gene—R168X, T158M, R306C, R255X, R294X, R270X and large C terminal deletions (Dragich et al., 2000). The remaining 24% had other individual mutations, with each less frequent mutation present in only one or two participants.

The age at regression averaged 15.6 months ($SD = 7.1$). At the time of testing, all children had completed active regression and were in Stage III, a period of stabilization or plateau. The age at diagnoses of RTT for all children was co-incident with the onset of regression.

The clinical characteristics of the sample were assessed with the Rett Syndrome Severity Scale (RSSS; Kaufmann et al., 2012). This scale comprises clinical ratings on seven parameters (seizure frequency/manageability, respiratory irregularities, scoliosis, ability to walk, hand use, speech, and sleep problems), with each parameter rated on a 4-point Likert scale from 0 (*absent/normal*) to 3 (*severe*). Composite scores on the RSSS, created by taking the mean of all subscales, averaged 8.79 ($SD = 2.93$), with 42% of the sample scoring in the *mild range* (0–7), and the remaining 58% scoring in *the moderate range* (8–14). Many (53.0%) were ambulatory (able to walk unaided or with support). Half (50%) of the children had a history of seizures (with 33% scoring in the *mild range,* 11% in the *moderate range*, and 6% in the *severe range* on the RSSS) and were taking one or more of the following anticonvulsants (valproic acid, topiramate, clonazepam, and/or levetiracetam). As is commonly found, symptom severity (as indexed by the RSSS score) tended to increase with age, $r = .34$, $p = .005$.

The protocol was approved by the institutional review board (IRB Protocol 3203) and written consent was obtained for all participants.

Apparatus

Stimuli were presented on a 23-in. flat panel monitor (resolution, 1024 × 768 pixels) in conjunction with a Tobii X2-60 infrared eyetracker (Tobii Technologies). Matlab, Psychtoolbox, and Talk2-Tobii software were used to allow for a gaze-contingent interface during stimulus presentation. Manufacturer-supplied algorithms for pupil, corneal reflection, and face identification were used during eye-tracking; gaze data were sampled at 60 Hz. Left and right eye gaze positions were recorded separately and then averaged for analyses. The reliability of the algorithms that were used during eye-tracking and methods for achieving a satisfactory calibration are dealt with in papers by Wass and colleagues (Leppänen et al., 2015; Wass et al., 2014).

Procedure

Testing was conducted in a quiet room; participants were tested individually, whilst seated approximately 45 cm from the monitor. Ambient light levels were reduced to diminish distraction. Verbal instructions, limited to "Look at the TV," were used at the beginning of the session. (All children oriented to the screen at this request. It should be noted that, even in the absence of any verbal instruction, infants orient to the TV screen in this situation.) To minimize body and head movement, participants were generally seated on their parent's lap. Parents kept their eyes closed during testing.

There were two blocks of trials for each task. The three tasks were interleaved within each block in a testing session that, in its entirety, took less than 20 min; breaks were given as necessary.

Calibration

At the start of the session, children completed a 5-point calibration procedure, in which pulsing colored blocks (1°–1.5°) appeared successively in the center and four corners of the screen, with each change in position accompanied by a sound ("Whee"). Point-of-gaze was calibrated by comparing each look to the known coordinates of the target; results were presented graphically. The quality of the calibration data was determined by the closeness of the fixation points to the calibration points. If the points did not cluster, or any targets were missed, the calibration was repeated until a satisfactory calibration was achieved. Each calibration attempt took less than a minute. The quality of the calibration was determined by the closeness of the fixation points to the calibration points.

Tasks and Measure

Sustained Attention

Trials started with a target, a butterfly (subtending 6°), presented on the left of a computer screen (Wass et al., 2012). When the child fixated the target, it moved, fluttering its wings and "flying" horizontally across the screen for as long as the child fixated it (moving 2°/s). Distractors, consisting of a house, a tree, and clouds (subtending 5°–15°) appeared in the periphery and scrolled in the opposite direction, moving at the same rate as the butterfly. Any time the child looked to one of the distracters, the target froze and the distracters disappeared, leaving only the static target on the screen. When the child re-fixated the target, it recommenced moving and the distracters reappeared and continued scrolling. Trials lasted approximately 15 s (dependent on performance) and were accompanied by an engaging sound track (the melody, Zip-a-Dee-Doo-Dah). There were two blocks of trials, with 9 trials/block, three trials each with 1, 2, and 3 distractors (with distractors presented in pseudo-random order), for a total of 18 trials. Each block lasted less than 2.5 min. *Measures*: percentage of time spent looking at the target on each trial; number of looks to the distractors; latency to initiate the first look to the target on each trial (SRT).

Disengaging and Shifting Attention

A gap/overlap task was used to assess the child's ability to disengage and shift attention. All trials began with a 500 ms presentation of a blank white screen, followed by the appearance of a colorful, animated, rotating clock face in the center of the screen (4.5°). As soon as the child fixated the central stimulus, a pulsating cartoon cloud (3°) appeared 20° to the left or right of the central stimulus; as soon as the child fixated the cloud, it disappeared and was replaced by a cartoon picture that performed a brief animation before disappearing. If the central stimulus was not fixated within 5,000 ms, the lateral target was presented anyway, and remained until fixated, or until 4,000 ms had elapsed. There were two types of trials: *baseline* (*gap*) *trials*, where the central target disappeared when the peripheral target was presented; and *overlap trials*, where the central target remained present, competing for attention with the peripheral one. Lateral targets appeared equally often, and unpredictably, to the left and right of center. There were two blocks of trials, with 12 trials/block, (6 baseline and 6 overlap) for a total of 24 trials. Each block lasted less than 2 min. The measures focus on the likelihood of shifting attention when disengagement is and is not required and the extent to which the presence of the competing stimulus slows reaction time to the peripheral target. *Measures*: % trials with gaze shifts from the central to the peripheral target; SRTs of gaze shifts to the peripheral target.

Selective Attention

A search task (adapted from Kaldy et al., 2011) was used to assess selective attention. Here, the child had to find a target (red apple) in an array containing varying numbers of randomly positioned distracters. In *single feature trials*, the distractors differ from the target in only one feature (color *or* shape), creating a pop-out effect; in *conjunction feature trials*, the distractors differed from the target in both features (color *and* shape), and finding the target was thought to require serial search. A figure illustrating the different trial types can be found in Rose et al. (2019b).

At the start of each trial the red apple (5°) appeared alone in the center of the screen for 1,000 ms (emitting an attractive "oh" sound), vanished, and then immediately re-appeared, randomly placed among distractors (blue apples; red cylinders). When the child looked to the target (or 4,000 ms elapsed), the trial ended and the target spun and made an engaging sound. Each block began with four familiarization trials, where the target was presented along with the two distracts (blue apple; red cylinder); test trials followed immediately, with single- and conjunction-feature trials intermixed. There were two blocks of test trials, with 13 test trials/block: 4 single-feature trials (2 containing five items; 2 containing nine items) and 9 conjunction trials (4 containing five items; 4 containing nine items; 1 containing 13 items). In all cases the arrays contained one target, with the rest of the items being distractors. There were a total of 26 trials for this task (13 per block). Each block lasted less than 2 min. *Measures*: number of targets located; time to fixate the target (SRT).

Data Analysis

General Considerations

All measures were examined for normality and outliers. Because their distributions were skewed, SRT latencies were log transformed for statistical analysis, while values more than 3 *SD* away from the mean were winsorized. (Raw scores are reported in the tables for ease of understanding; log transformed scores are presented in the histograms.) SPSS (version 25) was used for analyses.

Performance

Findings for each measure were analyzed using either paired sample *t*-tests or repeated measures analysis of variance (ANOVA), depending on the number of conditions involved. All effects were evaluated at a .05 level of significance and effect sizes for ANOVAs are reported as partial eta squared (η_p^2). Relations between clinical characteristics and performance variables were assessed with Pearson correlations.

Reliability

Cronbach's coefficient alpha (α), which is equivalent to performing split half analyses on all combinations of data points or trials, was used to assess reliability where scores fell on a continuum. This measure is a function of the number of test items and their average inter-correlation. Since α is sensitive to the number of items included in the analysis and, as noted above, it is often difficult to get a high number of usable trials from children with RTT, where possible these analyses were repeated with varying numbers of trials included in the analyses. Doing so allowed us to examine the impact of increasing trials on reliability in children with RTT and to determine the number of trials needed to achieve satisfactory reliability.

As is customary, split-half coefficients (with the Spearman–Brown correction) were used to assess the reliability when the scoring was binary (e.g., did vs. did not disengage). The split-half formula is specially designed for dichotomous data and the Spearman–Brown correction compensates for having only a single split involved in computing the coefficient. We used the popular odd–even split, comparing the sum of scores on even trials with the sum of scores on odd trials.

Finally, for the search task, where the number of trials available for each condition was quite limited, we assessed reliability in two ways. First, we considered each condition separately, using the individual trials as data points (as we had done for the first two tasks). Second, we used condition-averaged means as "items" for computing Cronbach's α (Thigpen et al., 2017; Van de Weijer-Bergsma et al., 2015). For this metric, high consistency between condition-averaged measures would indicate that individual participants retain the same rank ordering of performance across the different conditions. Such analysis is warranted when, as here, the condition-averaged means are important for hypothesis testing.

To evaluate reliability, we adopted the breakdown suggested by (Hinton et al. (2004) and recently used by Thigpen et al. (2017), which characterizes α levels as follows: <.50 (poor); .50–.70, (moderate; acceptable; adequate); .70–.90 (high; good); >.90 (excellent).

Supplementary Analyses

To examine the possibility that α may underestimate reliability, we also carried out a series of supplementary analyses, re-computing internal consistency reliability using Guttman's lambda 2 (λ_2), the lower bound estimate that has proven best in simulation studies (Guttman, 1945).

Results

Sustained Attention

Performance

On average, children looked at the target (butterfly) on 70.2% of the trials ($SD = 21.58$). Three measures of performance were examined for these trials: (a) the percentage of time looking at the target, (b) the number of looks away from the target to the distractors, and (c) latency of the first look to the target (see Table 1; Figure 1 presents histograms showing the response frequency for the various measures). Each measure was analyzed in a repeated measures ANOVA, with number of distractors (1, 2, or 3) as the repeated factor.

As can be seen in Table 1, the more distractors there were, the less time the children spent looking at the target, with the percentage of time looking dropping from 33% to 28% as the number of distractors increased from one to three. This drop off was significant, $F(2, 112) = 6.75$, $p = .002$, $\eta_p^2 = .11$ and followed a linear trajectory, as indicated by the significant linear component of the main effect, $F(1, 56) = 13.55$, $p = .001$, $\eta_p^2 = .20$.

The more distractors there were, the more frequently the children also looked away from the target, with the average number of looks away per trial increasing from 3.5 to 4.4 as the number of distractors increased from one to three, $F(2, 112) =10.35$, $p < .001$, $\eta_p^2 = .16$ The linear component of this effect was significant, $F(1, 56) = 14.41$, $p < .001$, $\eta_p^2 = .20$, as was the quadratic component, $F(1, 56) = 4.55$, $p < .05$, $\eta_p^2 = .08$, indicating that the number of looks away from the target was most pronounced when all three distractors were present.

Children took around 2.5 s to fixate the target at the outset of each trial. The latency to engage the target was independent of the number of distractors present, $F(2, 112) = .90$, ns, $\eta_p^2 = .02$, as would be expected, given that distractors did not appear until the child was already looking at the target.

Relation of Age at Test and RSSS to Performance. Correlations relating age at test and RSSS to the performance variables shown in Table 1 (combined over distractors), and to the number of trials completed, revealed two significant effects, both involving age. Older children showed greater sustained attention, as indexed by more time spend looking at the target ($r = .38$, $p < .01$) and were less impacted by distractors—more distractors had to be moving in the periphery to interfere with sustained attention, that is, to induce the child to look away from the target ($r = .45$, $p = .001$). RSSS was not related to any of performance variables and neither factor was related to the number of task trials completed.

Reliability

Alphas for the three measures are shown in Table 2. The alphas were computed for aggregates of 6, 9, and 12 trials, and are shown along with the average inter-trial correlations. All children had data on 6 trials (the minimum set for inclusion in the sample) and half the sample had data on at least 12 trials.

The alphas for two of the three measures—percentage of time looking to the target and number of looks to the distractors—were uniformly high, with alphas for the percentage of looking to the target ranging from .72 for 6 trials to .77 for 12 trials, and alphas for the number of looks to the distractors ranging from .71 to .88 over this span. Alpha coefficients tended to increase as successively more trials were aggregated, particularly for the number of looks away. The average inter-item correlations ranged from $r = .21$ to .38.

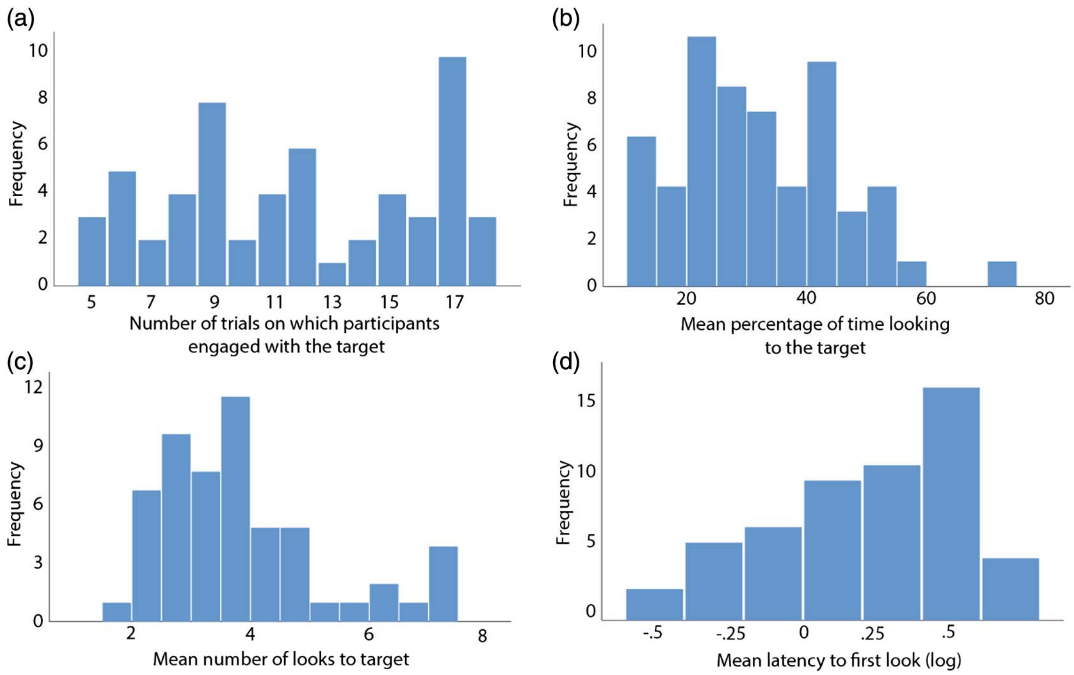
By contrast, alphas were lower for the latency of first look to the target, ranging from .53 (6 trials) to .65 (12 trials), with the average inter-item correlation remaining around $r = .15$. While the reliability for this measure increased as more trials were aggregated, it

Table 1
Sustained Attention: Basic Measures on Butterfly Task

		Number of distractors					
		One		Two		Three	
Measure	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Time spent looking at target (%)	57	33.51	16.06	32.07	17.68	28.25	14.08
Looks away from the target to the distractors (#)	57	3.52	1.43	3.60	1.87	4.39	1.74
Latency of first look to target (s)	57	2.29	1.59	2.21	1.70	2.56	2.01

Figure 1

Histograms for the Sustained Attention Task Showing: (a) the Number of Trials on Which Participants Engaged With the Target (Butterfly); (b) the Mean Percentage of Time Spent Looking at the Target on Each Trial; (c) the Mean Number of Looks to Target Per Trial; (d) the Mean Latency of the First Look Per Trial



Note. Results are averaged across trials with one, two, and three distractors. See the online article for the color version of this figure.

remained in the moderate range for all three levels of aggregation. The lower reliability for the latency measure is probably related to the shape of the distribution, which retained some skew even after a log transformation (see Figure 1d).

Supplementary Analysis. None of the Guttman Lambda 2 (λ_2) coefficients varied appreciably from the alphas shown in Table 2. The increase in these lower bound coefficients was on the order of .01–.02 for the various aggregates of (a) percentage time looking at the target and (b) number of looks away, and did not exceed .05 for any of the aggregates of the latency measure.

Disengaging and Shifting Attention

Performance

For trials to be scorable, the child had to fixate the center target at the start of the trial; the children met this criterion on 84.9% of the trials (SD = 15.74), indicating that the task effectively captured their interest and attention. Paired sample t-tests were used to compare performance on the baseline and overlap trials on the two key measures—the frequency of looks to the peripheral target and the time it took to look from the central target to the peripheral one (SRT). SRTs in the baseline and overlap conditions were measured as the time elapsed between the appearance of the peripheral target and the reported position of gaze leaving the central area (a 9° box around the central target). For an SRT to be considered valid the child had to: (a) look to the central stimulus location, and (b) make an eye movement to the peripheral target within 4000 ms of its onset. If there was no shift in gaze toward the peripheral target within this period, no SRT was recorded and the trial was considered a failure to disengage.

As can be seen in Table 3 and Figure 2, children shifted their attention from the central to the peripheral target on close to 75%

Table 2

Sustained Attention: Internal Consistency Reliability (Cronbach α)

Measure	Six trials			Nine trials			Twelve trials		
	N	α	Mean interitem correlation	N	α	Mean interitem correlation	N	α	Mean interitem correlation
Time spent looking at target (%)	57	.72	.30	42	.70	.21	28	.77	.22
Looks away from the target to the distractors (#)	57	.71	.29	42	.78	.29	28	.88	.38
Latency of first look to target (s)	57	.53	.16	42	.58	.15	28	.65	.15

Table 3
Disengagement of Attention: Basic Measures on the Gap/Overlap Task

Measure	*N*	*M*	*SD*
Frequency of shifts to peripheral target (%)			
Baseline (gap) condition	57	73.39	22.07
Overlap condition	57	58.33	26.81
Latency of shifts to peripheral target (ms)			
Baseline (gap) condition	57	581.73	237.14
Overlap condition	57	881.20	440.73

of the baseline trials, but on fewer than 60% of the overlap trials. A *t*-test comparing the two values was significant, $t(56) = 5.29$, $p < .001$, indicating that children had considerably more difficulty making the shift on overlap trials (where disengagement was required). Additionally, a t-test of the SRTs showed that even when the children did shift their gaze to the peripheral target, their SRT were significantly slower on overlap trials (881 ms) than on baseline trials (582 ms), $t(55) = 4.68$, $p < .001$ (see Figure 2c and d).

Relation of Age at Test and RSSS to Performance. The correlations relating age at test and RSSS to the performance variables shown in Table 3, and number of trials completed, were all low and non-significant for this task.

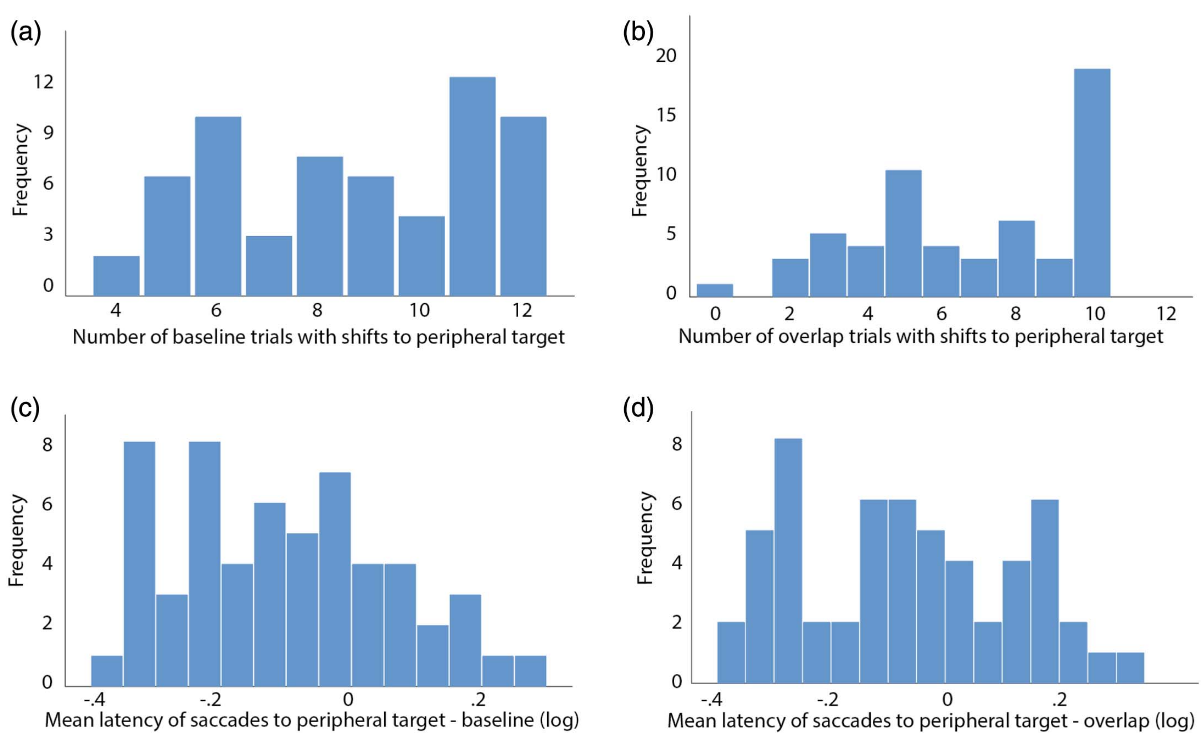
Reliability

The split-half reliability coefficients computed to examine the reliability of the frequency of looks to the peripheral target are shown in Table 4. As can be seen, reliability was excellent for both baseline and overlap trials (>.95). These coefficients, based on odd–even split of trials, indicate that children's tendency to shift attention from the central to the peripheral target was highly consistent for both types of trials.

Alphas for the SRTs are shown in Table 5, along with the average inter-trial correlations. Alphas were computed for aggregates of 5, 8, and 11 trials. (Aggregates here are based on fewer trials than for sustained attention because this task had fewer trials per condition.) As can be seen, on both baseline and overlap trials, reliability increased as more trials were aggregated. On baseline trials, alphas indicated reliability, while only moderate (.60) for the smallest aggregate, was high for both larger aggregates (.71 and .78). Alpha coefficients on overlap trials showed a similar increase as trials were aggregated, but reached a high level of reliability (.71) only with aggregates of 11 trials. (Average inter-item correlations ranged from $r = .15$ to .27, remaining mainly in the .20s.) The reliability of the SRT measures may be lower than that of the frequency scores because, once again, the SRT scores retained some skew even after transformation.

Supplementary Analysis. Again, Guttman Lambda 2 (λ_2) coefficients did not vary appreciably from the alphas shown in Table 5.

Figure 2
Histograms for the Disengaging/Shifting Attention Task Showing: (a) and (b) the Frequency of Shifts to the Peripheral Target on Baseline (a) and Overlap (b) Trials; (c) and (d) Mean Saccadic Reaction Times (SRT) for Those Trials on Which the Children Shifted Their Gaze for Baseline (c) and Overlap (d) Trials



Note. See the online article for the color version of this figure.

Table 4
Disengagement of Attention: Internal Consistency Reliability (Split-Half Coefficients)

Measure	*N*	Split-half coefficient
Frequency of shifts to peripheral target (%)		
Baseline condition	57	.96
Overlap condition	57	.97

The increase in reliability brought about by using these lower bound coefficients, instead of α, again ranged only from .01 to .05.

Selective Attention

Performance

Success rates for both single-feature and conjunction-feature trials are shown in Table 6. (For ease of presentation, data from both types of trials are averaged for the histograms shown in Figure 3). The success rates shown in Table 6 suggest that children found this task relatively difficult. On single feature trials, the overall mean success rate, though relatively low for both set sizes, was significantly lower for the five-item displays (39%) than the nine-item ones (50%), $t(53) = 2.34$, $p < .05$. For the conjunction-feature trials, which used arrays with 5, 9, and 13 items, success rates varied narrowly, from 35% to 41%, and did not differ significantly by array size, $F(2, 106) = 1.13$, $p = .33$, $\eta_p^2 = .02$.

For those trials in which children were successful in finding the target, the mean SRTs to do so are also shown in Table 6 (and in Figure 3c and d). There was no significant difference in SRTs as a function of array size in either condition: in the single-feature condition (which had 5- and 9-item arrays), $t(34) = .66$, $p = .51$; in the conjunction-feature condition (which had 5-, 9-, and 13-item arrays), $F(2, 50) = 2.10$, $p = .13$, $\eta_p^2 = .08$.

Relation of Age at Test and RSSS to Performance. The correlations relating age at test and RSSS to the performance variables shown in Table 6 were all low and non-significant.

Reliability

Split-half coefficients for success rates, and alphas for the SRTs, are shown in Table 7. Here alphas were computed for aggregates of 2, 3, and 4 trials. The values for both split-half coefficients and alphas are uniformly quite poor and some are negative. The low values are not surprising, given that there were relatively few trials for computing split half coefficients (2–8) in any of the conditions, and fewer yet for computing alphas for the SRTs. This falloff in available trials occurs because SRTs are calculated only for trials on which the child successfully found (fixated) the target, and children failed to find the target on close to 60% of the trials; and also, once again, the distribution of SRTs retained some skew even after transformation (see Figure 3c and d).

Reliability for this task we also assessed by re-computing Cronbach's α, this time using the five condition-averaged means as "items" (following Thigpen et al., 2017). For this metric, high consistency between condition-averaged measures would indicate that individual participants retain the same rank ordering of performance across the five conditions.

As can be seen in Table 8, this approach resulted in acceptable reliability for success rates (.70), indicating that the children tended to be consistent across the five conditions in their ability to find the target. However, reliability for the SRT measures remained low (.13). Overall then, reliability for this task was limited.

Supplementary Analysis. Again, the values for most of Guttman Lambda 2 (λ_2) coefficients were either unchanged, or little changed, from that of their alpha counterparts shown in Table 7. While three negative values did increase appreciably, unfortunately, none of the three λ_2 coefficients involved exceeded .12. Thus, reliability remained poor for these aggregates even with this alternative lower bound estimate. The values obtained when conditions were treated as items (Table 8) were similar whether reliability was assessed with Cronbach's α or Guttman's Lambda 2.

Discussion

The main purpose of this study was to determine whether internal consistency reliability could be established for children with Rett syndrome on three core aspects of attention that are critical for cognitive growth: sustained attention, disengaging and shifting attention, and selective attention. These aspects of attention were assessed using three experimental, gaze-based tasks, all of which bypass the motoric and verbal difficulties prominent in this disorder. In earlier work, where we compared the performance of children with RTT to that of their age- and gender-matched typically developing peers on these tasks, we identified the nature of the deficits associated with RTT on all three aspects of attention and quantified the extent of these deficits. However, for measures from this foundational work to serve as important endpoints in assessing the cognitive efficacy of clinical interventions, or to be considered targets for cognitive training, they must have good internal consistency reliability. That is, the responses of children must be shown to be consistent across trials or items.

The results indicate that children with Rett Syndrome showed acceptable to excellent reliability on almost all the key measures of attention on two of the three tasks—sustained attention and disengaging/shifting attention—with split-half coefficients and Cronbach

Table 5
Disengagement of Attention: Internal Consistency Reliability (Cronbach α)

	Five trials			Eight trials			Eleven trials		
Measure	*N*	α	Mean inter-item correlation	*N*	α	Mean inter-item correlation	*N*	α	Mean inter-item correlation
Latency of shifts to peripheral target (ms)									
Baseline condition	55	.60	.22	34	.71	.22	19	.78	.27
Overlap condition	44	.58	.21	27	.62	.15	13	.71	.27

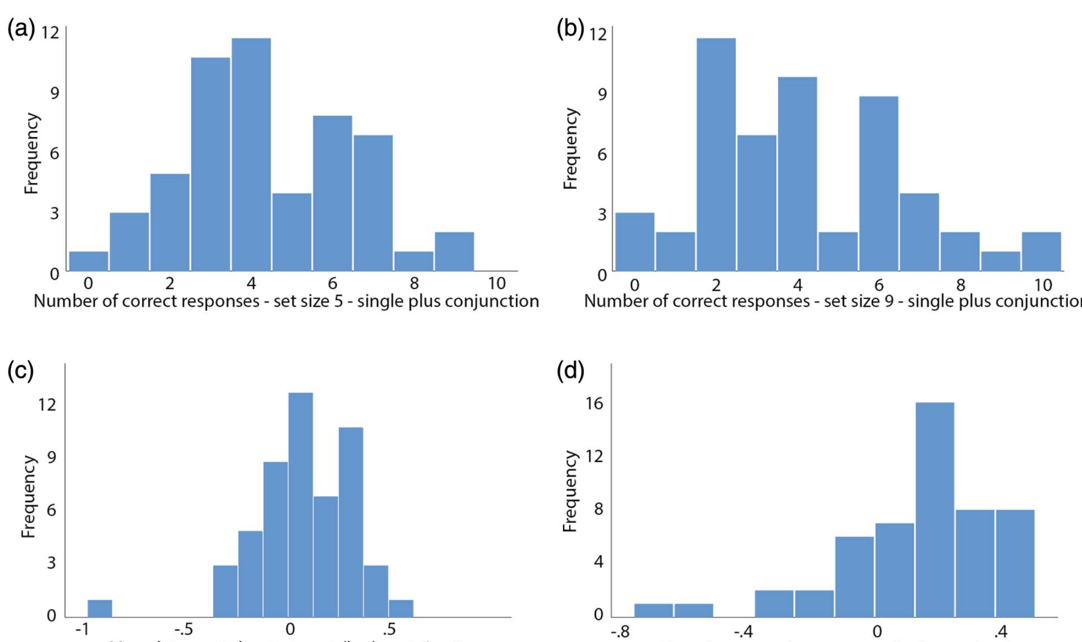
Table 6
Selective Attention: Basic Measures on the Search Task

Measure	Set size 5			Set size 9			Set size 13		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Success in finding target (% correct)									
Single feature condition	54	39.81	27.69	54	50.46	26.37	—	—	—
Conjunction feature condition	54	40.97	20.66	54	35.19	24.86	54	35.19	37.15
Latency to find target (ms)									
Single feature condition	43	1475	1022	43	1733	983	—	—	—
Conjunction feature condition	52	1458	731	49	1750	785	29	1358	1048

alphas ranging from .70 to .93. Thus, these two tasks, both of which measure key aspects of attention, appear to be appropriate for clinical and translational work. That is, we can use these measures without worrying that their predictive validity will be attenuated because of poor reliability. The results also provided guidelines on the number of trials needed on each to achieve acceptable levels of reliability for the specific measures from each task. Given the paucity of measures to evaluate cognition in this population, these finding constitutes a welcome advance for the area.

For *sustained* attention, where the child had to initiate and maintain attentional focus on a target (butterfly) to have it move/flutter across the screen (while ignoring distractors that moved in the periphery), children showed high levels of reliability (with $\alpha \geq .70$) on two key measures—the percentage of time looking at the target and the frequency with which their attention was drawn to the distractors—with aggregates containing as few as six trials; they showed acceptable levels of reliability in their latency to fixate the target with aggregates of 9 and 12 trials. The lower reliability for the latency variable appears to be due to the high degree of variability shown by individuals in the time taken to engage with the target, and to the scores having retained some skew even after log transformation. Because alpha tends to increase as the number of trials or items increase (assuming the added trials are equally good at measuring the variable of interest), we had estimated internal consistency over aggregates that included differing number of trials. In all cases, reliability continued to increase systematically as the number of trials aggregated increased. Overall then, the children showed high levels of reliability on the two measures reflecting sustained attention to the target, and a moderate level of reliability for the measure reflecting their latency to engage with the target.

Figure 3
Histograms for the Selective Attention Task Showing: (a) and (b) the Number of Correct Responses—i.e., Trials on Which Children "Found" the Target in the Search Task (i.e., Looked at It)—for Set Size 5 (a) and Size 9 (b) Trials; (c) and (d), the Mean Latency to Locate the Target on Trials Where the Children Were Successful for Set Size 5 (c) and Set Size 9 (d) Trials



Note. Trials have been averaged over single and conjunction feature trials for all four histograms shown here. See the online article for the color version of this figure.

Table 7
Selective Attention: Internal Consistency Reliability (Cronbach α and Split-Half Coefficients)

	α for SRTs—values for differing number of trials						Split-half coefficients for success rate	
	Two		Three		Four			
Condition	*N*	α	*N*	α	*N*	α	*N*	Split-half
Single feature								
Set size 5	29	.12	10	.50			54	.08
Set size 9	29	.07	8	.11			54	.36
Conjunction feature								
Set size 5	45	.27	28	.08	22	−.22	54	−.15
Set size 9	37	−.24	20	−.25	14	−.97	54	.65
Set size 13	8	.35					54	.20

For *disengaging and shifting attention*, where the task requires the child to switch attention from a central to a peripheral target, split-half reliabilities for the percentage of shifts in gaze to peripheral target were excellent (>.95), both for baseline trials (central target disappears with the onset of the peripheral one) and for overlap trials (central target remains present, competing for attention with the peripheral one). Cronbach's alphas for the SRTs reached levels of .70 or better for aggregates of 8 or more trials in the baseline condition, and for aggregates of 11 in the overlap condition. Here, too, α coefficients increased systematically in both baseline and overlap conditions as the number of trials aggregated increased. (Since reliabilities depend on the number of trials aggregated, it is not surprising that the reliability estimates are somewhat lower for the SRTs than success rates, since SRTs are based only on the fraction of trials in which a shift actually occurs, whereas success rates are based on all trials.) Overall then, the children showed excellent levels of reliability in the disengagement/shifting task on measures of success rate (frequency of looks to the peripheral target) and achieved high levels of reliability on measures assessing the rapidity of their responding.

For *selective attention*, where the child had to find a target hidden among a varying number of distractors that differed from it either in a single feature or in a conjunction of features, split-half reliabilities of success rates, and alpha coefficients of the SRTs were all quite poor. The situation improved somewhat when we computed Cronbach's alpha using the five condition-averaged means as "items." Here, Cronbach's alpha for success rate reached .70, but reliability continued to be low for the SRTs.

The low reliability for the SRT measures of search in both analyses may have been due to the small number of trials/items available for aggregation and to the fact that SRTs, when observed, were generally skewed (see Figure 3c and d). Since SRTs are available only for successful trials, and children did not succeed in finding the target on close to 60% of the trials, alphas for these measures are based on a small number of trials. With small sample sizes and small numbers of items, sampling error could produce low or negative average covariance in a given sample of cases, even when the true population covariances among items are positive. This problem could be addressed in future work by increasing the number of trials presented in each condition. It is also possible, of course, that children with RTT are highly variable from one trial to the next in their search strategies and thus that the trials/items do not truly have positive covariances. If the latter is the case, then the reliability of the SRTs would remain poor even with higher trial counts.

Table 8
Selective Attention: Internal Consistency Reliability (Cronbach α) With Conditions Treated as Items

	Five conditions treated as items		
Measure	*N*	α	Mean inter-item correlation
Success in finding target (% correct)	54	.70	.33
Latency to find target (ms)	19	.13	.04

Overall, the findings indicate that all the key measures from the tasks assessing sustained attention and disengaging/shifting attention show moderate to excellent reliability, whereas measures from selective attention showed limited reliability at best. It should be noted that none of the findings related to reliability were appreciably altered by substituting Guttman's lambda 2, an alternative lower-bound estimate, for Cronbach's alpha. Thus, there was no indication that the reliability of any of the measures examined here was underestimated by Cronbach's alpha.

The levels of performance seen on these tasks also complement the findings of earlier work, which had provided evidence of the atypicalities associated with RTT and evidence for the clinical validity of all three tasks (Rose et al., 2017, 2019a, 2019b). In that work, performance on each task was shown to differentiate children with Rett syndrome from that of their age- and gender-matched typically developing peers. The levels of impairment found in the present study are marked, and similar to those found earlier. For example, in the present study, children with RTT sustained attention on the butterfly only about half as long as did their TD counterparts in the earlier work (31% vs. 62%; see Rose et al., 2017)); shifted attention from the central to peripheral target in gap/overlap task on many fewer trials (66% vs. 92%; see Rose et al., 2019a); and were far less successful in finding the target in the search task (41% vs. 84%; see Rose et al., 2019b). While sustained attention improved over age, age had little impact on any of the measures of disengagement/shifting attention or selective attention. The severity of the children's physical symptoms, as assessed with the RSSS, was not related to performance on any of these cognitive tasks, and neither age nor the RSSS was related to the number of trials completed in a task.

Overall, the findings suggest that the levels of impairment found initially accurately characterize the abilities of this population. These atypicalities in three such basic aspects of attention are of particular concern because, as indicated earlier, they foreshadow deficits and deficiencies in areas that are key to cognitive growth and predictive of later IQ, executive functions, and academic achievement, as well as areas that are related to social/emotional functioning and behavioral regulation (Posner, 2001). Thus the deficits we found in RTT are clinically meaningful hallmarks of intellectual delay that have real-world implications. Recent reports that attention can be improved in Rett Syndrome through training in discrimination learning and digital games (Fabio et al., 2016, 2019) are encouraging and should be pursued. In addition, the training program developed by Wass (Forssman et al., 2014; Wass et al., 2011), which focuses exclusively on training attention control, using gaze-based measures, might well benefit these children.

Study Limitations

There are some limitations to our work. First, many children were unable to complete the tasks or accrue enough trials to reach acceptable levels of reliability. That is, there was a tradeoff between feasibility/test completion and reliability, as reflected in the increasing drop-out rates that occurred as more trials were aggregated. These children suffer from a variety of physical problems and state lability that often make it difficult for them to stay "on-task" or cooperate for extended periods of time. This problem could be addressed by increasing the number of trial blocks. This could be done either in a single session (provided sufficient breaks are given) or, as Ahonniska-Assa and colleagues did, by extending the assessment over several days (2018). Actually, neither approach would be burdensome, since each task is quite short (with a block of trials taking only 2–3 min). A second limitation revolves around the limited selection of tasks used here. Our tasks all focused on attention; future studies should include gaze-based tasks designed to assess other facets of cognition, including memory and executive function. Third, we examined only one psychometric property, namely internal consistency reliability. Future studies should be expanded to include assessments of test–retest reliability and stability; doing so would provide information as to how consistent the children are over time.

Conclusion

The findings indicate that key measures from tasks assessing two different aspect of attention—sustained attention and disengagement of attention—show moderate to excellent reliability in children with RTT, provide guidance as to the minimum number of trials needed for acceptable reliability for the various measures from both tasks, and indicate ways in which the limited reliability of the third task, visual search, might be improved. To our knowledge, these findings are unique, constituting the first in-depth examination of the psychometric properties of gaze-based measures in RTT. The findings also reinforce and extend our understanding of the nature and extent of the cognitive impairments that characterize RTT. Finally, it should be noted that the measures of attention we have been examining here have several features that make them well suited as potential outcome measures for clinical trials. These include (a) posing little administrative or respondent burden; (b) being cost effective; (c) being objective, performance-based assessments; (d) having potential applicability to multiple developmental disorders; and (e) showing comparable deficits in learning to those found in mouse models of RTT (Katz et al., 2012). Identifying the cognitive effects of therapeutic interventions is important for increasing the acceptance of these interventions by regulatory bodies.

References

Ahonniska-Assa, J., Polack, O., Saraf, E., Wine, J., Silberg, T., Nissenkorn, A., & Ben-Zeev, B. (2018). Assessing cognitive functioning in females with Rett syndrome by eye-tracking methodology. *European Journal of Paediatric Neurology*, *22*(1), 39–45. https://doi.org/10.1016/j.ejpn.2017.09.010

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. https://doi.org/10.1038/13810

Amso, D., & Scerif, G. (2015). The attentive brain: Insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, *16*(10), 606–619. https://doi.org/10.1038/nrn4025

Baptista, P. M., Mercadante, M. T., Macedo, E. C., & Schwartzman, J. S. (2006). Cognitive performance in Rett syndrome girls: A pilot study using eyetracking technology. *Journal of Intellectual Disability Research*, *50*(Pt 9), 662–666. https://doi.org/10.1111/j.1365-2788.2006.00818.x

Bendermacher, N. (2010). Beyond alpha: Lower bounds for the reliability of tests. *Journal of Modern Applied Statistical Methods*, *9*(1), 95–102. https://doi.org/10.22237/jmasm/1272687000

Blankenship, T. L., Slough, M. A., Calkins, S. D., Deater-Deckard, K., Kim-Spoon, J., & Bell, M. A. (2019). Attention and executive functioning in infancy: Links to childhood executive function and reading achievement. *Developmental Science*, *22*(6), Article e12824. https://doi.org/10.1111/desc.12824

Bornstein, M. H., Hahn, C. S., & Wolke, D. (2013). Systems and cascades in cognitive development and academic achievement. *Child Development*, *84*(1), 154–162. https://doi.org/10.1111/j.1467-8624.2012.01849.x

Bowden, S. C., & Finch, S. (2017). When is a test reliable enough and why does it matter? In S. C. Bowden (Ed.), *Neuropsychological assessment in the age of evidence-based parctice: diagnostic and treatment evaluations* (pp. 95–120). Oxford University Press.

Chahrour, M. (2007). 'A civilizing mission'? Austrian medicine and the reform of medical structures in the Ottoman Empire, 1838–1850. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, *38*(4), 687–705. https://doi.org/10.1016/j.shpsc.2007.09.005

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Cornish, K., Kogan, C., Turk, J., Manly, T., James, N., Mills, A., & Dalton, A. (2005). The emerging fragile X premutation phenotype: Evidence from the domain of social cognition. *Brain and Cognition*, *57*(1), 53–60. https://doi.org/10.1016/j.bandc.2004.08.020

Cronbach, L. J., & Warrington, W. G. (1951). Time-limit tests: Estimating their reliability and degree of speeding. *Psychometrika*, *16*(2), 167–188. https://doi.org/10.1007/BF02289113

Csibra, G., Johnson, M. H., & Tucker, L. A. (1997). Attention and oculomotor control: A high-density ERP study of the gap effect. *Neuropsychologia*, *35*(6), 855–865. https://doi.org/10.1016/S0028-3932(97)00016-X

Csibra, G., Tucker, L. A., & Johnson, M. H. (1998). Neural correlates of saccade planning in infants: A high-density ERP study. *International Journal of Psychophysiology*, *29*(2), 201–215. https://doi.org/10.1016/S0167-8760(98)00016-6

Cuevas, K., & Bell, M. A. (2014). Infant attention and early childhood executive function. *Child Development*, *85*(2), 397–404. https://doi.org/10.1111/cdev.12126

de Jong, M., Verhoeven, M., Hooge, I. T. C., & van Baar, A. L. (2016). Introduction of the Utrecht Tasks for Attention in toddlers using eye tracking (UTATE): A pilot study. *Frontiers in Psychology*, *7*, Article 669. https://doi.org/10.3389/fpsyg.2016.00669

Djukic, A., Holtzer, R., Shinnar, S., Muzumdar, H., Rose, S. A., Mowrey, W., Galanopoulou, A. S., Shinnar, R., Jankowski, J. J., Feldman, J. F., Pillai, S., & Moshé, S. L. (2016). Pharmacologic treatment of Rett syndrome with glatiramer acetate. *Pediatric Neurology*, *61*, 51–57. https://doi.org/10.1016/j.pediatrneurol.2016.05.010

Djukic, A., & McDermott, M. V. (2012). Social preferences in Rett syndrome. *Pediatric Neurology*, *46*(4), 240–242. https://doi.org/10.1016/j.pediatrneurol.2012.01.011

Djukic, A., Rose, S. A., Jankowski, J. J., & Feldman, J. F. (2014). Rett syndrome: recognition of facial expression and its relation to scanning patterns. *Pediatric Neurology*, *51*(5), 650–656. https://doi.org/10.1016/j.pediatrneurol.2014.07.022

Djukic, A., Valicenti McDermott, M., Mavrommatis, K., & Martins, C. L. (2012). Rett syndrome: Basic features of visual processing—A pilot study of eye-tracking. *Pediatric Neurology*, *47*(1), 25–29. https://doi.org/10.1016/j.pediatrneurol.2012.04.009

Dragich, J., Houwink-Manville, I., & Schanen, C. (2000). Rett syndrome: A surprising result of mutation in MECP2. *Human Molecular Genetics*, *9*(16), 2365–2375. https://doi.org/10.1093/hmg/9.16.2365

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., Johnson, M. H., & the British Autism Study of Infant Siblings Team. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*(3), 189–194. https://doi.org/10.1016/j.biopsych.2012.11.030

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., Bolton, P., Charman, T., Baird, G., & Johnson, M. H. (2009). Visual orienting in the early broader autism phenotype: Disengagement and facilitation. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *50*(5), 637–642. https://doi.org/10.1111/j.1469-7610.2008.02051.x

Fabio, R. A., Billeci, L., Crifaci, G., Troise, E., Tortorella, G., & Pioggia, G. (2016). Cognitive training modifies frequency EEG bands and neuropsychological measures in Rett syndrome. *Research in Developmental Disabilities*, *53–54*(1), 73–85. https://doi.org/10.1016/j.ridd.2016.01.009

Fabio, R. A., Giannatiempo, S., Antonietti, A., & Budden, S. (2009). The role of stereotypies in overselectivity process in Rett syndrome. *Research in Developmental Disabilities*, *30*(1), 136–145. https://doi.org/10.1016/j.ridd.2008.01.003

Fabio, R. A., Giannatiempo, S., & Capri, T. (2019). Attention and identification of the same and the similar visual stimuli in Rett Syndrome. *Life Span and Disability*, *22*(1), 113–127.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, *26*(2), 471–479. https://doi.org/10.1016/j.neuroimage.2005.02.004

Forssman, L., Wass, S., & Leppanen, J. M. (2014). *Training non-social attention control improves infants' socio-cognitive abilities* [Poster presentation]. The Biennial International Conference on Infant Studies, Berlin, Germany.

Forssman, L., & Wass, S. V. (2018). Training basic visual attention leads to changes in responsiveness to social-communicative cues in 9-month-olds. *Child Development*, *89*(3), e199–e213. https://doi.org/10.1111/cdev.12812

Foxe, J. J., Burke, K. M., Andrade, G. N., Djukic, A., Frey, H. P., & Molholm, S. (2016). Automatic cortical representation of auditory pitch changes in Rett syndrome. *Journal of Neurodevelopmental Disorders*, *8*(1), 34. https://doi.org/10.1186/s11689-016-9166-5

Grahn, J. A., & Manly, T. (2012). Common neural recruitment across diverse sustained attention tasks. *PLOS ONE*, *7*(11), Article e49556. https://doi.org/10.1371/journal.pone.0049556

Green, S. B., & Yang, Y. (2009). Commentary on coefficient α: A cautionary tale. *Psychometrika*, *74*, 121–135. https://doi.org/10.1007/s11336-008-9098-4

Green, S. B., Yang, Y., Alt, M., Brinkley, S., Gray, S., Hogan, T., & Cowan, N. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin & Review*, *23*(3), 750–763. https://doi.org/10.3758/s13423-015-0968-3

Guttman, L. (1945). A basis for analyzing test–retest reliability. *Psychometrika*, *10*, 255–282. https://doi.org/10.1007/BF02288892

Hendry, A., Johnson, M. H., & Holmboe, K. (2019). Early development of visual attention: Change, stability, and longitudinal associations. *Annual Review of Developmental Psychology*, *1*(1), 251–275. https://doi.org/10.1146/annurev-devpsych-121318-085114

Hendry, A., Jones, E. J., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review*, *42*, 1–33. https://doi.org/10.1016/j.dr.2016.06.005

Hinton, P., McMurray, I., & Brownlow, C. (2004). *SPSS explained* (2nd ed.). Routledge. https://doi.org/10.4324/9780203642597

Humphreys, L. G., & Drasgow, F. (1989). Some comments on the relation between reliability and statistical power. *Applied Psychological Measurement*, *13*(4), 419–425. https://doi.org/10.1177/014662168901300407

Kaldy, Z., Kraper, C., Carter, A. S., & Blaser, E. (2011). Toddlers with Autism Spectrum Disorder are more successful at visual search than typically developing toddlers. *Developmental Science*, *14*(5), 980–988. https://doi.org/10.1111/j.1467-7687.2011.01053.x

Karatekin, C., Marcus, D. J., & White, T. (2007). Oculomotor and manual indexes of incidental and intentional spatial sequence learning during middle childhood and adolescence. *Journal of Experimental Child Psychology*, *96*(2), 107–130. https://doi.org/10.1016/j.jecp.2006.05.005

Katz, D. M., Berger-Sweeney, J. E., Eubanks, J. H., Justice, M. J., Neul, J. L., Pozzo-Miller, L., Blue, M. E., Christian, D., Crawley, J. N., Giustetto, M., Guy, J., Howell, C. J., Kron, M., Nelson, S. B., Samaco, R. C., Schaevitz, L. R., St Hillaire-Clarke, C., Young, J. L., Zoghbi, H., & Mamounas, L. A. (2012). Preclinical research in Rett syndrome: Setting the foundation for translational success. *Disease Models & Mechanisms*, *5*(6), 733–745. https://doi.org/10.1242/dmm.011007

Kaufmann, W. E., Johnston, M. V., & Blue, M. E. (2005). MeCP2 expression and function during brain development: Implications for Rett syndrome's pathogenesis and clinical evolution. *Brain & Development*, *27*(Suppl 1), S77–S87. https://doi.org/10.1016/j.braindev.2004.10.008

Kaufmann, W. E., Tierney, E., Rohde, C. A., Suarez-Pedraza, M. C., Clarke, M. A., Salorio, C. F., Bibat, G., Bukelis, I., Naram, D., Lanham, D. C., & Naidu, S. (2012). Social impairments in Rett syndrome: Characteristics and relationship with clinical severity. *Journal of Intellectual Disability Research*, *56*(3), 233–247. https://doi.org/10.1111/j.1365-2788.2011.01404.x

Lebel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*(4), 570–583. https://doi.org/10.1177/0146167211400619

LeBlanc, J. J., DeGregorio, G., Centofante, E., Vogel-Farley, V. K., Barnes, K., Kaufmann, W. E., Fagiolini, M., & Nelson, C. A. (2015). Visual evoked potentials detect cortical processing deficits in Rett syndrome. *Annals of Neurology*, *78*(5), 775–786. https://doi.org/10.1002/ana.24513

Leppänen, J. M., Forssman, L., Kaatiala, J., Yrttiaho, S., & Wass, S. (2015). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods*, *47*(2), 538–548. https://doi.org/10.3758/s13428-014-0473-z

Neul, J. L., Kaufmann, W. E., Glaze, D. G., Christodoulou, J., Clarke, A. J., Bahi-Buisson, N., Leonard, H., Bailey, M. E., Schanen, N. C., Zappella, M., Renieri, A., Huppke, P., Percy, A. K., & the RettSearch Consortium. (2010). Rett syndrome: Revised diagnostic criteria and nomenclature. *Annals of Neurology*, *68*(6), 944–950. https://doi.org/10.1002/ana.22124

Nobre, A. C., Coull, J. T., Walsh, V., & Frith, C. D. (2003). Brain activations during visual search: Contributions of search efficiency versus feature binding. *NeuroImage*, *18*(1), 91–103. https://doi.org/10.1006/nimg.2002.1329

Posner, M. I. (2001). Developing brains: The work of the Sackler Institute. *Clinical Neuroscience Research*, *1*, 258–266. https://doi.org/10.1016/S1566-2772(01)00012-3

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Rose, S. A., Djukic, A., Jankowski, J. J., Feldman, J. F., Fishman, I., & Valicenti-McDermott, M. (2013). Rett syndrome: An eye-tracking study of attention and recognition memory. *Developmental Medicine and Child Neurology*, *55*(4), 364–371. https://doi.org/10.1111/dmcn.12085

Rose, S. A., Djukic, A., Jankowski, J. J., Feldman, J. F., & Rimler, M. (2016). Aspects of Attention in Rett Syndrome. *Pediatric Neurology*, *57*, 22–28. https://doi.org/10.1016/j.pediatrneurol.2016.01.015

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Attention and recognition memory in the 1st year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology*, *37*(1), 135–151. https://doi.org/10.1037/0012-1649.37.1.1351

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, *14*(5), 1161–1175. https://doi.org/10.1111/j.1467-7687.2011.01068.x

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of infant cognition for executive functions at age 11. *Psychological Science*, *23*(11), 1345–1355. https://doi.org/10.1177/0956797612444902

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2005). Pathways from prematurity and infant abilities to later cognition. *Child Development*, *76*, 1172–1184. https://doi.org/10.1111/j.1467-8624.2005.00842.x-i1

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2008). A cognitive cascade in infancy: Pathways from prematurity to later mental development. *Intelligence*, *36*, 367–378. https://doi.org/10.1016/j.intell.2007.07.003

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2011). Basic information processing abilities at 11 years account for deficits in IQ associated with preterm birth. *Intelligence*, *39*, 198–209. https://doi.org/10.1016/j.intell.2011.03.003

Rose, S. A., Wass, S., Jankowski, J. J., Feldman, J. F., & Djukic, A. (2017). Sustained attention in the face of distractors: A study of children with Rett syndrome. *Neuropsychology*, *31*(4), 403–410. https://doi.org/10.1037/neu0000369

Rose, S. A., Wass, S., Jankowski, J. J., Feldman, J. F., & Djukic, A. (2019a). Attentional shifting and disengagement in Rett syndrome. *Neuropsychology*, *33*(3), 335–342. https://doi.org/10.1037/neu0000515

Rose, S. A., Wass, S., Jankowski, J. J., Feldman, J. F., & Djukic, A. (2019b). Impaired visual search in children with Rett syndrome. *Pediatric Neurology*, *92*, 26–31. https://doi.org/10.1016/j.pediatrneurol.2018.10.002

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology*, *28*(2), 573–594. https://doi.org/10.1207/s15326942dn2802_2

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research. Brain Research Reviews*, *35*(2), 146–160. https://doi.org/10.1016/S0165-0173(01)00044-3

Scerif, G., Longhi, E., Cole, V., Karmiloff-Smith, A., & Cornish, K. (2012). Attention across modalities as a longitudinal predictor of early outcomes: The case of fragile X syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *53*(6), 641–650. https://doi.org/10.1111/j.1469-7610.2011.02515.x

Schmidt, F. L., & Humter, J. E. (1999). Theory training and measurement error. *Intelligence*, *27*(3), 183–198. https://doi.org/10.1016/S0160-2896(99)00024-0

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *Journal of Neurophysiology*, *97*(1), 229–237. https://doi.org/10.1152/jn.00677.2006

Sysoeva, O. V., Molholm, S., Djukic, A., Frey, H., & Foxe, J. J. (2020). Atypical processing of tones and phonemes in Rett Syndrome as biomarkers of disease progression. *Translational Psychiatry*, *10*(1), 188. https://doi.org/10.1038/s41398-020-00877-4

Thigpen, N. N., Kappenman, E. S., & Keil, A. (2017). Assessing the internal consistency of the event-related potential: An example analysis. *Psychophysiology*, *54*(1), 123–138. https://doi.org/10.1111/psyp.12629

Towers, D. N., & Allen, J. J. (2009). A better estimate of the internal consistency reliability of frontal EEG asymmetry scores. *Psychophysiology*, *46*(1), 132–142. https://doi.org/10.1111/j.1469-8986.2008.00759.x

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's α reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*. Advance online publication. https://doi.org/10.3389/fpsyg.2016.00769

van Baar, A. L., de Jong, M., Maat, M., Hooge, I. T. C., Bogičević, L., & Verhoeven, M. (2020). Reliability and validity of the Utrecht tasks for attention in toddlers using eye tracking (UTATE). *Frontiers in Psychology*, *11*, Article 1179. https://doi.org/10.3389/fpsyg.2020.01179

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online visual-spatial working memory task for self-reliant administration in school-aged children. *Behavior Research Methods*, *47*(3), 708–719. https://doi.org/10.3758/s13428-014-0469-8

Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology*, *21*(18), 1543–1547. https://doi.org/10.1016/j.cub.2011.08.004

Wass, S. V. (2015). Applying cognitive training to target executive functions during early development. *Child Neuropsychology*, *21*(2), 150–166. https://doi.org/10.1080/09297049.2014.882888

Wass, S. V., Forssman, L., & Leppänen, J. M. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analysis. *Infancy*, *19*(5), 427–460. https://doi.org/10.1111/infa.12055

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory—Is younger, better? *Developmental Review*, *32*(4), 360–387. https://doi.org/10.1016/j.dr.2012.07.001

Received November 16, 2020
Revision received March 18, 2021
Accepted March 18, 2021 ■