1-12-2023

Psychometric Properties of a Combined Go/No-Go and Continuous Performance Task Across Childhood

Caron A.C. Clark
University of Nebraska–Lincoln

Kaitlyn Cook
Harvard Medical School, kcook93@smith.edu

Rui Wang
Harvard Medical School

Michael Rueschman
Harvard Medical School

Jerilynn Radcliffe
University of Pennsylvania

See next page for additional authors

Recommended Citation

Authors

Caron A.C. Clark, Kaitlyn Cook, Rui Wang, Michael Rueschman, Jerilynn Radcliffe, Susan Redline, and H. Gerry Taylor

Psychometric Properties of a Combined Go/No-go and Continuous Performance Task across Childhood

Caron A.C. Clark[1], **Kaitlyn Cook**[2], **Rui Wang**[2,3,4], **Michael Rueschman**[3], **Jerilynn Radcliffe**[5,6], **Susan Redline**[3], **H. Gerry Taylor**[7]

[1.]Department of Educational Psychology, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

[2.]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, Massachusetts, USA

[3.]Division of Sleep Medicine and Circadian Disorders, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[4.]Department of Biostatistics, Harvard University T. H. Chan School of Public Health, Boston, Massachusetts, USA

[5.]Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[6.]Center for Human Phenomic Science, Children´s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[7.]Department of Pediatrics, Abigail Wexner Research Institute at Nationwide Children´s Hospital and The Ohio State University, Columbus, Ohio, USA

Abstract

Despite the critical importance of attention for children's self-regulation and mental health, there are few task-based measures of this construct appropriate for use across a wide childhood age range including very young children. Three versions of a combined Go/No-go and Continuous Performance Task (GNG/CPT) were created with varying length and timing parameters to maximize their appropriateness for age groups spanning early to middle childhood. As part of the baseline assessment of a clinical trial, 452 children aged 3 to 12 years (50% male, 50% female; 52% White, Non-Hispanic, 27% Black, 16% Hispanic/Latinx; 6% Other ethnicity/race) completed the task. Confirmatory factor analysis indicated that all task versions assessed two latent factors, labeled Response Inhibition and Sustained Attention. Versions for older children elicited lower overall accuracy while equating levels of inhibitory demand. All versions showed limited floor and ceiling effects, as well as developmental sensitivity. Boys showed higher commission error rates and children from lower income households showed lower performance across multiple task metrics. Task metrics, especially *d*-prime and accuracy summary scores, correlated with parent-reported executive function and externalizing behavior. Task scores show promise as valid and sensitive indicators of inhibition and sustained attention across heterogeneous pediatric age groups.

Corresponding author: Caron A.C. Clark, Rm. 241 Teachers' College Hall, Department of Educational Psychology, University of Nebraska-Lincoln, NE 68521, cclark4@unl.edu.

Keywords

Attention; Executive function; Developmental assessment; Inhibition

The capacity to coordinate and manage attention serves as a foundation for self-regulation and learning (Duncan et al., 2006; Erickson et al., 2015; Fisher, 2019; Stevens & Bavelier, 2012). Attention therefore is listed as a key variable in the National Institute of Mental Health Research Domain Criteria (Insel & Cuthbert, 2010) and is an important clinical endpoint in prevention science. Requirements for lifespan representation and repeated assessments in clinical trials research necessitate measures that are sensitive, valid and reliable across wide age and ability ranges. Given its slow developmental trajectory, performance on tasks designed to assess intrinsically regulated, self-directed attention may vary dramatically with age, making this construct especially difficult to measure consistently across different age groups (Kanaka et al., 2008; Mahone & Schneider, 2012). Go/No-go (GNG) and Continuous Performance Tests (CPT) are used widely to assess two key components of attention: sustained attention, the capacity to focus on a stimulus for prolonged intervals of time; and response inhibition, the capacity to inhibit or cancel an inappropriate behavioral response (Colombo & Cheatham, 2006; Wright et al., 2014). The aim of this study was to evaluate the measurement invariance, sensitivity, and validity of scores from a combined GNG and CPT task for age groups spanning 3 to 12 years.

The development of sustained attention and response inhibition

Sustained attention and response inhibition represent slowly maturing, endogenous forms of attention that involve the individual's effortful deployment of cognitive resources to resist task interference, as opposed to earlier-emerging, exogenous forms of attention that are driven reactively by the salience of stimuli in the environment (Ristic & Enns, 2015). Children show gradual improvements in sustained attention from age 9 months through middle childhood (Amso & Scerif, 2017; Kanaka et al., 2008; Kannass et al., 2006; Oakes et al., 2011; Ruff et al., 1990). Although the capacity to withhold or inhibit inappropriate responses also emerges in infancy, children show a sharp increase in response inhibition performance between 3 and 5 years of age, with less pronounced growth continuing through middle childhood and adolescence (Clark et al., 2012; Lewis, 2017; Wiebe et al., 2012). These forms of attention are closely interrelated, as infant and toddler sustained attention levels predict response inhibition performance in preschoolers (Brandes-Aitken et al., 2019; Reck & Hund, 2011; Veer et al., 2017). Growth in these forms of attention across childhood likely relates to increasing connectivity within, and differentiation of, associated neural networks, including the dorsal attention network, involved in focused attention; the ventral attention network, involved in orienting to sensory cues; and the fronto-parietal network, involved in maintaining goal-relevant information in working memory (Baum et al., 2017; Xie et al., 2019).

Challenges in measuring sustained attention and response inhibition

Given the prolonged developmental trajectories for sustained attention and response inhibition, similar measures may vary in their sensitivity to these constructs in different

age groups. In infants aged 3 to 6 months, lengthier periods of sustained focus on an object may reflect slower processing speed and are associated negatively with later measures of cognitive development (Colombo, 2001; Cuevas & Bell, 2014). Conversely, longer focal times after age 9 months correlate positively with later cognitive performance, suggesting that the same dependent variable may reflect poorer or more advanced attention contingent on the child's age. Likewise, a latent response inhibition construct is difficult to differentiate from related constructs, such as working memory, in confirmatory factor analyses (CFA) of tasks administered to preschool-aged children (Wiebe et al., 2011; Willoughby et al., 2012). Separate latent factors for inhibition and working memory can, however, be parsed in adolescent and adult cohorts (Karr et al., 2018; Lee et al., 2013). These non-invariant measurement properties pose challenges for research focusing on broad age bands, as it is not clear that tasks are measuring the same underlying constructs in different age groups.

When the same measures are administered to different age groups, it also often is necessary to employ different scaling methods or dependent variables to compensate for age differences in children's capacity to perform tasks. For instance, studies with adolescents and adults typically incorporate reaction time as the key measure of attention task performance; accuracy may be less sensitive because it often reaches ceiling levels in these age groups (Allan & Lonigan, 2015; Davidson et al., 2006). In younger children, however, it is typical to use accuracy, which may be appropriate only within constrained age bands. Indeed, dramatic gains in response inhibition over the preschool period mean that 3-year olds often show floor effects for accuracy, whereas 5-year olds perform at ceiling (Carlson, 2005; Clark et al., 2012). Commercially available measures of sustained attention and response inhibition, such as the Conner's CPT (Conners et al., 2003) and the Test of Variables of Attention (Greenberg & Waldman, 1993), include separate tasks appropriate for preschool and school-aged children, although it is not clear that scores from these tasks are comparable. Moreover, these tasks assess children aged 4 and over and therefore overlook a dramatic period of growth in response inhibition between age 3 and 4 years (Wiebe et al., 2012).

Go/No-Go and Continuous Performance Tasks as measures of endogenous attention

GNG and CPT tasks are widely used in research and clinical settings and require participants to make a button press response to a target stimulus or set of stimuli while withholding a response to non-target stimuli (Albrecht et al., 2015; Riccio et al., 2002). The ratio of target ('go') to non-target ('no-go') trials is an important theoretical determinant of the sensitivity of the tasks to distinct aspects of attention (Wright et al., 2014). GNG tasks typically include a higher proportion of target 'go' relative to non-target, 'no-go' trials, creating a prepotent tendency to respond on every trial and placing greater demands on response inhibition. The opposite typically is true of CPT tasks, where participants must maintain a high level of sustained attention in order to recognize the infrequent need for a response (although these labels are sometimes used interchangeably, with the Conner's CPT (Conners et al., 2003) having a GNG-like format). The timing of stimulus presentation also is integral to task sensitivity, as slower trials presumably dampen the prepotency of

the response and generally result in fewer erroneous responses to non-target stimuli (Metin et al., 2012). Young children, however, may require a long time window even to enact a response, meaning that tasks allowing a response time of less than one second may not be sensitive to differences in response inhibition in children below 4 years of age (Simpson & Riggs, 2006). The CPT originally was designed to capture lapses in attention over a prolonged test period (Albrecht et al., 2015). Young children, however, may be less able to tolerate very long test periods and are likely to show attentional lapses earlier than older participants. Measurement quality for these tasks therefore intersects with typical developmental expectations, making it especially important to consider developmental differences in studies with heterogenous age groups.

Another psychometric issue concerns the sensitivity of multiple metrics that can be derived from GNG and CPT tasks. Errors of commission or 'false alarms' to non-target stimuli theoretically capture impulsive, disinhibited or perseverative tendencies (Anderson et al., 2006; Lewis et al., 2017) and typically are higher in boys than in girls (Hasson & Fine, 2012; Wiebe et al., 2012). Conversely, errors of omission or target 'misses' may reflect inattention and tend to increase with task duration (Allan & Lonigan, 2015). Response times to individual trials may reflect the individual's processing speed and several researchers have argued that the fluctuations in participants' trial response times over the course of the task may be a sensitive marker of sustained attention (Antonini et al., 2013). Indeed, variance in reaction times is consistently higher among children with ADHD (Epstein et al., 2012). Measures of correct 'hits' of the target stimuli and total accuracy may provide more holistic performance indicators (Kanaka et al., 2008). Additionally, some researchers have advocated for the use of a d prime (d') signal detection score reflecting the probability of responding to target vs. non-target stimuli, as this index accounts for the fact that participants can demonstrate high overall accuracy despite numerous commission errors if the frequency of target trials is high (Epstein et al., 2003; Lin et al., 1999).

With respect to convergent validity, studies have reported significant differences in the mean performance of children with ADHD and typically developing samples on GNG and CPT tasks (Epstein et al., 2003; Vaughn et al., 2011), although the diagnostic sensitivity and specificity of the tasks relative to clinical interviews typically is below 70% (Albrecht et al., 2015; Edwards et al., 2007). Barnard et al. (2018) found that measures of omission, commission, and reaction time derived from a GNG paradigm for preschoolers did not correlate with parent-reported measures of executive control and attention. Other studies have shown similarly non-significant or small ($r \sim .1$ to .25) correlations with teacher or parent ratings of inattention and hyperactivity (Allan & Lonigan, 2015; Edwards et al., 2007; Vaughn et al., 2011). Overall, then, correlations between these task metrics and clinical criteria or observer rated behavior are small and it is not clear which task metrics provide the most parsimonious and sensitive outcome measures. The sensitivity of different dependent variables is especially critical for clinical trials, which hinge on a small number of pre-specified endpoints to evaluate treatment effects.

There is a need for developmentally appropriate, valid, and sensitive sustained attention and response inhibition measures that can be used across a wide age range. The overarching goal in this study was to determine whether a combined GNG/CPT task could yield valid

and sensitive measures of sustained attention and response inhibition across childhood, including in children as young as 3 years, for whom few measures are available. The task is designed to be appealing and meaningful even for very young children by characterizing it as a 'fishing game' (Wiebe et al., 2012). While the visual properties and response demands of the GNG/CPT task remained constant across age groups, we created age specific task versions that varied in their trial timing and length. Specific aims were to describe the psychometric properties of task scores; to establish whether scores from the different task versions reflected similar latent constructs; to determine the relation of demographic characteristics to children's performance; and to examine convergent the validity of task metrics with parent ratings of children's behavior. We hypothesized that the GNG and CPT task phases would assess distinct latent capacities for sustained attention and response inhibition in all age groups. We hypothesized that adjustments to task properties across different versions would serve to normalize performance and mitigate floor and ceiling effects, while affording sensitivity to individual differences in children's sociodemographic backgrounds and gender. In keeping with previous studies (Allan & Lonigan, 2015; Conners et al., 2003), we hypothesized that task metrics would show small correlations with parent ratings of children's behavior.

Method

Participants

The initial sample comprised 459 3- to 12.9-year-old children enrolled across six sites in a clinical trial (#02562040) designed to assess the effects of adenotonsillectomy on neurobehavioral functioning, sleep and other health-related outcomes in children with frequent snoring but without evidence of significant sleep apnea (apnea hypopnea index <= 3 and no significant oxygen desaturation during sleep; see Wang et al., 2020 for details). All protocols were approved by an Institutional Review Board and parents provided written consent to participate. Children were screened to ensure they did not have chronic health conditions (e.g., cardiopulmonary disorders, epilepsy), psychiatric disorders other than ADHD that required medication or therapy, genetic conditions, autism spectrum disorders, or developmental delay, and that the primary household language was English or Spanish. Children with ADHD were not excluded, as attention was an important outcome of interest in the study.

Seven children had missing data for the GNG/CPT task of interest in this study due to misunderstanding ($n = 1$), refusal to complete ($n = 4$), or technical issues that disrupted administration of ($n = 2$) the task. Therefore, the final analytic sample included 452 children with a mean (*SD*, range) age of 6.63 (2.28; 3.08 – 12.83) years; 49.8% of the children were female and 50.2% were male; and the ethnic breakdown of the sample was 51.5% White, Non-Hispanic; 26.3% Black/African American, Non-Hispanic; 16.4% Hispanic/Latinx; and 5.8% Other. Nineteen percent of mothers had not completed high school or had a maximum education of a high school diploma and 54.7% of the sample reported an annual household income <= $60,000.

Families attended a baseline study assessment at a pediatric research center, where children completed a health exam, a pegboard task to assess dexterity, and then the GNG/CPT

detailed below. Tasks were administered by research assistants trained by board-certified psychologists and typically lasted 15 minutes. Following this baseline assessment, children were assigned randomly to either surgery or watchful waiting with supportive care and then reassessed at 6– and 12– month follow-up points, although only baseline data is included here.

Measures

GNG/CPT task.

Wiebe et al.'s (2012) GNG task for preschoolers was adapted for this study. Children completed one of three separate task versions designed for those 3 to 4 years of age inclusive (Version 1), 5 through 6 years inclusive (Version 2), or age 7 and above (Version 3). The task was administered on a desktop PC using Eprime 2 (Psychology Software Tools, PA). Stimuli consisted of 10 different colored cartoon fish and 10 grey-colored sharks that appeared in the center of the screen. Children were informed that that they should 'catch' the fish by pressing a response pad as quickly as possible. The examiner demonstrated how to respond before allowing the child 3 practice trials to familiarize themself with the task speed. If children did not respond correctly on at least 2 of the 3 practice trials, they were given a maximum of two additional sets of 3 trials to reach this criterion before proceeding. Children were then introduced to a picture of several sharks and were instructed that they should withhold responses to the sharks because catching a shark would break their fishing net. Again, the examiner demonstrated how to correctly withhold a response before allowing the child up to 9 practice trials. Children's ability to distinguish the fish from the sharks was established by having them select the fish stimuli from an array of fish and shark pictures prior to proceeding to the test trials. Each button-press response was followed by a brief feedback stimulus. After a correct response, feedback included a bubbling noise and a picture of the fish in a net. After a commission error, i.e., responding to a non-target shark, feedback comprised a 'buzzer' sound and a picture of a broken net. During the first, 'GNG' phase of the task, 25% of stimuli were sharks. Halfway through administration, the task was paused and children were instructed that they would see many more sharks in subsequent trials. During the final, 'CPT' phase of the task, 75% of the stimuli were sharks, necessitating fewer responses. The task lasted about 8 minutes. Copies of the task are available from the first author for research purposes upon reasonable request.

Importantly, we manipulated several timing features for the different task versions to maximize their developmental appropriateness and minimize floor and ceiling effects (see Table 1). Version-specific response and inter-stimulus intervals were chosen based on attention task-based studies with varying age groups (Conners et al., 2003; Davidson et al., 2006; Wiebe et al., 2012). Specifically, Wiebe et al. (2012) and Orchinik et al. (2011) reported ceiling levels of performance in typically developing children by age 5 years using Version 1 parameters. We therefore reduced the stimulus display time and inter-stimulus interval for children aged 5 to 6 years. Studies indicate a slower rate of change in response inhibition task performance after age 7 (Brocki & Bohlin, 2004; Greenberg & Waldman, 1993). Therefore, the final, third version of the task used the same parameters for all children aged 7 years and over. For older children, we also increased trial numbers to

maximize sensitivity to sustained attention. To preserve target to non-target ratios in versions with varying trial numbers, trial sequences necessarily differed by version. For the purposes of evaluating measurement invariance and change in performance across the course of the task, each task phase, i.e., GNG and CPT, was divided into 4 consecutive subblocks and trial numbers within these subblock also varied by version (Table 1).

Dependent variables for each phase included the overall proportion of accurate trials; commission errors; omission errors; mean response time for correct trials in seconds (RT); RT variability, the standard deviation of reaction times to correct trials; and d', computed as the Z-scored proportion of incorrect commissions subtracted from the Z-scored proportion of correct responses to the target fish (Stanislaw & Todorov, 1999). Omission and commission rates of 0 or 100% were adjusted slightly to allow for the d' calculation (Stanislaw & Todorov, 1999).

Parent behavior ratings.

Parent reports of child behavior were obtained using age-appropriate versions of the Child Behavior Checklist (CBCL, Achenbach & Rescorla, 2000) and Behavior Rating Inventory of Executive Function (BRIEF, Gioia et al., 2003, 2015). Ratings of the broader spectrum of behavior problems were obtained using the CBCL/1.5–5 for children ages 3 – 5 years and the CBCL/6–18 for children aged 7 years and older. The scales include 99 behavior problems rated on a Likert scale of 0 = absent, 1 = occurs sometimes, and 2 = occurs often. Despite some differences in subscale content, both versions have scales for Internalizing and Externalizing behaviors, and Attention Problems, which were of interest in the current study given their mapping to hypothesized GNG/CPT task constructs. Standardized scores are based on child age and sex and research has confirmed high levels of reliability, as well as validity in relation to other behavior ratings (Achenbach & Rescorla, 2001). Internal consistencies for CBCL Attention Problems, Internalizing, and Externalizing scales are .83, .83, and .88, respectively.

Parent ratings of behavior problems suggestive of difficulties in executive functioning were obtained using the Behavior Rating Inventory of Executive Function – Preschool version (BRIEF-P, Gioia et al., 2003) for children 3 and 4 years of age and 5-year-olds not yet in kindergarten, and the BRIEF, Second Edition (BRIEF – 2, Gioia et al., 2015) for 5-year-olds attending kindergarten and children ages 6 years and older. The two BRIEF versions assess similar dimensions of behavior and include the Global Executive Composite (GEC) as a summary index. Each version consists of 63 items rated on Likert scale ranging from 0 = Never to 3 = Often. Normative data are based on large, nationally representative samples, with research confirming high levels of reliability as well as concurrent validity with other behavior ratings and measures of cognition. In this sample, Cronbach's α for the full scale was .97.

Analytic Approach

ANOVAs were used to evaluate the impact of task version and phase on various GNG/CPT task metrics. To test the factor structure of the task in the full sample, we treated the mean accuracy for each of eight task subblocks as items for analysis and tested, using

CFA, whether the underlying construct/s driving performance were best represented as one or two factors. Maximum likelihood estimation was used and model fit was compared using the Chi-squared difference test, RMSEA, CFI, TLI, and SRMR (Kline, 2011). The invariance of the GNG/CPT age versions was then tested by constraining the factor loadings (metric invariance), intercepts (scalar invariance) and item residuals (strict invariance) to be equal across versions and testing whether each constraint led to a statistically significant decline in model fit (Putnick & Bornstein, 2016). Specifically, if the chi-squared difference test comparing the constrained to the unconstrained model was significant, we rejected the assumption of measurement invariance. Where strict invariance was rejected, we progressively constrained individual parameters to test for partial invariance. Associations of GNG/CPT performance metrics with demographic characteristics were examined using multiple regression analyses, with each task version coded as a dummy variable. Given that 35 children had missing values for household income, we performed multiple imputation before conducting regression analyses with demographic predictor variables. We created 10 imputed datasets and combined point estimates and variances across each of these datasets using the standard Rubin (2004) method. Missing data for other variables was negligible ($<$ 2%) and reviews of reasons for missingness and descriptive analyses did not show evidence of bias in this missingness. Analyses were performed using SPSS 28, R 4.1.0, and MPLUS 8 with a significance threshold of α = .05. All data will be made available in the National Sleep Research Resource repository (https://sleepdata.org/datasets) at the study conclusion. Analytic code is available from the first author upon reasonable request. This study was not pre-registered.

Results

General descriptive indicators of performance

Table 2 describes children's mean performance on several GNG/CPT task metrics. Apart from age, there were no significant differences in the socio-demographic background characteristics of children completing different versions (see Supplementary Materials). Preliminary analyses also revealed no significant effect of study site on any of these metrics. Therefore, data were combined across sites. Generally, outcome metrics showed little evidence of skewness or kurtosis and there was a high degree of performance variability within each task version (see Supplementary Materials). As an exception, commission errors during the CPT phase were skewed (2.93) and kurtotic (9.47), as few children made such errors.

Factor structure of the GNG/CPT task

We evaluated the factor structure for the GNG/CPT using children's mean accuracy for each subblock as items for analysis (see Supplementary Materials for correlations and mean performance by version and sub-block). A model with all items loaded on the same factor showed a poor fit to the data, $\chi^2(24) = 343.71$, $p < .001$; RMSEA = .19, CFI = .82, TLI = .75, SRMR = .08 and this continued to be the case when the residuals for consecutive subblocks were allowed to covary, $\chi^2(14) = 102.48$, $p < .001$; RMSEA = .12, CFI = .95, TLI = .9, SMR = .05. A model that defined two correlated factors parsing items from the GNG (labeled Response Inhibition) and CPT (labeled Sustained Attention) phases showed

improved fit, $\chi^2(19) = 93.38$, $p < .001$; RMSEA = .09, CFI = .96, TLI = .94, SRMR = .04 and fit was excellent when the residuals for successive subblocks were allowed to covary in this 2-factor model, $\chi^2(13) = 13.97$, $p = .37$; RMSEA = .01, CFI = .99, TLI = .99, SRMR = 0 (see Figure 1). McDonald's ω reliability coefficients were .83 for the Response Inhibition and .85 for the Sustained Attention factor.

Measurement invariance for different versions of the GNG/CPT task

A key study aim was to determine whether changing the timing and duration of the same GNG/CPT task for different age groups would help to normalize age differences while preserving construct validity. As shown in Table 2, varying metrics from the different task versions suggested that children's performance was not equivalent, with performance on the versions designed for older children, i.e., versions 2 and 3, generally being poorer than in the version developed for preschool-aged children. In particular, children showed higher omission error rates for the third, age 7+ version of the task.

Table 3 shows the results of CFA models to evaluate the measurement invariance of the different task versions more formally. The two-factor model with autocorrelated residuals showed excellent fit across all task versions and configural invariance was supported. Notably, however, the strength of the correlation between the factors varied by version, being strongest for Version 2 ($\beta = .93$) and weakest for Version 3 ($\beta = .79$). Tests for metric invariance showed that the loading for item 4 on the CPT factor was significantly lower in Version 1, although it was similar for the latter two versions. The remaining loadings could be set equivalent for all versions without a significant reduction in model fit. Scalar invariance also held for most intercepts, although the intercept for GNG subblock 3 was significantly higher in Version 2. Finally, there was limited strict invariance, especially for the Sustained Attention factor, where residual variances were lower for task Version 1 (see Figure 1). McDonald's ω coefficients for Response Inhibition and Sustained Attention were high for task Versions 1 (.91, .83) and 3 (.84, .86), although lower for Version 2 (.67, .64).

Relations among GNG/CPT performance metrics

Table 4 describes the correlations among various dependent variables derived from each task phase, controlling for task version and child age. Accuracy and d' metrics correlated robustly with commission and omission error rates within and across task phases. Mean response times were weakly correlated with accuracy and d', and moderately correlated with omission and commission errors. Higher response time variability was associated with lower accuracy. Overall, the d' scores tended to correlate most consistently with other performance metrics.

Relation of GNG/CPT task performance to demographic factors

A series of linear regression models showed that children's age was related to all GNG/CPT task metrics independent of the effects of task version (p's < .05; see Supplementary Materials). Indeed, after accounting for task version, age explained an additional 13 to 28% of the variance in children's accuracy, d' and omission errors, whereas version effects explained 14 – 38% of the variance in these metrics in the same models. There were also interactions between version and age for all task metrics apart from for omission errors and

response times (partial R^2's = .02 – .11 p's < .05). Figure 2 illustrates this pattern more closely, showing that, within each version of the task, higher age was associated with better performance, and that age effects tended to be strongest for Version 1, designed for the youngest age group.

As shown in Table 5, after adjusting for task version, age, and other demographic characteristics, household income was the only demographic variable consistently related to task performance. Children from households with income levels < \$60,000 generally showed lower accuracy and d' scores, higher commission rates, and higher reaction time variability for both task phases, partial R^2 = 11 – 14%. In addition, boys showed higher commission rates, coupled with faster reaction times, during the GNG phase only, partial R^2 = 13 – 19%.

Relations of GNG/CPT task metrics to parent-reported behavior

Table 6 describes the correlations GNG/CPT task metrics and parent-reported behavior on the CBCL and BRIEF after controlling for task version and child age. Accuracy and d' scores from both phases correlated with the Global Executive Composite, Inhibit, Working Memory, and Plan/Organize scales from the BRIEF, such that children with higher scores were rated as having fewer difficulties in these areas. Omission errors during the CPT phase corresponded with higher BRIEF Shift and Working Memory scores. The d' and commission metrics from the GNG task phase correlated with Externalizing symptoms on the CBCL, as did the reaction time variability metrics from both task phases. There were no correlations between GNG/CPT scores and Attention Problems or Internalizing Problems rated on the CBCL. Similar analyses of selective parent-report scales recoded as binary outcome variables to compare children with high (T score > 65) and average scores showed that the d' metrics for each phase were negatively associated with high (more problematic) BRIEF Global Executive Composite scores, whereas commission errors during the GNG phase and higher RT variability during the CPT phase were positively associated with high CBCL Externalizing and Attention Problem ratings (see Supplementary Materials). Correlations with behavior generally were strongest for task Version 1 (see Supplementary Materials).

Discussion

Comparable assessment of children with widely varying ages is challenging, particularly for cognitive constructs that change dramatically with age, such as sustained attention and response inhibition. We manipulated the temporal demands of a child friendly GNG/CPT task (Wiebe et al., 2012) so that older children completed more trials at a faster event rate. Findings suggest that the different task versions assess similar latent constructs of Response Inhibition and Sustained Attention in children spanning as young as 3 to 12 years of age. Task metrics generally showed limited floor or ceiling effects even in the youngest children. The task is sensitive to developmental, gender and socioeconomic differences and scores correlate with parent-reported executive function and externalizing behavior. Although key metrics will require statistical adjustment for version difficulty, findings generally support the suitability of this task for varying age groups.

As hypothesized, the GNG and CPT task phases were best modeled as separate latent factors, which we deem 'Response Inhibition' and 'Sustained Attention'. Although the separation of these latent skills generally is aligned with the conceptualization of GNG and CPT tasks in the literature (Riccio et al., 2002; Wright et al., 2014), it should be noted that the correlation between the factors was high ($\beta = .72$), indicating substantial overlap in constructs assessed by these phases. The two-factor model showed excellent fit for all task versions, although the correlation between the factors varied, being weakest in the version designed for children aged 7+ years. The weaker correlation between the factors in the oldest children perhaps reflects a tighter concordance between these attentional capacities in early childhood or higher variation in performance associated with a broader age range for children completing Version 3. We should also note that, although items showed good internal consistency for the task as a whole, reliability coefficients for Version 2 were lower, perhaps due to slightly lower consistency in children's performance across subblocks for this version.

Further tests for metric, scalar, and strict invariance revealed some differences in measurement parameters across task versions. Versions generally were metric-invariant, with the exception that the factor loading for the final CPT subblock was lower for Version 1. There also was a difference in the intercepts for some subblocks of task Version 2, designed for children aged 5 to 7 years. Children performing this version of the task performed especially well in the final CPT block, perhaps because there was one fewer no-go trial in this block than in other blocks. Descriptive analyses also indicated that versions designed for older children elicited more omission errors than Version 1 and that accuracy was lower for these versions. Interestingly, this occurred within the context of relatively uniform commission error rates across versions during the GNG phase. That is, task adjustments may have equated inhibitory demands across versions while simultaneously increasing the tendency of older children to miss target trials. Differences in intercepts indicate that scores will need to adjusted to standardize these metrics across versions. It is reassuring, however, that all age versions appear to assess the same underlying constructs, as this offers opportunity for valid assessment of Sustained Attention and Inhibitory Control across heterogeneous age groups.

Within the different task versions, older children showed higher performance than younger children. Indeed, there were pronounced differences in performance with age in the version of task developed for the oldest children, suggesting that the task is sensitive to developmental differences even as children approach adolescence. Adjusting scores for version, the task may be helpful in examining developmental changes in sustained attention and inhibitory control across a broad age range.

After accounting for version effects, age, and gender, children from households with annual incomes of $60,000 or below showed poorer task performance, consistent with multiple studies describing poorer endogenous attention for children living in financially impoverished households (Allee-Herndon & Roberts, 2019; Brandes-Aitken et al., 2019). Poor sustained attention and inhibitory control in children from lower income households may be offset by relative strengths in other cognitive capacities (Ellis et al., 2022). However, the fact that these attentional capacities are strongly predictive of classroom behavior

and academic achievement (Hernández et al., 2017; Nelson et al., 2017) underscores the importance of policy and prevention strategies to address these income-related inequities. The fact that the task is sensitive to these sociodemographic factors indicates its potential sensitivity to interventions that target children's social environments. Similarly, there were gender differences in children's commission error rates during the GNG phase of the task, which may reflect a speed accuracy trade-off for girls, who showed slower response times. These gender differences are consistent with previous studies using similar tasks (Hasson & Fine, 2012).

Correlations between the GNG/CPT metrics and parent-reported behavior generally were small, consistent with previous studies (Allan & Lonigan, 2015; Edwards et al., 2007). Lower d' and accuracy scores on both task phases were linked to overall executive scores, and especially to working memory and planning, whereas RT variability correlated with externalizing behavior. Patterns of correlation hint that commission errors may be relatively more sensitive for the GNG phase, whereas omission errors are more sensitive for the CPT phase. Overall, however, these correlations suggest limited specificity of task metrics for isolating specific attention or inhibition behavior difficulties in a non-clinical sample and indicate that task metrics may instead offer a broad indication of children's general executive function. Modest correlations may also reflect the potential sensitivity of the task to aspects of attention not reflected in observer ratings that may serve as surrogates for biologically-based variations in attentional capacities. Although this remains to be seen, the correlations offer support for the convergent validity of this task in detecting individual differences in children's everyday behavior. Given that the task can be administered even to very young children, with further validation, it may represent a promising screening tool for early identification and follow-up of children in need of further clinical assessment and support for their self-management of attention.

Limitations and constraints on generality

Several limitations of the study should be noted. First, task measures of RT and RT variability cannot be compared across task versions, as response times were more variable in the youngest children, who were allocated more time to respond. Second, analyses of factorial invariance were based on average subblock accuracy. Higher trial numbers in versions designed for older children may mean that subblock averages from those versions are more reliable. Third, children who were administered Version 1 spanned only the 3 to 4 year age range, whereas Version 3 was administered to a children ranging from 7 to 12 years of age. This variation in age ranges for the task versions may also have impacted levels of invariance. A further limitation was that the order of the GNG and CPT blocks was not counterbalanced. While a standard phase order is critical to the internal validity of the broader clinical study, children may have been more fatigued during the second, CPT block, perhaps leading to performance changes across the blocks that are unrelated to the task's psychometric properties. Finally, it is important to note that the study included a primarily English-speaking, North American sample selected specifically for a clinical study of mild sleep disordered breathing. Participants represented East Coast, mid-west, and Texas regions in the USA. While the sample was racially, ethnically, and socioeconomically

diverse, further testing of the task across other regions and in neurodiverse samples will be important in establishing measurement validity.

Conclusions and directions for future research

Several future directions may help to maximize the validity, sensitivity and reliability of this task. First, it may be helpful in future studies to compare the same children's performance across task versions and to assess test-re-test reliability to determine whether relative performance remains consistent. Given the lower accuracy and d' scores for the oldest group, it also is possible that we over-adjusted presentation times in the final version of the task or that the timing parameters for the 5- to 6-year version may have been more appropriate for 7-year-old children. A major advantage of the task is its relative brevity; it would be especially helpful to compare metrics to those from other child attention tasks, which typically are longer in duration, to determine whether the task has comparable levels of sensitivity to children's behavior.

Despite these limitations, the study suggests that this GNG/CPT task with timing parameters titrated for varying age groups allows for the consistent assessment of critical aspects of attention over a wide age range. The measure fulfils a need for tasks that capture these attention skills comparably in children as young as 3 and as late as 12 years with limited floor and ceiling effects. Relations with demographic characteristics and parent-reported behavior are similar in magnitude to those reported in previous literature on similar tasks. Based on their robust correlations with other task metrics, their distributional properties, and their correlations with parent-reported executive function, the d' prime scores may provide the most robust general indicators of children's sustained attention and response inhibition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Achenbach T, & Rescorla L (2000). Manual for the ASEBA preschool forms and profiles. University of Vermont.

Albrecht B, Uebel-von Sandersleben H, Wiedmann K, & Rothenberger A (2015). ADHD History of the Concept: The Case of the Continuous Performance Test. Current Developmental Disorders Reports, 2(1), 10–22. 10.1007/s40474-014-0035-1

Allan DM, & Lonigan CJ (2015). Relations between response trajectories on the continuous performance test and teacher-rated problem behaviors in preschoolers. Psychological Assessment, 27(2), 678–688. 10.1037/pas0000054 [PubMed: 25419645]

Allee-Hemdon KA, & Roberts SK (2019). Poverty, Self-Regulation and Executive Function, and Learning in K-2 Classrooms: A Systematic Literature Review of Current Empirical Research. Journal of Research in Childhood Education, 33(3), 345–362. 10.1080/02568543.2019.1613273

Amso D, & Scerif G (2017). The attentive brain: Insights from developmental cognitive neuroscience. Physiology & Behavior, 176(3), 139–148. 10.1038/nrn4025.The [PubMed: 28363838]

Anderson V, Anderson D, & Anderson P (2006). Comparing attentional skills in children with acquired and developmental central nervous system disorders. Journal of the International Neuropsychological Society: JINS, 12(4), 519–531. 10.1017/S135561770606067X [PubMed: 16981604]

Antonini TN, Narad ME, Langberg JM, & Epstein JN (2013). Behavioral correlates of reaction time variability in children with and without ADHD. Neuropsychology, 27(2), 201–209. 10.1037/a003207 [PubMed: 23527648]

Barnard H, Rao R, Xu Y, Froehlich T, Epstein J, Lanphear BP, & Yolton K (2018). Association of the Conners' Kiddie Continuous Performance Test (K-CPT) Performance and Parent-Report Measures of Behavior and Executive Functioning. Journal of Attention Disorders, 22(11), 1056–1065. 10.1177/1087054715578271 [PubMed: 25846228]

Baum GL, Ciric R, Roalf DR, Betzel RF, Moore TM, Shinohara RT, Kahn AE, Vandekar SN, Rupert PE, Quarmley M, Cook PA, Elliott MA, Ruparel K, Gur RE, Gur RC, Bassett DS, & Satterthwaite TD (2017). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth. Current Biology, 27(11), 1561–1572.e8. 10.1016/j.cub.2017.04.051 [PubMed: 28552358]

Brandes-Aitken A, Braren S, Swingler M, Voegtline K, & Blair C (2019). Sustained attention in infancy: A foundation for the development of multiple aspects of self-regulation for children in poverty. Journal of Experimental Child Psychology, 184, 192–209. 10.1016/j.jecp.2019.04.006 [PubMed: 31039447]

Brocki KC, & Bohlin G (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. Developmental Neuropsychology, 26(2), 571–593. 10.1207/s15326942dn2602_3 [PubMed: 15456685]

Carlson SM (2005). Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology, 28, 595–616. 10.1207/s15326942dn2802 [PubMed: 16144429]

Clark CAC, Sheffield TD, Chevalier N, Nelson JM, Wiebe SA, & Espy KA (2012). Charting Early Trajectories of Executive Control With the Shape School. Developmental Psychology, 49(8), 1481–1493. 10.1037/a0030578 [PubMed: 23106846]

Colombo J (2001). The development of visual attention in infancy. Annual Review of Psychology, 52(337–367).

Colombo J, & Cheatham CL (2006). The emergence and basis of endogenous attention in infancy and early childhood. In Kail R (Ed.), Advances in Child Development and Behavior (pp. 283–322). Elsevier.

Conners K, Epstein JN, Angold A, & Klaric J (2003). Continuous Performance Test Performance in a Normative Epidemiological Sample. Journal of Abnormal Child Psychology, 31(5), 555–562. [PubMed: 14561062]

Cuevas K, & Bell MA (2014). Infant Attention and Early Childhood Executive Function. Child Development, 85(2), 397–404. 10.1111/cdev.12126 [PubMed: 23711103]

Davidson MC, Amso D, Anderson LC, & Diamond A (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia, 44(11), 2037–2078. 10.1016/j.neuropsychologia.2006.02.006 [PubMed: 16580701]

Duncan GJ, Dowsett C, Claessens A, Magnuson K, Huston A Cl., Klebanov, P., Feinstein, L., Engel, M., Brooks-Gunn, J., Seton, H., Duckworth, K., & Japel, C. (2006). School Readiness and Later Achievement. 43(6), 1–36. 10.1037/0012-1649.43.6.1428

Edwards MC, Gardner ES, Chelonis JJ, Schulz EG, Flake RA, & Diaz PF (2007). Estimates of the validity and utility of the conners' continuous performance test in the assessment of inattentive and/or hyperactive-impulsive behaviors in children. Journal of Abnormal Child Psychology, 35(3), 393–404. 10.1007/s10802-007-9098-3 [PubMed: 17295064]

Ellis BJ, Abrams LS, Masten AS, Sternberg RJ, Tottenham N, & Frankenhuis WE (2022). Hidden talents in harsh environments. Development and Psychopathology, 34(1), 95–113. 10.1017/S0954579420000887 [PubMed: 32672144]

Epstein JN, Erkanli A, Conners CK, Klaric J, Costello EJ, & Angold A (2003). Relations Between Continuous Performance Test Performance Measures and ADHD Behaviors. Journal of Abnormal Child Psychology, 31(5), 543–554. 10.1023/A:1025405216339 [PubMed: 14561061]

Epstein JN, Langberg JM, Rosen PJ, Graham A, Narad ME, Antonini TN, Brinkman WB, Froehlich T, & Simon JO (2012). Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. Neuropsychology, 25(4), 427–441. 10.1037/a0022155.Evidence

Erickson LC, Thiessen ED, Godwin KE, Dickerson JP, & Fisher AV (2015). Endogenously and exogenously driven selective sustained attention: Contributions to learning in kindergarten children. Journal of Experimental Child Psychology, 138, 126–134. 10.1016/j.jecp.2015.04.011 [PubMed: 26044539]

Fisher AV (2019). Selective Sustained Attention: A Developmental Foundation for Cognition. Elsevier, https://www.elsevier.com/open-access/userlicense/1.0/

Gioia GA, Espy KA, & Isquith PK (2003). BRIEF-P: Behavior Rating Inventory of Executive Function—Preschool Version. Psychological Assessment Resources.

Gioia GA, Isquith PK, & Guy S (2015). Behavior Rating Inventory of Executive Function—2. PAR.

Greenberg LM, & Waldman ID (1993). Developmental and normative data on the Test of Variables of Attention (TOVA). Journal of Child Psychology and Psychiatry, 34(6), 1019–1030. 10.1111/j.1469-7610.1993.tb01105.x [PubMed: 8408366]

Hasson R, & Fine JG (2012). Gender Differences Among Children With ADHD on Continuous Performance Tests: A Meta-Analytic Review. Journal of Attention Disorders, 16(3), 190–198. 10.1177/1087054711427398 [PubMed: 22290696]

Hernández MM, Valiente C, Eisenberg N, Berger RH, Spinrad TL, VanSchyndel SK, Silva KM, Southworth J, & Thompson MS (2017). Elementary students' effortful control and academic achievement: The mediating role of teacher–student relationship quality. Early Childhood Research Quarterly, 40, 98–109. 10.1016/j.ecresq.2016.10.004 [PubMed: 28684888]

Insel T, & Cuthbert B (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. American Journal of Psychiatry, 167, 748–751. 10.1176/appi.ajp.2010.09091379 [PubMed: 20595427]

Kanaka N, Matsuda T, Tomimoto Y, Noda Y, Matsushima E, Matsuura M, & Kojima T (2008). Measurement of development of cognitive and attention functions in children using continuous performance test. Psychiatry and Clinical Neurosciences, 62(2), 135–141. 10.1111/j.1440-1819.2008.01746.x [PubMed: 18412834]

Kannass KN, Oakes LM, & Shaddy DJ (2006). A Longitudinal Investigation of the Development of Attention and Distractibility. Journal of Cognition and Development, 7(3), 381–409. 10.1207/s15327647jcd0703_8

Karr JE, Areshenkoff CN, Rast P, Hofer SM, Iverson GL, & Garcia-Barrera MA (2018). The unity and diversity of executive functions: A systematic review and reanalysis of latent variable studies. Psychological Bulletin, 144(11), 1147–1185. 10.1037/bul0000160 [PubMed: 30080055]

Kline RB (2011). Principles and practice of structural equation modeling. Guildford Press.

Lee K, Bull R, & Ho RMH (2013). Developmental changes in executive functioning. Child Development, 84(6), 1933–1953. 10.1111/cdev.12096 [PubMed: 23550969]

Lewis FC, Reeve RA, Kelly SP, & Johnson KA (2017). Sustained attention to a predictable, unengaging Go/No-Go task shows ongoing development between 6 and 11 years. Attention, Perception, and Psychophysics, 79(6), 1726–1741. 10.3758/s13414-017-1351-4

Lin CCH, Hsiao CK, & Chen WJ (1999). Development of Sustained Attention Assessed Using the Continuous Performance Test among Children 6-15 Years of Age. Journal of Abnormal

Child Psychology, 27(5), 403–412. https://doi.org/0091-0627/99/1000-0403$16.00/0 [PubMed: 10582841]

Mahone EM, & Schneider HE (2012). Assessment of attention in preschoolers. Neuropsychology Review, 22(4), 361–383. 10.1007/s11065-012-9217-y [PubMed: 23090646]

Metin B, Roeyers H, Wiersema JR, Van Der Meere J, & Sonuga-Barke E (2012). A meta-analytic study of event rate effects on Go/No-Go performance in attention-deficit/hyperactivity disorder. Biological Psychiatry, 72(12), 990–996. 10.1016/j.biopsych.2012.08.023 [PubMed: 23062355]

Nelson TD, Nelson JM, James TD, Clark CAC, Kidwell KM, & Espy KA (2017). Executive control goes to school: Implications of preschool executive performance for observed elementary classroom learning engagement. Developmental Psychology, 53(5). 10.1037/dev0000296

Oakes LM, Hurley KB, Ross-Sheehy S, & Luck SJ (2011). Developmental changes in infants' visual short-term memory for location. Cognition, 118(3), 293–305. 10.1016/j.cognition.2010.11.007 [PubMed: 21168832]

Orchinik LJ, Taylor HG, Espy KA, Minich N, Klein N, Sheffield T, & Hack M (2011). Cognitive outcomes for extremely preterm/extremely low birth weight children in kindergarten. Journal of the International Neuropsychological Society, 17(6), 1067–1079. 10.1017/S135561771100107X [PubMed: 21923973]

Putnick DL, & Bornstein MH (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental Review, 41, 71–90. 10.1016/j.dr.2016.06.004 [PubMed: 27942093]

Reck SG, & Hund AM (2011). Sustained attention and age predict inhibitory control during early childhood. Journal of Experimental Child Psychology, 108(3), 504–512. 10.1016/j.jecp.2010.07.010 [PubMed: 20801457]

Riccio CA, Reynolds CR, Lowe P, & Moore JJ (2002). The continuous performance test: A window on the neural substrates for attention? Archives of Clinical Neuropsychology, 17, 235–272. [PubMed: 14589726]

Ristic J, & Enns JT (2015). The Changing Face of Attentional Development. Current Directions in Psychological Science, 24(1), 24–31. 10.1177/0963721414551165

Rubin DA (2004). Multiple imputation for non-response surveys (Vol. 81). John Wiley & Sons.

Ruff HA, Lawson KR, Parrinello R, & Weissberg R (1990). Long-term stability of individual differences in sustained attention in the early years. Child Development, 61(1), 60–75. 10.2307/1131047 [PubMed: 2307047]

Simpson A, & Riggs KJ (2006). Conditions under which children experience inhibitory difficulty with a "button-press" go/no-go task. Journal of Experimental Child Psychology, 94, 18–26. 10.1016/j.jecp.2005.10.003 [PubMed: 16325846]

Stanislaw H, & Todorov N (1999). Calculation of signal detection theory measures. Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc, 31(1), 137–149. 10.3758/BF03207704

Stevens C, & Bavelier D (2012). The role of selective attention on academic foundations: A cognitive neuroscience perspective. Developmental Cognitive Neuroscience, 2(SUPPL. 1). 10.1016/j.dcn.2011.ll.001

Vaughn AJ, Epstein JN, Rausch J, Altaye M, Langberg J, Newcorn JH, Hinshaw SP, Hechtman L, Arnold LE, Swanson JM, & Wigal T (2011). Relation between outcomes on a continuous performance test and adhd symptoms over time. Journal of Abnormal Child Psychology, 39(6), 853–864. 10.1007/s10802-011-9501-y [PubMed: 21476025]

Veer IM, Luyten H, Mulder H, van Tuijl C, & Sleegers PJC (2017). Selective attention relates to the development of executive functions in 2,5- to 3-year-olds: A longitudinal study. Early Childhood Research Quarterly, 41, 84–94. 10.1016/j.ecresq.2017.06.005

Wang R, Bakker JP, Chervin RD, Garetz SL, Hassan F, Ishman SL, Mitchell RB, Morrical MG, Naqvi SK, Radcliffe J, Riggan EI, Rosen CL, Ross K, Rueschman M, Tapia IE, Taylor HG, Zopf DA, & Redline S (2020). Pediatric Adenotonsillectomy Trial for Snoring (PATS): Protocol for a randomised controlled trial to evaluate the effect of adenotonsillectomy in treating mild obstructive sleep-disordered breathing. BMJ Open, 10(3). 10.1136/bmjopen-2019-033889

Wiebe SA, Sheffield TD, & Espy KA (2012). Separating the Fish From the Sharks: A Longitudinal Study of Preschool Response Inhibition. Child Development, 83(4), 1245–1261. 10.1111/j.1467-8624.2012.01765.x [PubMed: 22533453]

Wiebe SA, Sheffield T, Nelson JM, Clark CAC, Chevalier N, & Espy KA (2011). The structure of executive function in 3-year-olds. Journal of Experimental Child Psychology, 108(3). 10.1016/j.jecp.2010.08.008

Willoughby MT, Blair CB, Wirth RJ, & Greenberg M (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. Psychological Assessment, 24(1), 226–239. 10.1037/a0025361 [PubMed: 21966934]

Wright L, Lipszyc J, Dupuis A, Thayapararajah SW, & Schachar R (2014). Response inhibition and psychopathology: A meta-analysis of Go/No-Go task performance. Journal of Abnormal Psychology, 123(2), 429–439. 10.1037/a0036295 [PubMed: 24731074]

Xie W, Mallin BM, & Richards JE (2019). Development of brain functional connectivity and its relation to infant sustained attention in the first year of life. Developmental Science, 22(1). 10.1111/desc.12703

Public significance:

There are very few well-developed measures of children's management of their attention, although attention is fundamental to children's learning and well-being. This study describes and validates a task that can be used even with very young children to measure and monitor their attention skills. The task may be especially useful for clinicians and researchers hoping to understand and support attention in children of widely varying ages.
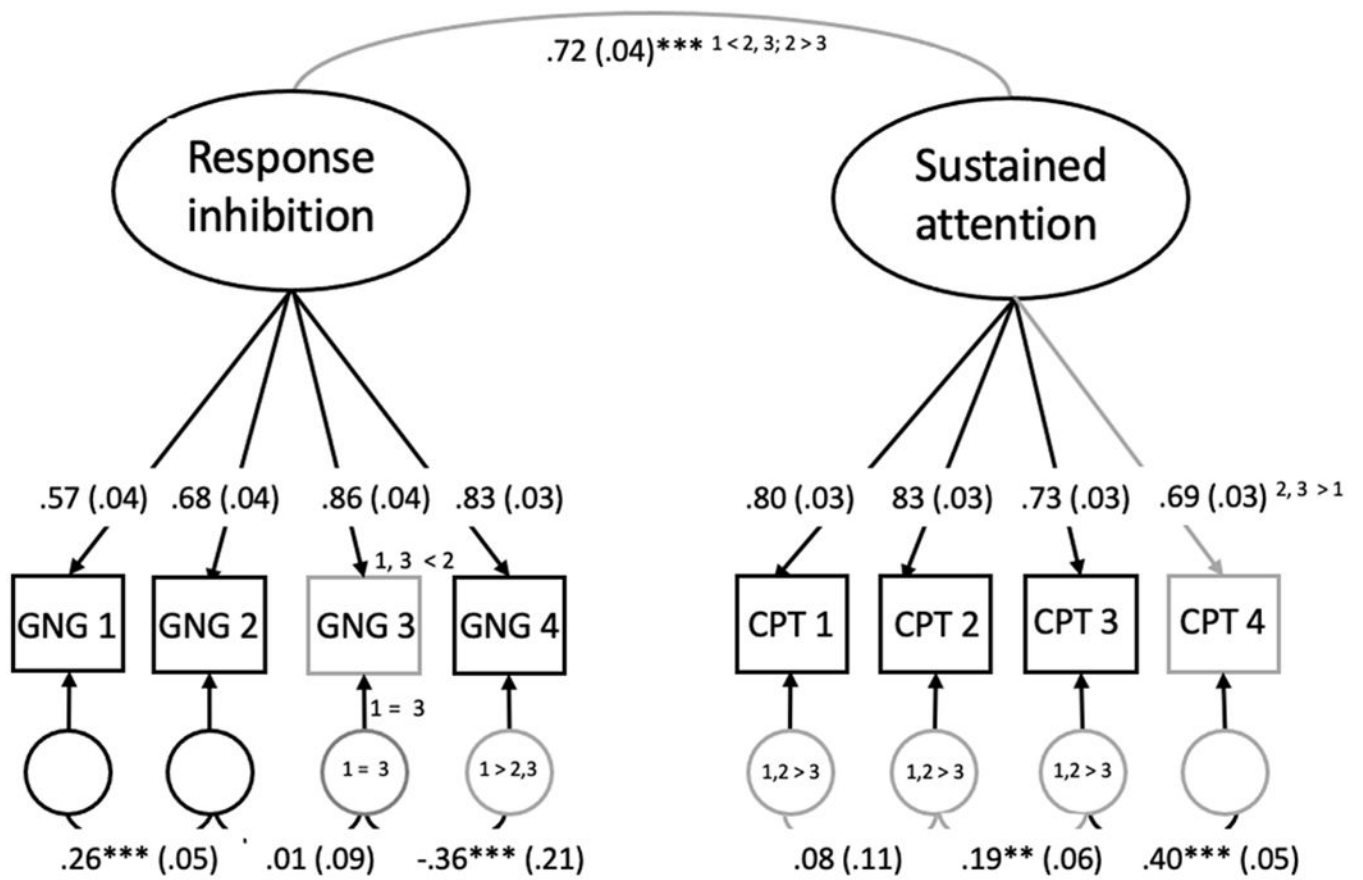
Figure 1:

Factor structure for the GNG/CPT task.

Note: The figure shows standardized parameters. Gray lines indicate parameters that could not be constrained equal across task versions, with significant differences numbered by task version. All factor loadings were significant at $p < .001$.
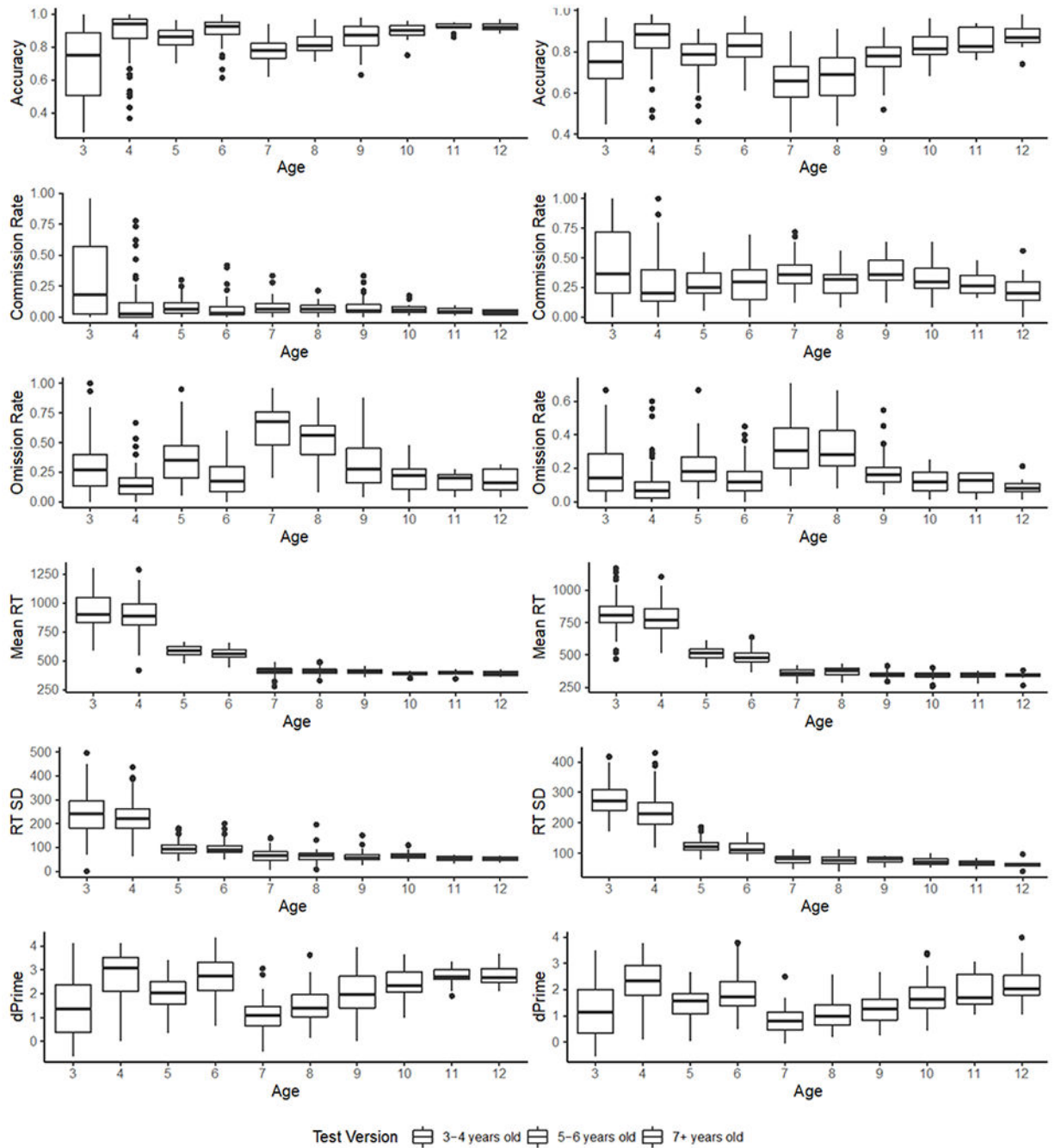
Figure 2:
Distribution of CPT (sustained attention; left) and GNG phase performance measures (inhibitory control; right) by test version and participant age.

Note: Age is shown here as a categorical indicator for illustrative purposes, but was treated as continuous in all analyses

Table 1:

Trial properties for three versions of the GNG/CPT task

	Test version		
	1: 3 - 4 years	2: 5 - 6 years	3: 7 + years
Stimulus display time	1500ms	750ms	500ms
Feedback stimulus time	750ms	750ms	750ms
Inter-trial interval [*]	1000ms	750ms	500ms
Trials per phase	60	80	100
Trials per subblock	15	20	25
# 'Go' trials per subblock – GNG phase	10 - 13	11 - 17	17 - 22
# 'Go' trials per subblock – CPT phase	3 - 4	4 - 6	6 - 7

[*] From the end of the feedback stimulus or stimulus to subsequent stimulus

Table 2:

Children's performance on various metrics from different versions of the GNG/CPT task

	GNG/CPT Task version			
	1. 3 – 4 years **(*n* = 134)**	**2. 5 – 6 years** **(*n* = 135)**	**3. 7+ years** **(*n* = 183)**	***p***
GNG Phase				
M (range) Accuracy	.82 [.45, .98]	.80 [.46, .98]	.73 [.41, .98]	<.001
M (range) *d′*	1.84 [−.53, 3.79]	1.68 [.04, 3.79]	1.24 [−.07, 3.99]	<.001
M (range) Commission errors	.33 [.00, 1.00]	.28 [.00, .70]	.33 [.00, .72]	.05
M (range) Omission errors	.14 [.00, .67]	.17 [.00, .67]	.25 [.01, .71]	<.001
M (range) RT	792.65 [469, 1171]	497.70 [365, 638]	356.57 [256, 433]	<.001
M (range) RT Variability	254.80 [119, 430]	119.64 [75, 186]	76.73 [41, 115]	<.001
CPT Phase				
M (range) Accuracy	.81 [.28, 1.00]	.87 [.61, 1.00]	.83 [.62, .98]	<.001
M (range) *d′*	2.23 [−.61, 4.12]	2.36 [.35, 4.35]	1.74 [−.43, 3.97]	<.001
M (range) Commission errors	.18 [.00, .96]	.07 [.00, .42]	.07 [.00, .33]	<.001
M (range) Omission errors	.22 [.00, 1.00]	.27 [.00, .95]	.44 [.00, .96]	<.001
M (range) RT	904.17 [418, 1301]	570.95 [443, 665]	405.68 [278, 490]	<.001
M (range) RT Variability [*]	234.72 [0, 497]	97.56 [41, 200]	63.58 [5, 196]	<.001

[*] 3 children who completed Version 1, 1 child who completed Version 2 and 2 children who completed Version 3 were missing RT variability scores due to insufficient responses in this phase of the task.

Table 3:

Model fit statistics from tests for invariance of age-specific GNG/CPT task versions

Model	χ^2	df	RMSEA	CFI	TLI	SRMR	χ^2	df	RMSEA	CFI	TLI	SRMR	Decision (accept or reject invariance)
Configural	47.92	39	.036	.995	.995	.028	-	-	-	-	-	-	Accept
Version 1	16.32	13	.044	.995	.989	.024	-	-	-	-	-	-	-
Version 2	18.88	13	.058	.979	.954	.036	-	-	-	-	-	-	-
Version 3	12.72	13	0	1	1	.020	-	-	-	-	-	-	-
Full metric	69.82 ***	51	.049	.988	.981	.089	21.90 *	12	.010	.007	.007	.061	Reject
Partial metric	66.16 **	50	.046	.990	.981	.077	18.24	11	.007	.005	.005	.049	Accept
Full scalar	121.31 *	63	.079	.964	.952	.131	73.39 ***	24	.040	.010	.036	.089	Reject
Partial scalar	76.76 *	59	.045	.989	.984	.074	29.27	20	.006	.006	.004	.046	Accept
Full strict	204.58 ***	72	.111	.919	.905	.201	212.47 **	33	.072	.108	.069	.173	Reject
Partial strict	87.26	67	.045	.988	.984	.112	39.82	28	.006	.006	.004	.082	Accept

*
$p < .05$;

**
$p < .01$;

***
$p < .00$

Table 4.

Correlations among various GNG/CPT task metrics

	1	2	3	4	5	6	7	8	9	10	11
1. GNG Accuracy											
2. GNG *d′*	.85 ***										
3. GNG Commission Errors	−.34 ***	−.73 ***									
4. GNG Omission Errors	−.90 ***	−.55 ***	−.11 *								
5. GNG *MRT*	−.08	.11 *	−.44 ***	.29 ***							
6. GNG RT Variability	−.47 ***	−.58 ***	.42 ***	.29 ***	−.14 **						
7. CPT Accuracy	.54 ***	.64 ***	−.54 ***	−.32 ***	.13 **	−.53 ***					
8. CPT *d′*	.61 ***	.64 ***	−.43 ***	−.44 ***	.06	−.54 ***	.88 ***				
9. CPT Commission Errors	−.33 ***	−.52 ***	.60 ***	.07	−.24 ***	.46 ***	−.91 ***	−.65 ***			
1. CPT Omission Errors	−.56 ***	−.37 ***	−.05	.61 ***	.22 ***	.24 ***	−.35 ***	−.67 ***	−.07		
11. CPT *MRT*	−.07	.14 **	−.31 ***	.07	.48 ***	−.07	.40 ***	.26 ***	−.46 ***	.07	
12. CPT RT Variability	−.28 ***	−.39 ***	.36 ***	.12 *	−.11 *	.47 ***	−.48 ***	−.38 ***	.51 ***	−.02	−.22 ***

Note: Correlations partial for children's age and GNG/CPT version; *d′*: d-prime; RT: Response time.

*
$p < .05$;

**
$p < .01$;

***
$p < .001$

Table 5:

Relation of sociodemographic factors to children's performance on the GNG/CPT task

	Accuracy	d'	Commission errors	Omission errors	MRT	RT Variability
			GNG Phase			
Intercept	.43 *** (.29, .56)	−1.53 ** (−2.51, −.56)	.91 *** (.68, 1.14)	.46 *** (.29, .63)	924.43 *** (824.05, 1024.81)	390.41 *** (343.40, 437.42)
Age & Test Version						
Age (in years)	.07 (−.17, .30)	1.10 (−.61, 2.82)	−.59 ** (−.99, −.18)	.11 (−.19, .41)	−301.95 *** (−478.84, −125.05)	−215.32 *** (−298.19, −132.46)
Test Version 2	−.12 (−.28, .04)	.34 (−.82, 1.49)	−.42 ** (−.70, −.15)	.30 ** (.10, .51)	−516.94 *** (−635.87, −398.01)	−293.11 *** (−348.81, −237.41)
Test Version 3	.11 *** (.07, .15)	.93 *** (.67, 1.20)	−.16 *** (−.22, −.10)	−.09 *** (−.14, −.04)	−37.98 *** (−65.19, −10.78)	−38.48 * (−51.22, 25.74)
Age x Version 2	−.05 * (−.10, .00)	−.55 ** (−.91, −.18)	.15 *** (.06, .24)	.02 (−.05, .08)	14.09 (−23.82, 52.00)	27.57 ** (9.81, 45.33)
Age x Version 3	−.06 *** (−.10, −.02)	−.65 *** (−.92, −.37)	.14 *** (.08, .21)	.03 (−.02, .08)	31.26 * (3.04, 59.47)	35.54 *** (22.32, 48.75)
Sex						
Female	.00 (−.02, .02)	.16 * (.02, .30)	−.07 ** (−.10, −.03)	.02 (.00, .05)	19.79 ** (5.44, 34.14)	−1.75 (−8.47, 4.97)
Maternal Education						
HS/GED or Less	.00 (−.03, .02)	−.04 (−.23, .15)	.01 (−.04, .05)	.00 (−.03, .04)	−3.72 (−23.31, 15.87)	1.69 (−7.49, 10.87)
Household Income						
$60,000	−.03 * (−.05, .00)	−.25 * (−.42, −.08)	.05 ** (.01, .09)	.02 (−.01, .05)	−2.96 (−20.08, 14.16)	9.16 * (1.07, 17.26)
			CPT Phase			
Intercept	.18 ** (.04, .32)	−2.53 *** (−3.66, −1.40)	.83 *** (.66, 1.01)	.79 *** (.54, 1.03)	1032.57 *** (912.13, 1153.01)	302.53 *** (236.29, 368.77)
Age & Test Version						
Age (in years)	.39 ** (.14, .64)	.68 (−1.32, 2.67)	−.64 *** (−.95, −.34)	.37 (−.06, .80)	−306.49 *** (−516.10, −96.88)	−207.37 *** (−322.60, −92.13)
Test Version 2	.38 *** (.22, .55)	1.02 (−.32, 2.36)	−.70 *** (−.91, −.50)	.57 *** (.28, .86)	−586.79 *** (−728.62, −444.97)	−225.71 *** (−303.78, −147.64)
Test Version 3	.18 *** (.14, .21)	1.32 *** (1.01, 1.63)	−.18 *** (−.23, −.14)	−.16 *** (−.23, −.09)	−34.09 * (−66.62, −1.57)	−19.53 * (−37.41, −1.65)
Age x Version 2	−.12 *** (−.17, −.07)	−0.55 ** (−.98, −.12)	.16 *** (.09, .23)	.00 (−.10, .09)	6.71 (−38.30, 51.72)	18.65 (−6.08, 43.38)
Age x Version 3	−.14 *** (−.18, −.10)	−0.93 *** (−1.25, −.62)	.17 (.13, .22)	.05 (−.02, .12)	29.98 (−3.72, 63.68)	17.23 (−1.30, 35.75)
Sex						
Female	.01 (−.01, .03)	.14 (−.02, .30)	−.02 (−.05, .00)	.01 (−.02, .05)	9.74 (−7.22, 26.70)	−2.87 (−12.17, 6.44)
Maternal Education						
HS/GED or Less	.00 (−.02, .03)	.02 (−.21, .24)	−.01 (−.04, .03)	.00 (−.05, .05)	−11.25 (−34.39, 11.89)	−9.93 (−22.58, 2.72)

	Accuracy	d′	Commission errors	Omission errors	MRT	RT Variability
Household Income						
$60,000	−.03 ** (−.06, −.01)	−.25 * (−.45, −.05)	.04 ** (.01, .07)	.01 (−.03, .05)	−8.41 (−29.05, 12.24)	15.08 ** (3.81, 26.34)

Note. HS: High school; Categories not shown, including male gender, maternal education > HS and household income > $60, 000 are the reference categories; race/ethnicity was also included in the models;

*
$p < .05$;

**
$p < .01$;

***
$p < .001$

Table 6:

Partial correlations between CPT/GNG task metrics and parent-reported behavioral problems

	CBCL : Int	CBCL : Ext	CBCL : Atte	BRIEF GEC	BRIEF Inh	BRIEF Shift	BRIEF EC	BRIEF WM	BRIEF PO
GNG Phase									
Accuracy	−.04	−.04	−.04	−.11*	−.05	−.06	−.07	−.17***	−.11*
d′	'.01	'.10*	'.07	'.14**	'.10*	'.06	'.08	'.20***	'.15**
Commission Errors	.01	.13*	.08	.12*	.09	.04	.06	.15**	.12**
Omission Errors	.04	−02	.02	.06	.01	.04	.05	.11*	.06
*M*RT	.06	−.05	−.06	.01	.07	.04	.01	−.02	−.02
RT Variability	.01	.10*	.05	.06	.07	−.01	.01	.10*	.06
CPT Phase									
Accuracy	−.03	−.08	−.04	−.10*	−.11*	−.05	−.05	−.13**	−.08*
d′	−.01	−.08	−.05	−.13**	−.14**	−.07	−.07	−.16**	−.10*
Commission Errors	.01	.08	.03	.06	.08	.01	.03	.09	.05
Omission Errors	.07	.02	.02	.11*	.08	.10*	.06	.12*	.08
*M*RT	−.06	−.10	−.08	−.02	−.04	.01	−.01	−.04	.02
RT Variability	.01	.10*	.07	.08	.08	.03	.05	.08	.10*

Note: Correlations partial for children's age and test version. Int: Internalizing problems; Ext: Externalizing problems; Atte: Attention difficulties; GEC: General Executive Composite; Inh: Inhibitory control difficulties; EC: Emotional Control difficulties; WM: Working memory difficulties; PO: Planning and Organization difficulties.

**
p < .01;

*
p < .05