Research on Child and Adolescent Psychopathology (2024) 52:505–520
https://doi.org/10.1007/s10802-023-01162-w

Explaining Brain-Behavior Relations: Inhibitory Control as an Intermediate Phenotype Between the N2 ERP and the Externalizing Spectrum in Childhood

Alexis Hosch[1] · Benjamin Swanson[2] · Jordan L. Harris[1] · Jacob J. Oleson[3] · Eliot Hazeltine[1] · Isaac T. Petersen[1]

Accepted: 12 December 2023 / Published online: 15 January 2024

Abstract

Identifying neural and cognitive mechanisms in externalizing problems in childhood is important for earlier and more targeted intervention. Meta-analytic findings have shown that smaller N2 event-related potential (ERP) amplitudes, thought to reflect inhibitory control, are associated with externalizing problems in children. However, it is unclear *how* (i.e., through which cognitive processes) N2 amplitudes relate to externalizing problems. We examined whether inhibitory control may be a cognitive process that links N2 amplitudes and externalizing problems in early childhood. Children ($N = 147$, 74 girls) were assessed at four time points, spanning 3–7 years of age. Children's externalizing behavior was assessed via questionnaires completed by mothers, fathers, and teachers/secondary caregivers. Children's inhibitory control was assessed using eleven performance-based tasks and two questionnaires. Developmental scaling linked differing measures of inhibitory control and externalizing behavior across ages onto the same scale. Children's N2 amplitudes were extracted from electroencephalography data collected during a go/no-go task. Smaller N2 amplitudes were associated with externalizing problems and poorer inhibitory control. A concurrent analysis of indirect effects revealed that poorer inhibitory control partially explained the association between smaller N2 amplitudes and externalizing problems, even when controlling for the child's age, sex, and socioeconomic status. This is among the first studies to link N2 amplitudes, inhibitory control, and externalizing problems during early childhood. Findings suggest that smaller N2 amplitudes may be an early neural indicator of inhibitory control deficits and externalizing psychopathology. Moreover, inhibitory control may be an important target for early intervention in the development of externalizing psychopathology.

Keywords N2 ERP · Externalizing problems · Inhibitory control · Children · EEG · RDoC

Externalizing behavior problems consist of children's outward behaviors and reactions to external cues from the environment, such as aggression, inattention, hyperactivity, and conduct problems (Liu, 2004). The worldwide prevalence of externalizing disorders is ~ 5.7%, or 113 million children (Polanczyk et al., 2015). Moreover, individual differences in externalizing behaviors tend to be relatively stable throughout the life span and are associated with severe outcomes, including academic underachievement (Hinshaw, 1992), substance use (Petersen et al., 2015), and criminality (White et al., 1990). Therefore, it is crucial to identify mechanisms in the development of externalizing behaviors before these behaviors develop into severe problems later in life. It may be especially important to identify biological and cognitive mechanisms underlying externalizing behavior, because a given behavior (e.g., deficient self-regulation) can reflect different underlying substrates and can appear across several disorders. That is, the same behavior can occur for different reasons. Thus, behavior ratings are not sufficient to make conclusions about mechanisms in the development of psychopathology (Insel, 2014).

✉ Alexis Hosch
alexis-hosch@uiowa.edu

[1] Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA

[2] Department of Psychological Science, University of Arkansas, Fayetteville, USA

[3] Department of Biostatistics, University of Iowa, Iowa City, USA

Accordingly, the Research Domain Criteria (RDoC) initiative from the National Institute of Mental Health seeks to advance understanding of neural substrates of psychopathology (Insel et al., 2010). The RDoC initiative was developed to address shortcomings of traditional categorical conceptualizations of psychopathology that are based on a set of symptoms. Instead of using diagnostic categories, the RDoC framework uses dimensional conceptualizations of psychopathology, in which typical and atypical behavioral development are examined across the lifespan. The RDoC framework provides an organizational structure (i.e., matrix) for researchers to characterize the nature of psychopathology across units of analysis. The matrix specifies six major domains (e.g., Cognitive Systems) consisting of constructs (e.g., cognitive control), subconstructs (e.g., inhibitory control), and units of analysis (e.g., physiological, self-report, and behavioral data). Within the RDoC framework, it is important to identify neurobiological (and other) processes underlying the development of psychopathology. Notably, one initial criticism of the RDoC framework was that it largely excluded developmental data in its formulation (Conradt et al., 2021; Durbin et al., 2022). However, recent studies have sought to integrate developmental perspectives into the RDoC framework. For example, Vogel et al. (2021) used an RDoC approach to examine how trajectories of emotion dysregulation in positive (i.e., excitability) and negative (i.e., irritability) affect in childhood predicted emotion dysregulation in adolescence. Similarly, Damme et al. (2022) used the RDoC framework to examine associations between patterns of irritability in early childhood (i.e., preschool age and early school age) and later psychopathology (internalizing and externalizing symptoms) and neural outcomes in preadolescence. Thus, the RDoC framework provides a structure for researchers to examine trajectories of constructs across units of analysis, allowing researchers to examine the development of typical and atypical behavior (e.g., psychopathology) over time.

One possible biological process in the development of externalizing behavior is neural activity in the prefrontal cortex or anterior cingulate cortex, as indexed by the N2 event-related potential (ERP). The N2 ERP is commonly examined using tasks designed to assess inhibitory control. Inhibitory control is the ability to inhibit responses to irrelevant stimuli in pursuit of a cognitively represented goal (Simpson & Carroll, 2019). In the RDoC framework, inhibitory control is a subconstruct of the cognitive systems domain and cognitive control construct. A widely used paradigm to assess inhibitory control is the go/no-go task. During go/no-go tasks, children are presented with two stimuli: a go stimulus, which is paired with response activation (e.g., a button press), and a no-go stimulus, which is paired with response inhibition. Go stimuli are often presented more frequently than no-go stimuli to elicit a prepotent response and make inhibition more difficult. The N2 ERP, extracted using electroencephalography (EEG), is the second negative deflection in the waveform that occurs approximately 300–500 ms post stimulus in children to both go and no-go stimuli (Hoyniak, 2017). Importantly, the inhibitory (no-go/stop) N2 component has been widely associated with externalizing behavior in children. Meta-analytic work has shown that smaller (i.e., less negative) no-go N2 amplitudes are associated with more externalizing behavior in children (Hoyniak & Petersen, 2019). However, the mechanism underlying the association between no-go N2 amplitudes and externalizing problems is unclear.

Given the costs and practical challenges of using neural substrates in intervention and prevention efforts, it is important to identify cognitive intermediate phenotypes in the association between biological processes and externalizing behavior. Cognitive intermediate phenotypes between inhibitory N2 amplitudes and externalizing problems may capture early neural risk processes, while providing practical targets for intervention. Furthermore, identifying links between the brain and behavior is aligned with the RDoC framework, which encourages researchers to build a bridge that spans the same construct across multiple units of analysis. A cognitive process that may be a potential intermediate phenotype between no-go N2 amplitudes and externalizing problems is inhibitory control.

Although the functional interpretation of the N2 component is not yet established, research supports the interpretation of the N2 as an index of inhibitory control in inhibitory tasks (Jing et al., 2021; Jodo & Kayama, 1992), such as go/no-go tasks (Hoyniak, 2017; Hoyniak & Petersen, 2019). The N2 is larger to inhibition (i.e., no-go) stimuli than to activation (i.e., go) stimuli. Moreover, when experimentally manipulating effort by setting limits on adults' reaction time, Jodo and Kayama (1992) found that no-go N2 amplitudes were larger when greater effort was required to withhold the prepotent response. Research has localized the no-go N2 component to areas thought to support inhibitory control (Steele et al., 2013), including the anterior cingulate cortex, orbitofrontal cortex, ventral prefrontal cortex, and dorsolateral prefrontal cortex (Bokura et al., 2001; Lavric et al., 2004). Taken together, these findings provide evidence for the interpretation of the no-go N2 as an index of inhibitory control. However, other functional interpretations of the N2 have been proposed, including that the inhibitory (i.e., no-go and stop) N2 reflects conflict monitoring or attention to mismatched stimuli (Enriquez-Geppert et al., 2010; Folstein & Van Petten, 2008; Smith et al., 2010). Given the association between no-go N2 amplitudes and externalizing problems, it is important to clarify the functional interpretation of the N2 component.

Several studies have examined the association between N2 amplitudes and inhibitory control in children (Brydges

et al., 2014; Espinet et al., 2012; Jing et al., 2021; Kaiser et al., 2006). A larger difference between go and no-go N2 amplitudes is thought to reflect more advanced inhibitory capacities (Jodo & Kayama, 1992). Thus, it might be expected that larger no-go N2 amplitudes would be associated with better inhibitory control, because smaller N2 no-go amplitudes may reflect insufficient recruitment of neural resources necessary for inhibition. However, mixed findings in children have emerged. Several studies have shown that larger no-go N2 amplitudes are associated with better inhibitory control in children (Grabell et al., 2017; Hoyniak, 2017; Ruberry et al., 2017). By contrast, meta-analytic work has found that *smaller* no-go N2 amplitudes are associated with better inhibitory control (Buss et al., 2011; Espinet et al., 2012; Hoyniak & Petersen, 2019). Given these inconsistent findings in children, it is important to consider developmental changes in N2 amplitudes and self-regulatory processes. On average, no-go N2 amplitudes decrease and self-regulation abilities increase with age during childhood (Berger, 2011; Hoyniak, 2017). Thus, if inhibition is successful, smaller no-go N2 amplitudes may reflect more efficient and mature neural processing. Further, it is possible that children who remain less efficient in the neural processes required for inhibition have larger no-go N2 amplitudes than typically developing children. That is, older children with poorer inhibitory control skills via less efficient neural processing may have larger no-go N2 amplitudes. This may explain why some studies have found that smaller no-go N2 amplitudes (more efficient processing) are associated with better inhibitory control, whereas others have found that larger no-go N2 amplitudes (more advanced inhibitory capacity) are associated with better inhibitory control. More research is needed to clarify the nature of the association between no-go N2 amplitudes and inhibitory control in children.

Nevertheless, inhibitory control deficits are robustly associated with externalizing problems (Schoemaker et al., 2013). Inhibitory control deficits predict later externalizing problems in children (Kahle et al., 2018) and growth in externalizing behaviors across development (Perry et al., 2018).

Given that (a) the no-go N2 may reflect inhibitory control processes, (b) inhibitory control deficits are associated with externalizing behavior, and (c) no-go N2 amplitudes are associated with externalizing behavior, inhibitory control may be an intermediate cognitive phenotype that explains the relation between the N2 component and externalizing behavior. To date, no studies have examined whether inhibitory control processes may be a mechanism that accounts for the association between no-go N2 amplitudes and externalizing behavior problems in children. Identifying the cognitive processes underlying the association between no-go N2 amplitudes and externalizing behavior could help identify intervention targets that are more clinically practical for intervention than neural processes.

The Present Study

The aim of the present study is to identify neural and cognitive processes underlying externalizing problems in childhood, and to determine whether inhibitory control is a cognitive intermediate phenotype between the no-go N2 ERP and externalizing problems in childhood. In the RDoC framework, it is important to identify intermediate phenotypes that explain how neural processes relate to behavior. Consistent with this framework, we aimed to examine the same construct (i.e., cognitive control or disinhibition) across multiple units of analysis, including physiology, paradigms, and behavior. To do so, we examined whether inhibitory control concurrently mediated the association between N2 amplitudes on inhibition (i.e., no-go) trials and externalizing problems in 3–7-year-old children. We hypothesized that no-go N2 amplitudes would be positively associated with externalizing behavior, such that smaller, less negative no-go N2 amplitudes would be associated with greater externalizing problems, consistent with prior studies (Hoyniak & Petersen, 2019). Second, we hypothesized that no-go N2 amplitudes would be associated with inhibitory control, but we had no a priori hypothesis about the sign of the association given mixed findings. Third, we hypothesized that inhibitory control would be negatively associated with externalizing behavior, such that better inhibitory control would be associated with fewer externalizing problems, consistent with prior research (e.g., Buss et al., 2014; Kahle et al., 2018; Perry et al., 2018). Finally, we hypothesized that inhibitory control would partially mediate the association between no-go N2 amplitudes and externalizing behavior. Specifically, we hypothesized that smaller, less negative no-go N2 amplitudes would be associated with inhibitory control, whose deficits in turn would be associated with externalizing problems.

The N2 may also be related to broader executive function-related processes, of which inhibitory control is a component (Espinet et al., 2012). However, the present study focuses on brain activity in response to a particular trial condition (no-go trials) of a specific inhibitory control paradigm (i.e., go/no-go). ERPs assess neural processes at a particular timing and are thought to index particular cognitive processes. Thus, we sought to identify the cognitive process related to the no-go N2 component with as much precision as possible. Because prior research suggests that no-go N2 amplitudes may reflect inhibitory control processes, we focus on inhibitory control in this study. This approach aligns closely with the dimensional approach of the RDoC framework, in which researchers

are encouraged to relate biological processes to simpler, lower-order, narrower subdimensions of psychological constructs (Macedo et al., 2021), including the subconstruct of inhibitory control.

Method

Participants

A community sample of children ($N = 147$, $M_{age} = 4.81$ years, SD = 1.18 years, 74 girls), and their caregivers participated in an ongoing accelerated longitudinal study. Participants were recruited at four ages: 36 ($n = 40$), 45 ($n = 38$), 54 ($n = 32$), or 63 ($n = 37$) months. The full sample of children spanned 3 to 7.5 years of age. The inclusion criterion to be recruited for the study was that the child was one of the target ages (described above). Exclusion criteria were: the child's primary caregiver did not speak English, or the child did not have a permanent guardian, did not have normal or corrected-to-normal vision and hearing, or was not capable of communicating or following basic instructions in English. Participants were recruited in 2018–2023 from the greater Iowa City community via university listservs, databases acquired from the University of Iowa Hospital and Clinics, local daycares and preschools, community flyers, social media, snowball sampling, and by word of mouth. Reasons for participant ineligibility and a flowchart of the final sample are in Supplementary Fig. S1. Reasons for missingness and tests of systematic missingness are in Supplementary Appendix S1. The sample consisted of children, their primary caregiver, the primary caregiver's parenting partner (as applicable), and a teacher/secondary caregiver (e.g., nanny, babysitter, or someone else who knew the child well).

The sample of children was 66.7% Non-Hispanic White, 9.5% Hispanic or Latino, 6.8% Black or African American, 4.8% Asian, 6.1% multiracial, and 6.1% other race. For the consented primary caregivers ($n = 148$), and parenting partners ($n = 139$), 97% were biological parents, 1% were stepparents, 1% were adoptive parents, and less than 1% were grandparents or other caregivers. The level of educational attainment across primary caregivers and parenting partners was: 7.8% doctoral degree, 7.5% professional degree, 21.3% master's degree, 30.1% bachelor's degree, 11.3% associate degree, 14.4% some college, 5.6% high school graduate, 1.9% some high school (Grades 9–12, no degree). Additionally, among primary caregivers and parenting partners, 85.5% were married, 8.3% were single/never married, 3.4% were divorced, 1.5% were re-married, and 1.2% were separated.

Procedures

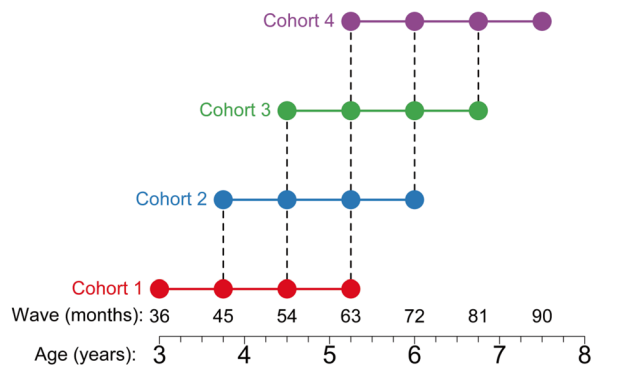
Children and their primary caregiver completed two lab visits, one week apart, every 9 months for four time points (see Fig. 1). During the first lab visit ($M_{minutes} = 152.51$, $SD = 20.80$ min), children completed behavioral tasks, including inhibitory control tasks, while the primary caregiver completed questionnaires, including ratings of their child's inhibitory control and externalizing problems. Parenting partners and secondary caregivers rated the child's externalizing problems via online questionnaires. During the second lab visit ($M_{minutes} = 89.75$, $SD = 18.83$ min), children completed several tasks, including a computerized go/no-go task while electroencephalography (EEG) was recorded. Video examples of procedures are available on Databrary (https://nyu.databrary.org/volume/1559).

Measures

The present study is part of a larger study, the School Readiness Study. Measures and hypotheses for the School Readiness Study were preregistered: https://osf.io/jzxb8. Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are published online: https://osf.io/e2nkr. Descriptive statistics and bivariate correlations for all study variables are in Table 1. Estimates of reliability (inter-rater, internal consistency, cross-time stability) for study measures are in Supplementary Table S3.

Inhibitory Control

Thirteen measures, including questionnaires and laboratory tasks, were used to assess inhibitory control. Laboratory



Note. Accelerated longitudinal research design with four cohorts. The longitudinal design follows any given child for 2¼ years, with testing every nine months; the whole data set spans the ages of 3–7½ years. Circles reflect measurement points (four waves) for each cohort. Dashed lines indicate common measurement points across cohorts.

Fig. 1 Accelerated Longitudinal Research Design

Table 1 Correlations among Predictors, Outcomes, and Covariates

Variable	Age	Sex	SES	N2	IC	EXT
Age	—					
Sex	0.07	—				
SES	0.15**	-0.03	—			
N2	-0.16*	-0.11	0.03	—		
IC	0.60****	0.23****	0.25****	-0.23****	—	
EXT	-0.45****	-0.11**	-0.16****	0.17****	-0.38****	—
Obs	336	590	586	158	307	647
M	4.81	0.50	-0.10	-3.16	0.81	0.46
SD	1.18	0.50	0.80	6.10	0.67	15.70
Min	2.92	0.00	-3.38	-24.26	-0.65	-61.05
Max	7.80	1.00	3.47	15.24	2.63	43.77

Age is in years. Sex is coded such that 1 = female and 0 = male. Given the strong, cross-time rank-order stability of SES ($r = 0.90$, $p < .001$), we interpolated missing SES values at a given time point by carrying a participant's last observation forward. Sex values at later time points were interpolated by carrying a participant's last observation forward. That is, values for sex and SES were included in the data even when the child had not yet come in for a visit, which explains the larger number of observations of sex and SES compared to age, N2 amplitudes, and inhibitory control in the data used for analysis. 69% of children had available N2 data at one or more timepoints. Much of the greater missingness in the N2 amplitudes relative to questionnaire ratings of inhibitory control was due to COVID (see Supplementary Appendix S1). Correlations and descriptive statistics are presented using the data structured in long form. Correlations and descriptive statistics for all variables, except for externalizing problems, have one observation per participant at a given time point in the study (up to four timepoints). Correlations and descriptive statistics of externalizing problems included observations from up to three raters (i.e., primary caregiver, parenting partner, and secondary caregiver) at a given time point for a given participant (up to four time points), which explains the larger number of observations than the other variables

SES socioeconomic status, *N2* N2 amplitudes for correct no-go trials, *IC* Inhibitory control, *EXT* Externalizing problems, *Obs* number of observations for a given variable, *Min* minimum observed value, *Max* maximum observed value

*$p < 0.10$; **$p < 0.05$; ***$p < 0.005$; ****$p \leq 0.001$, all ps are two-tailed

tasks included: Bear/Dragon, Day/Night, Grass/Snow, Hand Game, Knock/Tap, Less is More, Peg Tapping, Shape Stroop, and Simon Says. Computerized inhibitory control tasks included: Fish/Sharks and Stop-Signal. Additionally, caregivers reported on their child's inhibitory control using the Behavioral Rating of Executive Function (BRIEF) and Children's Behavior Questionnaire (CBQ). Detailed descriptions of each measure are in Supplementary Appendix S2.

For developmental scaling, scores of each measure of externalizing problems and inhibitory control were converted to proportion of maximum (POM) scores to have the same possible range (0–1), with higher scores reflecting greater externalizing problems and inhibitory control, respectively. Proportion scores are widely recommended by longitudinal researchers for studying growth with different measures (Little, 2013; Moeller, 2015). For measures that had a minimum and maximum possible score, the POM score reflected the proportion of the maximum possible score. For measures that did not have a minimum or maximum possible score (i.e., Stop-Signal task), the POM score reflected the proportion of the maximum *observed* score. POM scores were calculated as: $\frac{\text{score}-\text{minimum}}{\text{maximum}-\text{minimum}}$, where minimum and maximum were the minimum and maximum possible or observed score. Tasks (Stop-Signal Task) and questionnaires (BRIEF) were adapted to accommodate the developmental capacity of the child and the changing expression of inhibitory control with age.

Bear/Dragon Bear/Dragon (Kochanska et al., 1996) is a go/no-go task. The child was asked to follow instructions from a bear puppet, and to ignore instructions from a dragon puppet, and then rules were reversed. There were 12 trials. Each no-go trial was scored from 1 to 4 (1 = full commanded movement, 2 = partial movement, 3 = wrong movement, and 4 = no movement). Scoring was reversed for go trials, consistent with Carlson and Moses (2001). Consistent with Eisenberg et al. (2013), a composite of children's inhibition was computed by multiplying mean scores from six inhibition (no-go) and six activation (go) trials; children who activated a behavior on go trials and inhibited on no-go trials received the highest scores, whereas children who never activated (or always activated) a behavior received low scores.

Day/Night In Day/Night (Gerstadt et al., 1994), the child was instructed to say "day" when they saw the card with the black moon and say "night" when they saw the card with the yellow sun. Sixteen test trials were scored incorrect (0), initially incorrect, but changed to correct (1), or correct (2). Scores were averaged across trials (0–2).

Fish/Sharks Fish/Sharks (Wiebe et al., 2012) is a go/no-go task and was administered on a computer using E-Prime software (version 2.0.10.356; Schneider et al., 2012). The child was shown cartoon images of fish (go stimuli) and sharks (no-go stimuli) on a touch screen and was instructed to touch the fish and not to touch the sharks. A composite of children's inhibition was computed by multiplying the proportion of correct inhibition (no-go) trials (20 trials) by the proportion of correct activation (go) trials (60 trials), consistent with Eisenberg et al. (2013).

Grass/Snow In Grass/Snow (Carlson & Moses, 2001), the child was instructed to touch a white square when they heard the word "grass" and a green square when they heard the word "snow." Twelve test trials were scored either correct (1) or incorrect (0), consistent with Carlson and Moses (2001). Final scores were averaged across trials (0–1).

Hand Game In Hand Game (Luria et al., 1964), the child was instructed to point a finger when the experimenter made a fist, and to make a fist when the experimenter pointed a finger. Fifteen test trials were scored incorrect (0), initially incorrect, but changed to correct (1), or correct (2), consistent with Kochanska et al.'s (1997) scoring of other inhibitory tasks. Scores were averaged across all trials (0–2).

Knock/Tap In Knock/Tap (Klenberg et al., 2001), the child was instructed to knock on the table when the experimenter tapped, and to tap the table whenever the experimenter knocked. In the second part of the task, the instructions changed. The child was instructed to make a side fist when the experimenter knocked, to knock when the experimenter made a side fist, and to do nothing when the experimenter tapped the table. Fifteen test trials were scored incorrect (0), initially incorrect, but changed to correct (1), or correct (2), consistent with Kochanska et al.'s (1997) scoring of other inhibitory tasks. Scores were averaged across trials (0–2).

Less is More Less is More is a motivationally salient symbolic representation task that assesses affective ("hot") inhibitory control (Carlson et al., 2005). The child chose a preferred treat from two options, white marshmallows and uniformly colored jellybeans. In front of the child were two bowls, one of which had a "naughty monkey" puppet, and the other bowl was the child's bowl. The child was told that "the monkey wants all the treats for himself." On each trial, two bags were presented to the child: one bag with five treats and one bag with two treats. The child was instructed to point to a bag among the two bag options presented and that the bag they point to goes to the monkey's bowl, and that the child receives the other bag (i.e., the bag they did not point to). Responses were scored as: 0 = child points to large treats bag; 1 = child initially points to the large treats bag, then changes to the small treats bag; 2 = child points to the small treats bag, consistent with Kochanska et al.'s (1997) scoring of other inhibitory tasks. Scores were averaged across 16 test trials (0–2).

Peg Tapping In Peg Tapping (Luria et al., 1964), the child observed sequences of a specific number of pencil taps on a table (either one or two) and was instructed to tap a pencil the opposite number of times of what they observed. For example, if the experimenter taps the pencil once, the child is to tap the pencil twice and vice versa. Sixteen trials were scored correct (1) or incorrect (0). Scores were averaged across trials (0–1).

Shape Stroop Shape Stroop (Kochanska et al., 2000) assesses children's perceptual inhibitory control. The task assessed the child's ability to identify a picture of a small fruit embedded within a picture of a different, larger fruit. Six test trials were scored from 0 to 2 (0 = incorrect, 1 = initially incorrect, but changed response to correct, 2 = correct). Scores were averaged across the three small fruit trials (0–2).

Simon Says In Simon Says (Strommen, 1973), the child was instructed to perform simple motor actions (e.g., clap your hands, stomp your feet) and was told to perform the action only if the instructions are preceded by the phrase "Simon Says." Each no-go trial was scored from 1 to 4 (1 = full commanded movement, 2 = partial movement, 3 = wrong movement, and 4 = no movement), consistent with Carlson and Moses (2001) scoring of a simplified version of Simon Says (Bear/Dragon); scoring was reversed for go trials. A composite score was computed by multiplying mean scores from 10 go trials and 10 no-go trials (20 trials total), consistent with Eisenberg et al. (2013).

Stop-signal Task In a stop-signal task adapted from Berger et al. (2013), the child was told to give purple food to the purple pig and green food to the green goat by touching the animal on the screen. The child was then shown a cartoon wizard and told that the wizard will try to trick them and turn the food into a car. The child was instructed not to feed cars to animals and not to touch the screen when they saw a car. Blocks two and three had the same structure, with different animals. The latency of stop signal after go stimulus onset (i.e., stop-signal delay [SSD]) was manipulated based on the child's performance to obtain as close to a 50% error

rate as possible on stop trials, which helped normalize task difficulty across ages. Stimuli were presented via E-Prime software (Schneider et al., 2012). Response inhibition was operationalized as the stop-signal reaction time (SSRT). The SSRT was calculated as the median reaction time on correct go trials minus the mean SSD from Blocks 2 and 3. Block 1 was not included in the calculation to allow the algorithm time to converge upon a 50% error rate on stop trials. Cases were excluded if the SSRT was negative (i.e., the median go reaction time was faster than the mean SSD). Scores were reverse scored so that higher scores reflected greater inhibitory control.

Behavior Rating Inventory of Executive Function (BRIEF) The BRIEF assesses children's executive functioning within the context of their everyday environment. Two versions were used based on the child's age. Parents completed the BRIEF–Preschool Version (BRIEF–P; Gioia et al., 1996) if the child was 3–5 years old or the BRIEF–2 (Gioia et al., 2015) if the child was 6–7 years old. Scores on the Inhibitory Control subscale were used for both versions of the questionnaire. Twenty-four items were rated on a 3-point scale (1 = never, 2 = sometimes, 3 = often) in terms of how often, in the last six months, the child's behavior had been a problem. To account for missing responses in the sum score, scores were averaged across items and then multiplied by the number of items. Scores were reverse scored so that higher scores reflected greater inhibitory control. Mothers' and fathers' ratings on the Inhibitory Control Composite were correlated ($r[152] = 0.38$, $p < 0.001$).

Children's Behavior Questionnaire (CBQ) The CBQ assesses children's temperament (i.e., reactivity and regulation). Parents completed the CBQ (Putnam & Rothbart, 2006). Secondary caregivers completed the CBQ–Teacher Short Form (CBQ–TSF, Teglasi et al., 2015). Scores from the Inhibitory Control scale (CBQ: 47 items; CBQ–TSF: 26 items) were used. Items were rated on a 7-point Likert scale (1 = extremely untrue, 7 = extremely true). Scores were averaged across items. Mothers' ratings on the Inhibitory Control scale were associated with ratings by fathers ($r[164] = 0.46$, $p < 0.001$) and secondary caregivers ($r[165] = 0.31$, $p < 0.001$). Fathers' ratings were associated with ratings by secondary caregivers ($r[112] = 0.35$, $p < 0.001$).

Externalizing Problems

Achenbach System of Empirically Based Assessment The Achenbach System of Empirically Based Assessment (ASEBA) assesses children's emotional and behavioral problems. Items were rated on a 3-point Likert scale according to how well the item described the child (0 = not true, 1 = somewhat or sometimes true, 2 = very true). Multiple versions were used based on the child's age and rater type. Parents completed the Child Behavior Checklist 1.5–5 (CBCL 1.5–5; Achenbach & Rescorla, 2000) if the child was 3–5 years old or the Child Behavior Checklist 6–18 (CBCL 6–18; Achenbach & Rescorla, 2001) if the child was 6–7 years old. Secondary caregivers completed the Caregiver–Teacher Report Form (C–TRF; Achenbach & Rescorla, 2001) if the child was 3–5 years old or the Teacher's Report Form (TRF; Achenbach & Rescorla, 2001) if the child was 6–7 years old. Scores on the Externalizing scale were used. Mothers' ratings on the Externalizing scale were associated with ratings by fathers ($r[178] = 0.56$, $p < 0.001$) and secondary caregivers ($r[174] = 0.46$, $p < 0.001$). Fathers' ratings were associated with ratings by secondary caregivers ($r[123] = 0.44$, $p < 0.001$). Age and sex norm-referenced T-scores had a mean of 46.35 ($SD = 9.73$). Using T-scores of 65 or greater as a clinical cutoff, ~ 2.4% of ratings in the study were in the at-risk or clinical range on the Externalizing scale; ~ 6.8% of children were in the at-risk or clinical range at one or more timepoints based on ratings from one or more raters.

Covariates

We examined models with and without covariates. Covariates included the child's age, sex, and family socioeconomic status (SES). Socioeconomic status was calculated as the average of three z-scored (relative to the sample) indices: income-to-needs ratio, parent educational attainment, and parent occupational prestige. Given the strong, cross-time rank-order stability of SES ($r = 0.90$, $p < 0.001$), we interpolated missing SES values at a given time point by carrying a participants last observation forward. A full description of covariates is in Supplementary Appendix S3.

Electrophysiological Recordings and Data Processing

Electrophysiological data were collected using an Electrical Geodesic, Inc (EGI) 128-electrode Hydrocel Geodesic Sensor Net with a Net Amps 400 series amplifier. Net Station Acquisition Software 5.4.2 (Electrical Geodesics Inc., 2018) was used to collect the continuous EEG data. EEG data were collected during a go/no-go task (i.e., Fish/Sharks), which was administered using E-Prime 2.0.10.356 (Schneider et al., 2012). A detailed description of the collection and pre-processing of the EEG data is in Supplementary Appendix S4.

Data were pre-processed in Net Station Tools 5.4.3 (Electrical Geodesics Inc., 2018). Continuous data were band-pass filtered from 0.1 to 30.0 Hz. Correct go and correct no-go trials were selected and segmented into 1200 ms epochs that began 200 ms prior to the presentation of each stimulus. Epochs were then automatically inspected for artifacts, which included identifying and removing "bad" channels. Epochs were marked bad if they contained more than

20 bad channels, an eye blink, or an eye movement. Channels were marked bad across all epochs if 20% or greater of the epochs were marked bad. Channels marked bad across all epochs were removed. Removed channels were interpolated based on the waveforms of surrounding electrodes. If a child did not have at least 10, artifact-free trials in each condition after automatic processing, epochs were manually examined for artifacts. After manually identifying and removing artifacts and bad channels, epochs were subjected to the same automatic inspection procedure described above. Epochs were then averaged within participants, and re-referenced to an average reference (i.e., the average of all scalp electrodes). Finally, epochs were baseline corrected by subtracting the average activity from each epoch's 200 ms baseline.

Data were excluded from analyses if the child did not have at least 8 correct, artifact-free trials in each condition after manual processing, consistent with prior studies with children (e.g., Hoyniak et al., 2018). Data were also excluded if the child refused to wear the EEG net, refused to complete the task (i.e., Fish/Sharks), or if there were technical errors during the EEG collection. A total of 102 children (69% of the full sample of children [$N = 147$]), had available EEG data. EEG data were more likely to be missing for children with poorer inhibitory control and for children from lower SES families. EEG data were not missing as a function of age, sex, ethnicity, or externalizing problems.

Following pre-processing, we conducted temporospatial principal component analysis (tsPCA) to decompose the EEG waveform. All PCA analyses were conducted using the ERP PCA Toolkit (version 2.98, Dien, 2010). We performed tsPCA separately for each condition (i.e., go versus no-go trials), consistent with prior research which found that a combined PCA (i.e., including both trial types in one PCA) misallocated substantial variance (Barry et al., 2018). For sensitivity analyses, we also performed tsPCA separately for two age groups across conditions, consistent with Scharf et al. (2022). PCAs were conducted for younger (36–54 months, $n = 81$ observations) and older children (63–90 months, $n = 77$ observations). However, the no-go N2 amplitudes from the age-combined versus age-separated PCA were strongly correlated ($r[472] = 0.64$, $p < 0.001$). Thus, we used the N2 amplitudes from the age-combined PCA to help ensure that we extracted the same ERP component across ages. A description of the tsPCA analyses and results is in Supplementary Appendix S5.

The grand averaged waveform is depicted in Supplementary Fig. S2. The temporospatial component, thought to correspond with the N2 component, was selected based on a priori hypotheses about the latency (typically 300–500 ms post stimulus onset), topography, and morphology of the component. The selected N2 component was characterized by a frontocentral negativity (see Supplementary Fig. S3) that peaked at 427 ms in the go condition, and 466 ms in the no-go condition. N2 amplitudes on inhibition (i.e., no-go) trials were extracted and used in analyses. Amplitudes were extracted from a cluster of electrodes whose loading on the N2 temporospatial component was 0.5 or greater (see Supplementary Fig. S3) at the peak latency (i.e., 466 ms; Scharf et al., 2022). The N2 tsPCA component waveform is depicted in Supplementary Fig. S4.

Statistical Analysis

Exploratory Factor Analysis

We first examined whether inhibitory control measures' scores were able to be modeled with item response modeling by examining their scores in exploratory factor analysis (EFA). Results of the EFA models supported item response modeling; see Supplementary Appendix S6.

Developmental Scaling Approach

We used developmental scaling to link scores from the different measures across ages onto the same scale (Hosch et al., 2022). In this way, we could estimate meaningful individual differences in inhibitory control and externalizing problems from age-differing measures across 3–7 years of age. To perform developmental scaling, we used a two-parameter Bayesian longitudinal item response model in a mixed modeling item response theory (IRT) framework. Details of the developmental scaling approach are in Supplementary Appendix S7.

Mediation Models

Mediation models were fit in a structural equation modeling (SEM) framework. First, we fit separate models to estimate unadjusted associations (i.e., not controlling for covariates) of N2 amplitudes with inhibitory control and externalizing problems. Second, we estimated concurrent mediation models (i.e., analyses of indirect effects of concurrent associations) that included all three variables. We fit SEM models using the sem() function of the lavaan 0.6–16 package (Rosseel, 2012) in R 4.2.0 (Team, 2022). SEM models were fit with FIML estimation, which uses all available data and is the gold standard approach for handling missingness when data are missing at random or completely at random (Enders & Bandalos, 2001). Models were fit with a robust maximum likelihood estimator that provides robust standard errors to account for nonnormally distributed data. Following recommendations, the indirect effect was estimated with bias-corrected bootstrapped confidence intervals (Hayes, 2009; Shrout & Bolger, 2002). Confidence intervals were estimated from 10,000 bootstrap samples. Models were saturated—i.e., there were no degrees of freedom because only manifest

variables were included. Thus, fit indices indicated perfect model fit. Given the range of ages included in the study, we included the child's age as a covariate. To account for the nonindependence of data owing to multiple observations from the same participant, we also conducted a Bayesian multilevel mediation analysis with random intercepts for each child.

The effect size of the indirect effect was calculated with three estimates: (1) the standardized regression coefficient (beta, β) of the indirect effect, (2) the proportion of the effect that was mediated (P_M), which is the ratio of the indirect effect to the total effect (Wen & Fan, 2015), and (3) the proportion of variance in externalizing problems that was accounted for jointly by N2 amplitudes and inhibitory control (upsilon, υ; Lachowicz et al., 2018). Upsilon was estimated using the upsilon() function of the MBESS 4.9.2 package (Kelley, 2007) in R.

Data Structure

To leverage all time points of data for all participants for greater power, we stacked the data in long form for the structural equation models, so that each combination of participant, timepoint, and rater uniquely identified each row. Participants could have observations from up to three raters (i.e., primary caregiver, parenting partner, and/or secondary caregiver) and up to four time points (i.e., waves). Thus, each participant could have up to 12 rows of observations. When transforming the data from wide to long format, N2 and inhibitory control scores were unique for each combination of participant and timepoint and were thus applied to all rows of a given timepoint (i.e., wave) for that participant. Externalizing problem scores were unique for each combination of participant, timepoint, and rater. The structure of the data is depicted in Supplementary Table S6. Given modest cross-informant associations of externalizing problems, the long form data structure allowed us to make use of all raters' perspectives and all available information without losing information by averaging or aggregating across raters.

Sensitivity Analyses

We conducted several sensitivity analyses to (1) include cluster-robust standard errors, (2) account for additional covariates, (3) examine moderated mediation by sex, (4) examine models using latent variables of inhibitory control estimated by performance-based tasks and questionnaires separately, (5) examine the specificity of the N2 component, and (6) examine whether results changed when using N2 amplitudes extracted from different electrodes.

Results

N2 Amplitudes and Externalizing Problems

As expected, N2 amplitudes were positively associated with externalizing problems (β = 0.17, $p < 0.018$) in an unadjusted model, such that smaller (less negative) N2 amplitudes were associated with greater externalizing problems. This association was somewhat attenuated and only marginally significant after controlling for the child's age (β = 0.10, $p = 0.071$). Externalizing problems decreased with age (β = 0.042, $p < 0.001$).

N2 Amplitudes and Inhibitory Control

N2 amplitudes were negatively associated with inhibitory control in an unadjusted model (β = -0.25, $p < 0.001$), such that larger (more negative) N2 amplitudes were associated with better inhibitory control. This association held controlling for the child's age (β = -0.17, $p = 0.021$). Inhibitory control increased with age (β = 0.57, $p < 0.001$).

Inhibitory Control and Externalizing Problems

Consistent with hypotheses, inhibitory control was negatively associated with externalizing problems in an unadjusted model (β = -0.38, $p < 0.001$). That is, poorer inhibitory control was associated with greater externalizing problems. The association held controlling for the child's age (β = -0.17, $p = 0.006$).

Mediation Models

Without Covariates

First, we examined the indirect effect of N2 amplitudes on externalizing problems via inhibitory control without controlling for covariates. There was a significant total effect (β = 0.17, 95% CI [0.17, 0.66], $p = 0.001$). After accounting for inhibitory control, the direct effect of N2 amplitudes on externalizing problems was nonsignificant (β = 0.08, 95% CI [-0.05, 0.43], $p = 0.122$). Moreover, we observed an indirect effect of inhibitory control (β = 0.09, 95% Cl [0.14, 0.33], $p < 0.001$), which accounted for approximately 54% ($P_M = 0.536$) of the variance in the association between the N2 amplitudes and externalizing problems. The upsilon value estimate of the effect size of the indirect effect was $u_{adj} = 0.00752$.

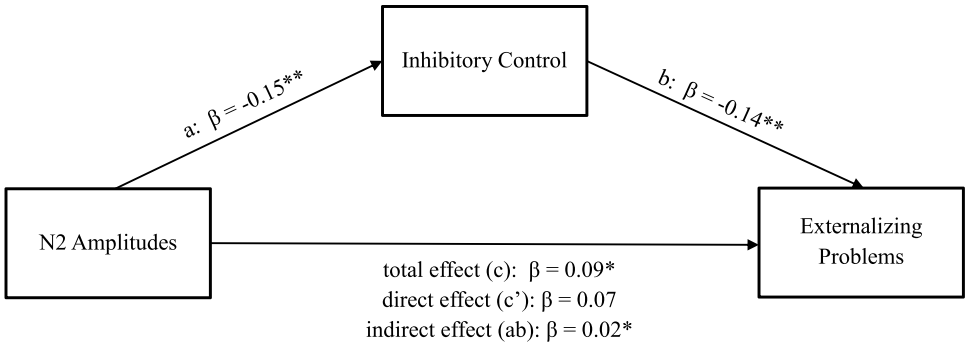


Fig. 2 Mediation Model with Covariates

Note. Mediation model controlling for child's age, sex, and family SES. * $p < .05$. ** $p < .01$.

With Covariates

After controlling for the child's age and sex, there remained a significant total effect (β = 0.09, 95% CI [0.002, 0.46], $p = 0.049$). The direct effect of N2 amplitudes on externalizing problems accounting for inhibitory control was not significant after controlling for covariates (β = 0.07, 95% CI [-0.06, 0.41], $p = 0.149$). There was a significant indirect effect of inhibitory control (β = 0.02, 95% Cl [0.02, 0.11], $p = 0.031$), which accounted for approximately 23% ($P_M = 0.229$) of the variance in the association between the N2 amplitudes and externalizing problems after controlling for covariates. The mediation model is depicted in Fig. 2. The regression coefficients for the mediation models (i.e., with and without covariates) are in Table 2.

Bayesian Multilevel Mediation Model

Without Covariates To account for the nonindependence of data owing to multiple observations from the same participant, we conducted an additional Bayesian multilevel mediation analysis with random intercepts for each child. Results were largely the same. N2 amplitudes remained associated with externalizing problems (i.e., total effect; $B = 0.48$, 95% ETI [0.20, 0.76]). After accounting for inhibitory control, the association between the N2 and externalizing problems was significant (i.e., direct effect; $B = 0.33$, 95% ETI [0.03, 0.64]). Additionally, there was a significant indirect effect of inhibitory control ($B = 0.14$, 95% ETI [0.01, 0.30]), which accounted for approximately 29% ($P_M = 0.288$) of the variance in the association between the N2 amplitudes and externalizing problems.

Table 2 Regression Coefficients for the Mediation Models

Pathway	*B*	β	SE	*p*	95% CI
Model without Covariates					
N2 → IC	-0.02	-0.23	0.004	< 0.001	[-0.03, -0.02]
IC → EXT	-9.22	-0.38	1.33	< 0.001	[-11.85, -6.62]
N2 → EXT (direct effect)	0.19	0.08	0.12	0.122	[-0.05, 0.43]
N2 → EXT (total effect)	0.41	0.17	0.13	0.001	[0.17, 0.66]
indirect effect	0.15	0.09	0.05	< 0.001	[0.14, 0.33]
Model with Covariates					
Age → N2	-0.73	-0.14	0.27	0.007	[-1.24, -0.19]
Age → IC	0.32	0.57	0.02	< 0.001	[0.29, 0.36]
Age → EXT	-4.53	-0.34	0.60	< 0.001	[-5.71, -3.34]
Sex → N2	-1.21	-0.10	0.55	0.029	[-2.32, -0.13]
Sex → IC	0.23	0.17	0.03	< 0.001	[0.16, 0.30]
Sex → EXT	-1.03	-0.03	1.16	0.376	[-3.26, 1.28]
N2 → IC	-0.02	-0.15	0.004	< 0.001	[-0.02, -0.01]
IC → EXT	-3.27	-0.14	1.15	0.005	[-5.57, -1.07]
N2 → EXT (direct effect)	0.17	0.07	0.12	0.149	[-0.06, 0.41]
N2 → EXT (total effect)	0.23	0.09	0.12	0.049	[0.002, 0.46]
indirect effect	0.05	0.02	0.02	0.031	[0.02, 0.11]

Age is in years. Sex is coded such that 1 = female and 0 = male. The models with and without covariates were fit separately

N2 N2 amplitudes for correct no-go trials, *IC* Inhibitory Control, *EXT* Externalizing problems, *CI* Confidence Interval

With Covariates After controlling for the child's age and sex, the total effect was no longer significant ($B = 0.24$, 95% ETI [-0.03, 0.52]). The direct effect of N2 amplitudes on externalizing problems accounting for inhibitory control remained significant after controlling for covariates ($B = 0.31$, 95% ETI [0.03, 0.59]). Additionally, the indirect effect was somewhat attenuated and was significant at a trend level after accounting for covariates ($B = -0.07$, 95% ETI [-0.16, 0.001]). Unexpectedly, inhibitory control scores were *positively* associated with externalizing problems in the mediation model (despite having a negative bivariate association: $r = -0.38$). It appears that the unexpected sign of the association between inhibitory control and externalizing problems was likely due to controlling for a variable (age) that was strongly associated with both (age and inhibitory control: $r = 0.60$; age and externalizing problems: $r = -0.45$). After removing age as a covariate, inhibitory control was negatively associated with externalizing problems, as expected. The indirect effect held ($B = 0.14$, 95% ETI [0.001, 0.289]), and it accounted for approximately 28% ($P_M = 0.284$) of the variance in the association between the N2 amplitudes and externalizing problems.

Sensitivity Analyses

We conducted several sensitivity analyses, as described in Supplementary Appendix S8. Notably, results did not substantially differ when controlling for additional covariates, examining questionnaire- and task-based measures of inhibitory control separately, or when extracting N2 amplitudes from Fz or the peak negative channel. Indirect effects were specific to the timing of the N2 but were not specific to the frontocentral region; similar effects were observed with a positive-going waveform that was likely a dipole of the N2 and was centrally distributed. The child's sex did not moderate the indirect effect.

Power Analysis

We conducted a post-hoc power analysis to estimate the statistical power of detecting an indirect effect given our sample size and observed effect sizes (see Supplementary Appendix S9). We had power of 0.83 to detect significance in the observed indirect effect.

Discussion

The current study integrated information across physiological (i.e., the N2 ERP) and cognitive (i.e., inhibitory control) units of analysis to inform our understanding of externalizing behavior problems in early childhood. Previous research has shown that smaller N2 amplitudes may be an early neural biomarker of externalizing behavior problems in children (Hoyniak & Petersen, 2019). However, is it unclear *how* N2 amplitudes relate to externalizing problems. Cognitive intermediate phenotypes in the association between neural substrates (e.g., the N2 component) and externalizing behavior may provide more practical targets for intervention than neural substrates. Furthermore, the RDoC framework encourages researchers to examine the same construct (i.e., common process) across multiple units of analysis. Consistent with this framework, a potential cognitive intermediate phenotype between N2 amplitudes and externalizing behavior is inhibitory control. Research supports the interpretation of the N2 component as an index of inhibitory control when examined using go/no-go tasks (Jodo & Kayama, 1992), particularly in children (Hoyniak, 2017; Hoyniak & Petersen, 2019). Moreover, deficits in inhibitory control have been widely associated with externalizing problems (Buss et al., 2014; Kahle et al., 2018; Perry et al., 2018). Thus, we investigated whether the RDoC subconstruct inhibitory control concurrently mediated the association between physiological processes (i.e., the N2 ERP) and disinhibited behavior (i.e., externalizing problems). That is, we aimed to examine the same construct—cognitive control or disinhibition—across several units of analysis, including physiology, behavior, and paradigms, consistent with the RDoC framework.

We examined these associations in a community sample of young children (ages 3–7). A community sample is relevant because externalizing problems are considered a dimensional spectrum (Markon et al., 2011). Examining basic processes that underlie dimensional differences in externalizing problems is consistent with aims of the RDoC framework, which uses dimensional conceptualizations of psychopathology and encourages researchers to examine typical and atypical behavioral development. It also aligns with other emerging nosologies, such as the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017).

As expected, we found a positive association between the N2 component and externalizing behavior. This finding replicates prior meta-analytic work, which has found that smaller, less negative N2 amplitudes are associated with more externalizing behavior problems in children (Hoyniak & Petersen, 2019). N2 amplitudes were negatively associated with inhibitory control, such that smaller N2 amplitudes were associated with poorer inhibitory control. We also observed negative associations between inhibitory control and externalizing behavior. This is consistent with prior literature, which has found that poorer inhibitory control skills predicted greater externalizing problems (Buss et al., 2014; Kahle et al., 2018; Olson et al., 2005; Perry et al., 2018). Moreover, we found an indirect effect of N2 amplitudes on externalizing problems via inhibitory control. That is,

inhibitory control—including questionnaire and task-based operationalizations—partially mediated the association between N2 amplitudes and externalizing problems. Moderation models demonstrated that the indirect effect did not differ for boys and girls, suggesting that the mediation process operated similarly for boys and girls. Interestingly, the indirect effect showed some specificity to the timing of the N2 but not specificity to the frontocentral region; similar effects were observed with a likely positive-going dipole of the N2 that was centrally distributed.

Implications

This is among the first studies to identify associations between N2 amplitudes, inhibitory control, and externalizing problems during early childhood. Findings provide additional evidence for N2 amplitudes as an early neural indicator of externalizing psychopathology, consistent with prior research (Hoyniak & Petersen, 2019). Notably, however, the magnitude of the association between N2 amplitudes and externalizing behavior (i.e., the total effect) was small ($\beta s = 0.09 - 0.17$). Thus, it may be important to consider additional neural risk processes for externalizing problems. For instance, work in adults has shown that the P3 ERP may better capture processes related to response inhibition when using a stop-signal paradigm (Wessel & Aron, 2015). There are likely other neural processes, e.g., the error-related negativity (ERN) ERP, that contribute to the development of externalizing problems (Lutz et al., 2021).

We also found that larger N2 amplitudes were associated with better inhibitory control. This is consistent with some previous research (Grabell et al., 2017; Ruberry et al., 2017), but inconsistent with other studies, including a prior meta-analysis (Hoyniak & Petersen, 2019). It is possible that age-related differences in the N2 could explain why larger N2 amplitudes are associated with better inhibitory control in some children but not others. However, there are inconsistencies between studies examining children of similar ages. Studies with similar age ranges (i.e., ages 3–5) have found that both larger (e.g., Grabell et al., 2017; Ruberry et al., 2017) and smaller (e.g., Buss et al., 2011; Espinet et al., 2012) N2 amplitudes are associated with better inhibitory control. Thus, there may be reasons other than age for the inconsistent findings. It is possible that the relation between N2 amplitudes and inhibitory control is non-linear, in which extreme variation in either direction (i.e., small or large N2 amplitudes) may confer risk for inhibitory control deficits. Speculatively, excessively small N2 amplitudes may reflect the insufficient recruitment of neural resources that are necessary for inhibition. By contrast, excessively large N2 amplitudes may reflect over-recruitment of neural resources, reflecting inefficient processing. Ultimately, more research is needed to clarify the nature of the association between N2 amplitudes and inhibitory control in children.

Results from the present study suggest that inhibitory control may be a cognitive intermediate phenotype between N2 amplitudes and externalizing problems. Inhibitory control deficits may capture early neural risk processes for externalizing psychopathology. Moreover, these findings suggest that inhibitory control may be a key target for early intervention in the development of externalizing problems. Thus, interventions targeting inhibitory control skills, or self-regulation, may be useful for the prevention of later externalizing problems. Studies suggest that curriculum-based interventions, such as Tools of the Mind or Red Light, Purple Light, may be effective and practical interventions to improve self-regulation in children (Diamond et al., 2019; McClelland et al., 2019; Pandey et al., 2018). Notably, inhibitory control accounted for approximately 23–28% of the variance in the association between N2 amplitudes and externalizing problems. Thus, it is important to consider additional cognitive processes through which neural processes such as the N2 ERP may lead to externalizing problems. Future research should examine additional cognitive processes, such as conflict monitoring or attention processes, that may also help explain the association between N2 amplitudes and externalizing psychopathology (Enriquez-Geppert et al., 2010; Folstein & Van Petten, 2008; Smith et al., 2010).

Strengths

The study had several strengths. First, the predictor (i.e., N2 amplitudes), mediator (i.e., inhibitory control), and outcome (i.e., externalizing problems) were assessed via distinct methods, which reduces the extent to which the indirect effect could be accounted for by method bias. Moreover, our assessment of inhibitory control included several measurement methods, including behavioral tasks and questionnaires to reduce the effects of common method variance. In addition, our latent variable of inhibitory control was estimated using several tasks beyond the task in which N2 amplitudes were extracted (i.e., Fish/Sharks), reducing potential measure-specific bias. Second, questionnaire data were collected from multiple informants, including mothers, fathers, and teachers or other caregivers to gain a more accurate estimate of children's real-world functioning and behavior across contexts. Third, we applied methods (i.e., developmental scaling) to maintain the developmental relevance of measures across a wide age range. We linked scores from age-differing measures onto the same scale, which allowed us to examine individual differences in inhibitory control and externalizing behavior across ages 3–7. Fourth, findings held across many sensitivity analyses, providing greater confidence in our inferences.

Finally, it is notable that we observed these associations in a community sample, in which externalizing behaviors were less prevalent compared to a clinical sample. It will be valuable for future work to replicate and extend these findings in clinical samples. We also make our data and analysis scripts freely available to promote dissemination.

Limitations

The study also had limitations. First, the study was correlational and examined concurrent associations. Thus, we cannot make causal inferences. Because of the ongoing nature of this study, we would currently be underpowered to examine lagged associations. Specifically, we were constrained by limited data available at time points three and four ($n = 51$), which would be needed for a fully longitudinal mediation model. Future research should examine the longitudinal relations between the N2 component, inhibitory control, and externalizing problems. Second, we had some missingness in ERP data, much of which was due to COVID. Nevertheless, the study had a larger sample of participants with ERP data ($n = 102$) than many studies of young children. Third, there were some differences in missingness as a function of demographic characteristics. N2 amplitudes were more likely to be missing for children with poorer inhibitory control and for children from lower SES families. Inhibitory control scores were more likely to be missing for children from lower SES families and for boys. Externalizing problems ratings were more likely to be missing for older children, for children from lower SES families, for boys, and for children of "other" race. Systematic missingness may limit the generalizability of our findings. However, effect sizes of systematic missingness were small, and findings did not substantially change after including the child's age, sex, and family SES as control variables in our models. Fourth, cross-informant associations of inhibitory control ($rs = 0.31 – 0.46$) and externalizing behavior ($rs = 0.44 – 0.56$) were modest. However, these associations are similar in magnitude to those observed in prior studies (Carneiro et al., 2021). Our modeling approaches estimated latent variables from the common variance of measures, which may miss context-specific behavior. Thus, it may be beneficial in future studies to examine these associations separately for parents and teachers.

Conclusion

Small N2 amplitudes are a commonly studied neural marker of externalizing behavior in children. However, the mechanisms that explain how the N2 is associated with externalizing problems are unclear. In the RDoC framework, it is important to identify intermediate phenotypes that explain how neural processes relate to behavior. Intermediate phenotypes (e.g., cognitive processes) may provide more practical targets for intervention. Thus, the current study examined whether inhibitory control mediated the association between N2 amplitudes and externalizing problems in young children. We found that smaller, less negative N2 amplitudes were related to externalizing problems, consistent with prior research (Hoyniak & Petersen, 2019). Smaller N2 amplitudes were also associated with poorer inhibitory control, which in turn was associated with externalizing behavior problems. That is, inhibitory control partially mediated the association between N2 amplitudes and externalizing problems. This study is the first to examine cognitive intermediate phenotypes of the association between neural processes and externalizing psychopathology in childhood. Findings suggest that inhibitory control deficits may be an early indicator of biological risk for externalizing psychopathology. Inhibitory control may be an important target for early intervention.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10802-023-01162-w.

Funding The project was funded by Grants HD098235 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), T32GM108540 from the National Institute of General Medical Sciences (NIGMS), and UL1TR002537 from the National Center for Advancing Translational Sciences (NCATS).

Data Availability The present study is part of a larger study, the School Readiness Study. Measures and hypotheses for the School Readiness Study were pre-registered: https://osf.io/jzxb8. Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are published online: https://osf.io/e2nkr.

Compliance with Ethical Standards

Conflict of Interest We have no conflicts of interest to disclose.

Ethical Approval The present study was approved by the University of Iowa Institutional Review Board (Study #: 201708761).

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms and Profiles: An integrated system of multi-informant assessment*. University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont, Department of Psychiatry.

Barry, R. J., De Blasio, F. M., & Fogarty, J. S. (2018). A processing schema for children in the auditory equiprobable go/nogo task: ERP components and behaviour. *International Journal of Psychophysiology, 123*, 74–79. https://doi.org/10.1016/j.ijpsycho.2017.10.014

Berger, A. (2011). Self-regulation: Brain, cognition, and development. *American Psychological Association*. https://doi.org/10.1037/12327-000

Berger, A., Alyagon, U., Hadaya, H., Atzaba-Poria, N., & Auerbach, J. G. (2013). Response inhibition in preschoolers at familial risk for attention deficit hyperactivity disorder: A behavioral and electrophysiological stop-signal study. *Child Development, 84*(5), 1616–1632. https://doi.org/10.1111/cdev.12072

Bokura, H., Yamaguchi, S., & Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a go/nogo task. *Clinical Neurophysiology, 112*(12), 2224–2232. https://doi.org/10.1016/s1388-2457(01)00691-5

Brydges, C., Fox, A., Reid, C., & Anderson, M. (2014). Predictive validity of the N2 and P3 ERP components to executive functioning in children: A latent-variable analysis. *Frontiers in Human Neuroscience, 8*, 80. https://doi.org/10.3389/fnhum.2014.00080

Buss, K. A., Dennis, T. A., Brooker, R. J., & Sippel, L. M. (2011). An ERP study of conflict monitoring in 4–8-year old children: Associations with temperament. *Developmental Cognitive Neuroscience, 1*(2), 131–140. https://doi.org/10.1016/j.dcn.2010.12.003

Buss, K. A., Kiel, E. J., Morales, S., & Robinson, E. (2014). Toddler inhibitory control, bold response to novelty, and positive affect predict externalizing symptoms in kindergarten. *Social Development, 23*(2), 232–249. https://doi.org/10.1111/sode.12058

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science, 16*(8), 609–616. https://doi.org/10.1111/j.1467-9280.2005.01583.x

Carneiro, A., Soares, I., Rescorla, L., & Dias, P. (2021). Meta-analysis on parent–teacher agreement on preschoolers' emotional and behavioural problems. *Child Psychiatry & Human Development, 52*(4), 609–618. https://doi.org/10.1007/s10578-020-01044-y

Conradt, E., Crowell, S. E., & Cicchetti, D. (2021). Using development and psychopathology principles to inform the Research Domain Criteria (RDoC) framework. *Development and Psychopathology, 33*(5), 1521–1525. https://doi.org/10.1017/S0954579421000985

Damme, K. S. F., Norton, E. S., Briggs-Gowan, M. J., Wakschlag, L. S., & Mittal, V. A. (2022). Developmental patterning of irritability enhances prediction of psychopathology in preadolescence: Improving RDoC with developmental science. *Journal of Psychopathology and Clinical Science, 131*(6), 556–566. https://doi.org/10.1037/abn0000655

Diamond, A., Lee, C., Senften, P., Lam, A., & Abbott, D. (2019). Randomized control trial of Tools of the Mind: Marked benefits to kindergarten children and their teachers. *PLoS ONE, 14*(9), e0222447. https://doi.org/10.1371/journal.pone.0222447

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods, 187*(1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009

Durbin, C. E., Wilson, S., & MacDonald, A. W. (2022). Integrating development into the Research Domain Criteria (RDoC) framework: Introduction to the special section. *Journal of Psychopathology and Clinical Science, 131*(6), 535–541. https://doi.org/10.1037/abn0000767

Eisenberg, N., Edwards, A., Spinrad, T. L., Sallquist, J., Eggum, N. D., & Reiser, M. (2013). Are effortful and reactive control unique constructs in young children? *Developmental Psychology, 49*(11), 2082–2094. https://doi.org/10.1037/a0031745

Electrical Geodesics Inc. (2018). *Net Station 5.4 EEG software* [Computer software].

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Enriquez-Geppert, S., Konrad, C., Pantev, C., & Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *NeuroImage, 51*(2), 877–887. https://doi.org/10.1016/j.neuroimage.2010.02.043

Espinet, S. D., Anderson, J. E., & Zelazo, P. D. (2012). N2 amplitude as a neural marker of executive function in young children: An ERP study of children who switch versus perseverate on the dimensional change card sort. *Developmental Cognitive Neuroscience, 2*, S49–S58. https://doi.org/10.1016/j.dcn.2011.12.002

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition, 53*(2), 129–153. https://doi.org/10.1016/0010-0277(94)90068-x

Gioia, G. A., Andrwes, K., & Isquith, P. K. (1996). *Behavior Rating Inventory of Executive Function: Preschool version (BRIEF-P) [Measurement Instrument].* Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function: Second edition (BRIEF2) [Measurement Instrument].* Odessa, FL: Psychological Assessment Resources.

Grabell, A. S., Olson, S. L., Tardif, T., Thompson, M. C., & Gehring, W. J. (2017). Comparing self-regulation-associated event related potentials in preschool children with and without high levels of disruptive behavior. *Journal of Abnormal Child Psychology, 45*(6), 1119–1132. https://doi.org/10.1007/s10802-016-0228-7

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408–420. https://doi.org/10.1080/03637750903310360

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*(1), 127–155. https://doi.org/10.1037/0033-2909.111.1.127

Hosch, A., Oleson, J. J., Harris, J. L., Goeltz, M. T., Neumann, T., LeBeau, B., Hazeltine, E., & Petersen, I. T. (2022). Studying children's growth in self-regulation using changing measures to account for heterotypic continuity: A Bayesian approach to developmental scaling. *Developmental Science, 25*, e13280. https://doi.org/10.1111/desc.13280

Hoyniak, C. P. (2017). Changes in the nogo N2 event-related potential component across childhood: A systematic review and meta-analysis. *Developmental Neuropsychology, 42*(1), 1–24. https://doi.org/10.1080/87565641.2016.1247162

Hoyniak, C. P., & Petersen, I. T. (2019). A meta-analytic evaluation of the N2 component as an endophenotype of response inhibition and externalizing psychopathology in childhood. *Neuroscience & Biobehavioral Reviews, 103*, 200–215. https://doi.org/10.1016/j.neubiorev.2019.06.011

Hoyniak, C. P., Petersen, I. T., Bates, J. E., & Molfese, D. L. (2018). The neural correlates of temperamental inhibitory control in toddlers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1744). https://doi.org/10.1098/rstb.2017.0160

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2014). Mental disorders in childhood: Shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA, 311*(17), 1727–1728. https://doi.org/10.1001/jama.2014.1193

Jing, J., Zhang, Z., Qi, M., & Gao, H. (2021). The fronto-central N2 but not parietal P3 reflects response inhibition in the count/no-count task. *Psychological Research Psychologische Forschung*. https://doi.org/10.1007/s00426-021-01571-w

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a go/no-go task. *Electroencephalography and Clinical Neurophysiology, 82*(6), 477–482. https://doi.org/10.1016/0013-4694(92)90054-L

Kahle, S., Utendale, W. T., Widaman, K. F., & Hastings, P. D. (2018). Parasympathetic regulation and inhibitory control predict the development of externalizing problems in early childhood. *Journal of Abnormal Child Psychology, 46*(2), 237–249. https://doi.org/10.1007/s10802-017-0305-6

Kaiser, S., Weiss, O., Hill, H., Markela-Lerenc, J., Kiefer, M., & Weisbrod, M. (2006). N2 event-related potential correlates of response inhibition in an auditory go/nogo task. *International Journal of Psychophysiology, 61*(2), 279–282. https://doi.org/10.1016/j.ijpsycho.2005.09.006

Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods, 39*(4), 979–984. https://doi.org/10.3758/BF03192993

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20*(1), 407–428. https://doi.org/10.1207/S15326942DN2001_6

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*(2), 490–507.

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68*(2), 263–277. https://doi.org/10.2307/1131849

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*(2), 220–232. https://doi.org/10.1037/0012-1649.36.2.220

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., & Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Lachowicz, M. J., Preacher, K. J., & Kelley, K. (2018). A novel measure of effect size for mediation analysis. *Psychological Methods, 23*(2), 244. https://doi.org/10.1037/met0000165

Lavric, A., Pizzagalli, D. A., & Forstmeier, S. (2004). When 'go' and 'nogo' are equally frequent: ERP components and cortical tomography. *European Journal of Neuroscience, 20*(9), 2483–2488. https://doi.org/10.1111/j.1460-9568.2004.03683.x

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.

Liu, J. (2004). Childhood externalizing behavior: Theory and implications. *Journal of Child and Adolescent Psychiatric Nursing, 17*(3), 93–103.

Luria, A. R., Pribram, K. H., & Homskaya, E. D. (1964). An experimental analysis of the behavioral disturbance produced by a left frontal arachnoidal endothelioma (meningioma). *Neuropsychologia, 2*(4), 257–280. https://doi.org/10.1016/0028-3932(64)90034-X

Lutz, M. C., Kok, R., Verveer, I., Malbec, M., Koot, S., van Lier, P. A. C., & Franken, I. H. A. (2021). Diminished error-related negativity and error positivity in children and adults with externalizing problems and disorders: A meta-analysis on error processing. *Journal of Psychiatry and Neuroscience, 46*(6), E615–E627. https://doi.org/10.1503/jpn.200031

Macedo, I., Pasion, R., Barbosa, F., & Ferreira-Santos, F. (2021). A dimensional approach to the neuronal correlates of anxiety, depression, and perfectionism: A transdiagnostic dissociation of error-related brain activity. *Behavioural Brain Research, 408*, 113271. https://doi.org/10.1016/j.bbr.2021.113271

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137*(5), 856–879. https://doi.org/10.1037/a0023678

McClelland, M. M., Tominey, S. L., Schmitt, S. A., Hatfield, B. E., Purpura, D. J., Gonzales, C. R., & Tracy, A. N. (2019). Red Light, Purple Light! Results of an intervention to promote school readiness for children from low-income backgrounds. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02365

Moeller, J. (2015). A word on standardization in longitudinal studies: don't. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01389

Olson, S. L., Sameroff, A. J., Kerr, D. C. R., Lopez, N. L., & Wellman, H. M. (2005). Developmental foundations of externalizing problems in young children: The role of effortful control. *Development and Psychopathology, 17*(1), 25–45. https://doi.org/10.1017/s0954579405050029

Pandey, A., Hale, D., Das, S., Goddings, A.-L., Blakemore, S.-J., & Viner, R. M. (2018). Effectiveness of universal self-regulation–based interventions in children and adolescents: A systematic review and meta-analysis. *JAMA Pediatrics, 172*(6), 566–575. https://doi.org/10.1001/jamapediatrics.2018.0232

Perry, N. B., Calkins, S. D., Dollar, J. M., Keane, S. P., & Shanahan, L. (2018). Self-regulation as a predictor of patterns of change in externalizing behaviors from infancy to adolescence. *Development and Psychopathology, 30*(2), 497–510. https://doi.org/10.1017/S0954579417000992

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology, 27*(3), 791–818. https://doi.org/10.1017/S0954579414000789

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, 56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://www.jstatsoft.org/v48/i02/

Ruberry, E. J., Lengua, L. J., Crocker, L. H., Bruce, J., Upshaw, M. B., & Sommerville, J. A. (2017). Income, neural executive processes, and preschool children's executive control. *Development and Psychopathology, 29*(1), 143–154. https://doi.org/10.1017/S095457941600002X

Scharf, F., Widmann, A., Bonmassar, C., & Wetzel, N. (2022). A tutorial on the use of temporal principal component analysis in developmental ERP research – Opportunities and challenges. *Developmental Cognitive Neuroscience, 54*, 101072. https://doi.org/10.1016/j.dcn.2022.101072

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime reference guide*. Psychology Software Tools Inc.

Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of Abnormal Child Psychology, 41*(3), 457–471. https://doi.org/10.1007/s10802-012-9684-x

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Simpson, A., & Carroll, D. J. (2019). Understanding early inhibitory development: Distinguishing two ways that children use inhibitory control. *Child Development, 90*(5), 1459–1473. https://doi.org/10.1111/cdev.13283

Smith, J. L., Smith, E. A., Provost, A. L., & Heathcote, A. (2010). Sequence effects support the conflict theory of N2 and P3 in the go/nogo task. *International Journal of Psychophysiology, 75*(3), 217–226. https://doi.org/10.1016/j.ijpsycho.2009.11.002

Steele, V. R., Aharoni, E., Munro, G. E., Calhoun, V. D., Nyalakanti, P., Stevens, M. C., Pearlson, G., & Kiehl, K. A. (2013). A large scale (N=102) functional neuroimaging study of response inhibition in a go/nogo task. *Behavioural Brain Research, 256*, 529–536. https://doi.org/10.1016/j.bbr.2013.06.001

Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon says." *Child Development, 44*(4), 849–853. https://doi.org/10.2307/1127737

Team, R. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Teglasi, H., Schussler, L., Gifford, K., Annotti, L. A., Sanders, C., & Liu, H. (2015). Child behavior questionnaire–short form for teachers: Informant correspondences and divergences. *Assessment, 22*(6), 730–748. https://doi.org/10.1177/1073191114562828

Vogel, A. C., Tillman, R., El-Sayed, N. M., Jackson, J. J., Perlman, S. B., Barch, D. M., & Luby, J. L. (2021). Trajectory of emotion dysregulation in positive and negative affect across childhood predicts adolescent emotion dysregulation and overall functioning. *Development and Psychopathology, 33*(5), 1722–1733. https://doi.org/10.1017/S0954579421000705

Wen, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure. *Psychological Methods, 20*(2), 193–203. https://doi.org/10.1037/met0000029

Wessel, J. R., & Aron, A. R. (2015). It's not too late: The onset of the frontocentral P3 indexes successful response inhibition in the stop-signal paradigm. *Psychophysiology, 52*(4), 472–480. https://doi.org/10.1111/psyp.12374

White, J. L., Moffitt, T. E., Earls, F., & Robins, L. (1990). How early can we tell: Predictors of childhood conduct disorder and adolescent delinquency. *Criminology, 28*, 507.

Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development, 83*(4), 1245–1261. https://doi.org/10.1111/j.1467-8624.2012.01765.x

Supplementary Appendix S1. Description of Missing Data and Tests of Systematic Missingness.

The number of observations by variable, i.e., the inverse of the extent of missingness, is in Table 1. Among possible participant-by-wave instances, 28.9% had missing scores because the child was not yet eligible for a given wave. A total of 158 participant-by-wave instances had data with at least 8 correct, artifact-free trials in each condition, and thus had N2 amplitudes that were used in analyses. Among missing EEG visits at a given wave for which the child reached eligibility, reasons for missingness included: not interested (15%), too busy (15%), moved/relocated (2%), unable to contact (13%), coronavirus (COVID-19) pandemic (44%), and other (12%). Thus, over half of missing instances were due to the COVID-19 pandemic or to not yet being eligible. We suspended lab visits for 14 months during the COVID-19 pandemic (March 2020 – April 2021). We continued to collect online questionnaires from families during the pandemic but were unable to collect EEG assessments during this period. Among those with EEG visits, reasons for missingness were as follows: child refused to wear the EEG cap (6%), child refused to play the task (2%), not enough good channels (4%), not enough good trials (18%), and another technical problem (3%). The number of children with 1, 2, 3, and 4 timepoints of data for the N2 ERP, inhibitory control, and externalizing problems in in Supplementary Table S1. In addition, the number of children with data for the N2 ERP, inhibitory control, and externalizing problems at each wave is in Supplementary Table S2.

We examined whether missingness was systematic in the predictor (N2 amplitudes), outcome (externalizing problems), or hypothesized mediator (inhibitory control). There was no systematic missingness in N2 amplitudes as a function of age, sex, ethnicity, or externalizing problems. However, N2 amplitudes were more likely to be missing for families with lower

socioeconomic status (SES; compared to children from families with higher SES; t[353.59] = 3.16, p = .002). Moreover N2 amplitudes were more likely to be missing for children with poorer inhibitory control (compared to children with greater inhibitory control; t[301.22] = 5.59, p < .001).

There was no systematic missingness in inhibitory control as a function of age, ethnicity, or externalizing problems. However, inhibitory control scores were more likely to be missing for children from families with lower SES (compared to children from families with higher SES; t[412] = -2.30, p = .022). Moreover, inhibitory control scores were more likely to be missing for boys than girls (χ^2[1] = 5.00, p = .025).

Externalizing problems ratings were more likely to be missing for older children than younger children (t[776.10] = -3.22, p = .001), likely due to some COVID-related attrition. Moreover, externalizing problem ratings were more likely to be missing for children from families with lower SES (compared to children from families with higher SES; t[1,204.60] = 4.11, p < .001). Externalizing problem ratings were more likely to be missing for boys than girls (χ^2[1] = 8.68, p = .003). Externalizing problem ratings also showed systematic missingness as a function of ethnicity (χ^2[1] = 11.34, p = .045). Pairwise comparisons revealed that missingness in externalizing problem ratings did not differ between Asian, Black or African American, Non-Hispanic White, Hispanic or Latino, or Multiracial groups. However, those with “other” race showed higher rates of missingness than Non-Hispanic Whites (χ^2[1] = 7.58, p = .006).

Effect sizes of differences were small. To account for systematic missingness, we included the child’s age, sex, and the family’s SES as covariates in models.

Supplementary Appendix S2. Detailed Descriptions of Study Measures.

The present study is part of a larger study, the School Readiness Study. Measures and hypotheses for the School Readiness Study were pre-registered: https://osf.io/jzxb8. Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are published online: https://osf.io/e2nkr. Video examples of procedures are available on Databrary (https://nyu.databrary.org/volume/1559). Estimates of reliability (inter-rater, internal consistency, cross-time stability) for study measures are in Supplementary Table S3.

For developmental scaling, scores of each inhibitory control measure were converted to proportion of maximum (POM) scores to have the same possible range (0–1), with higher scores reflecting greater inhibitory control. Proportion scores are widely recommended by longitudinal researchers for studying growth with different measures (Little, 2013; Moeller, 2015). For measures that had a minimum and maximum possible score, the POM score reflected the proportion of the maximum possible score. For measures that did not have a minimum or maximum possible score (i.e., Stop-Signal task), the POM score reflected the proportion of the maximum *observed* score. POM scores were calculated as: $\frac{\text{score} - \text{minimum}}{\text{maximum} - \text{minimum}}$, where minimum and maximum were the minimum and maximum possible or observed score. Tasks (Stop-Signal Task) and questionnaires (BRIEF) were adapted to accommodate the developmental capacity of the child and the changing expression of inhibitory control with age.

Inhibitory Control Tasks

Fourteen measures, including questionnaires and laboratory tasks, were used to assess inhibitory control. Laboratory tasks included: Bear/Dragon, Day/Night, Grass/Snow, Hand Game, Knock/Tap, Less is More, Peg Tapping, Shape Stroop, and Simon Says. Computerized inhibitory control tasks included: Fish/Sharks and Stop Signal. Additionally, caregivers reported

on children's inhibitory control using the Behavioral Rating of Executive Function (BRIEF) and Children's Behavior Questionnaire (CBQ).

Bear/Dragon. Bear/Dragon (Kochanska et al., 1996) is a go/no-go task that assesses children's inhibitory control and shifting. It involves activation on a subset of go trials and inhibition on a subset of no-go trials, based on the cue (i.e., puppet), with a rule reversal. The child was asked to follow instructions from a bear puppet, and to ignore instructions from a dragon puppet. The child completed three go and three no-go practice trials and was reminded of the rule if they failed a trial. Next, they were presented with 12 mixed test trials, including six go (i.e., bear) trials and six no-go (i.e., dragon) trials. Subsequently, the experimenter changed the rules and instructed the child to now follow the dragon's directions and to ignore the bear. After six practice trials, the child completed a second set of 12 test trials (6 go trials and 6 no-go trials) in a pseudo-random order. Each no-go trial was scored from 1 to 4 (1 = full commanded movement, 2 = partial movement, 3 = wrong movement, and 4 = no movement). Scoring was reversed for go trials, consistent with Carlson and Moses (2001). Scores were averaged across trials within condition (no-go versus go). Because children could receive a high score on no-go trials by performing no action, we examined the degree to which children inhibited a response on no-go trials and activated a response on go trials. Consistent with Eisenberg et al. (2013), a composite of children's inhibition was computed by multiplying mean scores from inhibition (no-go) and activation (go) trials (1–16). Therefore, children who activated a behavior on go trials and inhibited on no-go trials received the highest scores, whereas children who never activated (or always activated) a behavior received low scores. Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater inhibitory control.

Day/Night. Day/Night (Gerstadt et al., 1994) assesses inhibitory control by inhibition of

a prepotent association (that the sun is associated with daytime and the moon is associated with nighttime) and generation of a competing response. In this task, the child was shown two kinds of cards: one with a picture of a sun on a white background and the other with a picture of a moon on a black background. The child was instructed to say "day" when they see the card with the black moon and say "night" when they see the card with the yellow sun. The task began with two practice trials, during which the experimenter praised the child for correct responses. If the child responded incorrectly to either practice trial, the experimenter reminded them of the rules and repeated the trials. After completing the practice trials, the child was presented with 16 test trials, eight of each word, in a fixed, quasi-random order. During the test trials, the experimenter did not provide feedback. Each trial was scored incorrect (0), initially incorrect, but changed to correct (1), or correct (2), consistent with Kochanska et al. (1997) scoring of other inhibitory tasks. The final score was the average score across trials (0–2). Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater inhibitory control.

Fish/Sharks. Fish/Sharks (Wiebe et al., 2012) is a go/no-go task that assesses inhibitory control and was administered on a computer using E-Prime software (version 2.0.10.356; Schneider et al., 2012). During the task, the child was shown cartoon images of fish (go stimuli) and sharks (no-go stimuli) on a touch screen. Stimuli included ten fish and three sharks. However, on any given trial, only one fish or one shark was presented. The child was instructed to touch the fish to catch the fish in their net and not to touch the sharks because the sharks are too big and would break their net. On go trials in which the child touched the fish, positive feedback was presented: an image of the fish in the net and pleasant bubble sounds. On no-go trials in which the child touched the shark, an image of the shark breaking the net and an unpleasant buzzer sound was presented. No feedback was given if the child successfully

inhibited (except during practice trials). The task began with four practice blocks, each with eight practice trials, in the following order: go trials only, no-go trials only, go-trials only, and mixed (i.e., both go and no-go trials) practice block. The experimenter gave feedback during the practice trials. After the child successfully completed the practice trials, the test trials began. The test trials consisted of 80 trials: 60 go trials and 20 no-go trials. The task was split into ten blocks of eight trials. Each block included six go trials and two no-go trials that were randomly presented. The stimuli (i.e., fish or sharks) were presented for a maximum of 3000 milliseconds or until the child touched the screen. Feedback stimuli were presented after the child touched the screen and were displayed for 750 ms. The overall inter-stimulus interval was 1500 ms. The experimenter provided rule reminders during the test trials but did not provide corrective feedback. Behavioral responses that occurred less than 200 ms after stimulus onset were discarded from analyses because this would be too rapid for the child to have responded deliberately to the target stimulus.

A composite of children's inhibition was computed by multiplying the proportion of correct inhibition (no-go) trials by the proportion of correct activation (go) trials, consistent with Eisenberg et al. (2013). Children who activated a behavior on go trials and inhibited on no-go trials received the highest scores, whereas children who never activated (or always activated) a behavior received low scores. Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater inhibitory control.

Grass/Snow. Grass/Snow (Carlson & Moses, 2001) assesses inhibitory control by inhibition of a prepotent association (that the word "grass" is associated with the color green and the word "snow" is associated with the color white) and generation of a competing response. In the task, the child was instructed to touch a white square when they hear the word "grass" and a

green square when they hear the word “snow.” The task began with several practice trials, during which the experimenter praised the child for correct responses. If the child responded incorrectly to a practice trial, the experimenter reminded the child of the rules and repeated the trials. Following these practice trials, the child was presented with 12 trials, six of each word, in a fixed, quasi-random order, and each trial was scored either correct (1) or incorrect (0), consistent with Carlson and Moses (2001). Final scores were averaged across trials (0–1), which reflected a proportion of maximum possible score. Higher scores reflected greater inhibitory control.

Hand Game. Hand game (Luria et al., 1964) assesses inhibitory control. In this task, the child was instructed to either point a finger or make a fist, in response to the experimenter’s hand movement. During the six initial imitation checks, the child copied the experimenter’s hand movements to ensure the child had the motor abilities to complete the task. Subsequently, the child was asked to point a finger when the experimenter made a first, and to make a fist when the experimenter pointed a finger. The task began with two comprehension check trials, one for each movement, followed by six practice trials. The experimenter praised the child for correct trials. If the child responded incorrectly, the experimenter reminded the child of the rules and repeated the trial. After completing the practice trials, the child was presented with 15 test trials, in a fixed, quasi-random order. During these test trials, the experimenter did not provide feedback. Each trial was scored incorrect (0), initially incorrect, but changed to correct (1), or correct (2), consistent with Kochanska et al.'s (1997) scoring of other inhibitory tasks. Scores were averaged across all trials (0–2). Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater inhibitory control.

Knock/Tap. Knock/Tap (Klenberg et al., 2001) assesses inhibitory control and shifting and consists of two parts. Prior to starting the task, two imitation trials were administered to

ensure the child had the motor abilities to complete the task. During these imitation trials, the child copied how the experimenter knocked or tapped the table. The child was then instructed to knock the table, whenever the experimenter tapped, and to tap the table whenever the experimenter knocked. After two comprehension checks and two practice trials, 15 pseudorandom test trials were administered. In the second part of the task, the instructions changed. The child was instructed to make a side fist when the experimenter knocked, and to knock when the experimenter made a side fist. However, when the experimenter tapped the table, the child was instructed to do nothing. After six practice trials, 15 test trials were administered. During test trials, the experimenter did not provide feedback. Each trial was scored incorrect (0), initially incorrect, but changed to correct (1), or correct (2 consistent with Kochanska et al.'s (1997) scoring of other inhibitory tasks. Scores were averaged across trials (0–2). Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater inhibitory control.

Less is More. Less is More is a motivationally salient symbolic representation task that assesses affective (“hot”) inhibitory control (Carlson et al., 2005). The child chose a preferred treat from two options, white marshmallows and uniformly colored jelly beans. The preferred treats were pre-bagged in transparent bags with some bags containing two treats and others containing five treats. The child was asked if they prefer the bag of two treats or five treats. Children who preferred the two treat bags at the beginning of the trial were excluded. In front of the child were two bowls, one of which had a “naughty monkey” puppet, and the other bowl was the child’s bowl. The child was told that “the monkey wants all the treats for himself.” On each trial, two bags are presented to the child: one bag with five treats and one bag with two treats. The child was instructed to point to a bag among the two bag options presented. The child was

instructed that the bag they point to goes to the monkey's bowl, and that they receive the treats in the other bag (i.e., the bag they did not point to). Each time the child chose a bag, the experimenter put the bag the child chose in the monkey's bowl, and the other bag in the child's bowl. After up to three comprehension check trials with corrective feedback, there were eight test trials in the first trial set. The monkey was then moved to the opposite bowl to avoid a side bias. Then, another comprehension check and eight more trials were administered with the same rules as the first trial set. Responses were scored as: 0 = child points to large treats bag; 1 = child initially points to the large treats bag, then changes to the small treats bag; 2 = child points to the small treats bag, consistent with Kochanska et al.'s (1997) scoring of other inhibitory tasks. Scores were averaged across 16 test trials (0–2). Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater affective inhibitory control.

Peg Tapping. Peg Tapping (Luria et al., 1964) assesses inhibitory control. The child observed sequences of a specific number of pencil taps on a table (either one or two) and was instructed to tap a pencil the opposite number of times of what they observed. The experimenter explained the rules: when the experimenter taps the pencil once and then hands the pencil to the child, the child is to tap the pencil twice. When the experimenter taps the pencil twice, the child is to tap the pencil once. The child received two practice trials and then received 16 test trials in which the experimenter followed a fixed, quasi-random order to tap once or twice. The child was given corrective feedback on the practice trials but not the test trials. Trials were scored correct (1) or incorrect (0). Final scores were averaged across trials, which reflected a proportion of maximum possible score. Higher scores reflected greater inhibitory control.

Shape Stroop. Shape Stroop (Kochanska et al., 2000) assesses children's perceptual inhibitory control. The task assessed the child's ability to identify a picture of a small fruit

embedded within a picture of a different, larger fruit. To verify that the child knew the names of the fruits in the pictures, the child was first presented with three pictures, each containing one large fruit: an apple, banana, or orange. In the first three trials, the child was asked to point to a large fruit (e.g., the large apple). After successfully identifying these three fruits, the child was presented with three new pictures, each containing a small fruit embedded within a different, larger fruit image (e.g., a small banana embedded within a larger apple image). The following three trials, the child was instructed to point to a small fruit (e.g., the small banana). Trials were scored from 0 to 2 (0 = incorrect, 1 = initially incorrect, but changed response to correct, 2 = correct; Kochanska et al., 2000). Scores were averaged across the three small fruit trials (0–2). Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater perceptual inhibitory control.

Simon Says. Simon Says (Strommen, 1973) assesses children's inhibitory control in response to verbal and motor cues. The task involved a series of activation (i.e., "go") and inhibition (i.e., "no-go") trials, in which the child was instructed to inhibit their behavioral response to instructions unless the instructions are accompanied by a verbal cue. The child was presented with a series of instructions to perform simple motor actions (e.g., clap your hands, stomp your feet) and was told to perform the action only if the instructions are preceded by the phrase "Simon Says." The child completed two go practice trials and two no-go practice trials, followed by 20 test trials, including ten go trials and ten no-go trials, presented in a fixed, pseudo-random order. Each no-go trial was scored from 1 to 4 (1 = full commanded movement, 2 = partial movement, 3 = wrong movement, and 4 = no movement), consistent with Carlson and Moses (2001) scoring of a simplified version of Simon Says (Bear/Dragon); scoring was reversed for go trials. Scores were averaged across trials within condition (no-go versus go;

ranged 1–4). Because children could receive a high score on no-go trials by simply not responding, a composite score of children's inhibitory control was computed by multiplying mean scores from ten go trials and ten no-go trials, consistent with Eisenberg et al. (2013). Children who inhibited behavior across all trials thus received a lower score compared to children who correctly inhibited behavior across inhibition (no-go) trials and activated behavior across activation (go) trials. Final scores were converted to a proportion of the maximum possible score. Higher scores reflected greater inhibitory control.

Stop-signal task. The stop-signal task is a widely used experimental procedure to assess the ability to inhibit inappropriate actions (Verbruggen et al., 2019). The Food Finder stop-signal task was adapted from Berger et al. (2013) to be more appropriate for children as young as three years of age with child-friendly stimuli, an engaging storyline, animations, touchscreen, and a progress bar. Children performed a two-alternative forced choice task, but on some trials, they were given a cue (stop signal) to withhold responding. If the stop signal appeared too late after the go stimulus, children were unable to withhold the response. Latency of the stop signal after go stimulus onset (stop-signal delay [SSD]) was manipulated to determine a child's speed of response inhibition.

The task included three blocks that followed the same structure: presentation of go stimuli, practice go trials, presentation of stop signal, mixed practice trials, and test trials. Each trial began with a flickering star in the center of the screen that served as a fixation point. In Block 1, trials included a picture of a green food (e.g., lime) or purple food (e.g., grapes) in the middle of the screen. On the bottom of the screen was a picture of a green goat and purple pig. The child was told to give purple food to the purple pig and green food to the green goat by touching the animal on the screen. The child was told to touch the purple pig when they see

purple food and to touch the green goat when they see green food. The child then completed the practice go trials and experimenters provided praise for correct responses. After completing the practice go trials, the child was shown a cartoon wizard and was told that wizard will try to trick them and turn the food into a car. On stop trials, the food and animals were shown, and after some delay (i.e., SSD) the food and animals were replaced by a car. The child was instructed not to feed cars to the animals and not to touch the screen when they saw a car. The child was instructed to go as fast as they can. The child then completed mixed practice trials, i.e., both go and stop trials. After the mixed practice trials, the child completed the test trials which consisted of 60 trials in each block: 42 go trials and 18 stop trials.

The task used a staircase dynamic-tracking paradigm that adjusted the SSD based on the child's performance on previous stop trials. The algorithm adjusting the SSD attempted to obtain a 50% error rate on stop trials, which helped normalize task difficulty across ages. The SSD was set at 400 ms for the first trial of Block 1 so the task would be relatively easy in the beginning and become more challenging over time. The delay modification after each stop trial was 100 ms during Block 1 and was 50 ms in Blocks 2 and 3. The delay modification was higher in Block 1 than Blocks 2 and 3 to converge upon the 50% error rate more quickly. If the participant successfully inhibited on a stop trial, the delay modification was added to the SSD on the next stop trial to make stopping more difficult. If the participant failed to inhibit on a stop trial or if they responded before the stop signal, the delay modification was subtracted from the SSD on the next stop trial to make stopping easier. The running SSD at the end of each block carried forward to the next block.

The trial stimuli (i.e., food, animals, and cars) were presented for a maximum of 5000 ms or until the child touched the screen. Auditory feedback lasting ~540–700 ms was provided after

every trial. Feedback was a “yippee” sound for all correct trials: correct responses on go trials and successful omissions on stop trials. For correct responses on go trials, animation showed the food moving toward the selected animal. Feedback was a “hmm” sound for all incorrect trials: omission errors on go trials, incorrect categorizations on go trials—i.e., touching the picture of the wrong animal, and commission errors on stop trials.

To reduce habituation, the animals and foods changed in Blocks 2 and 3. In Block 2, the child was told to give orange food to the orange owl and red food to the red rabbit. In Block 3, the child was told to feed blue food to the blue bird and pink food to the pink penguin. The cartoon wizard and cars were kept the same for both Blocks 2 and 3. There were three foods of each color. In Blocks 2 and 3, the children completed the test trials only, for a total of 180 trials (126 go and 54 stop trials). Again, feedback was provided on every trial. Stimuli were presented via E-Prime software (Schneider et al., 2012).

We performed several processing steps to ensure data were high-quality. The length and difficulty of task blocks caused some children to fail to perform the task for some subsets of trials. We attempted to identify these subsets of children and trials to retain as many children and trials in the analyses as possible while eliminating trials that did not tap response inhibition processes and children who had insufficient valid trials. No algorithm will be perfectly accurate in adjudicating valid responding, but the following criteria were adopted to restrict the analysis to trials in which the child appeared to be performing the task as instructed while allowing for temporary lapses. The same criteria were applied to all children.

First, responses that occurred less than 200 ms after go stimulus onset were discarded from analyses because this would be too rapid for the child to have responded deliberately to the target stimulus. We excluded subsets of trials during which the child appeared to be temporarily

deviating from the instructions but later returned to the task. In some cases, children consistently delayed their response to wait for the stop signal, causing the SSD to become so long that it was no longer relevant for task performance. These subsets of trials were identified by sequences of six or more stop trials in which the child responded before the stop signal appeared. For these trial subsets, we kept only those data prior to the first instance of responding before the stop signal (in that sequence of six consecutive stop trials), and we retained trials after the child had three consecutive stop trials in which they did not respond before the stop signal. If the child had a sequence of trials in which they appeared not to be participating (i.e., they did not respond on four or more consecutive go trials), we kept only those trials prior to their first missed go trial in that sequence of consecutive missed go trials. If the child started participating again, as operationalized by three failed stops in a sequence of six stop trials, we retained the subsequent trials.

We excluded children at a given measurement occasion who had insufficient valid trials due to excessive use of the strategy of delaying their response to wait for the stop signal. We set the threshold for insufficient trials due to this strategy as the child having 20% or more of their go trials in which their reaction times was shorter than the running SSD (i.e., the SSD at that point in the task). We also excluded children who did not have any failed stop trials, indicating that they responded infrequently or after a long delay. In addition, we excluded children who intentionally touched the stop signal (thus not following the rules), resulting in an unreasonably quick SSD. We set this threshold to exclude children whose mean SSD was less than 100 ms.

We operationalized response inhibition as the stop-signal reaction time (SSRT). The SSRT was calculated as the median reaction time on correct go trials minus the mean SSD from Blocks 2 and 3. Block 1 was not included in the calculation to allow the algorithm time to

converge upon a 50% error rate on stop trials. Cases were excluded if the SSRT was negative (i.e., the median go reaction time was faster than the mean SSD). Final scores were converted to a proportion of the maximum observed score and were reverse scored. Higher scores reflected greater inhibitory control.

Behavior Rating Inventory of Executive Function (BRIEF). The BRIEF assesses children's executive functioning within the context of their everyday environment. Two versions were used based on the child's age. Parents completed the BRIEF–Preschool Version (BRIEF–P; Gioia et al., 1996) if the child was 3–5 years old or the BRIEF–2 (Gioia et al., 2015) if the child was 6–7 years old. Scores on the Inhibitory Control subscale were used for both questionnaires' versions. Twenty-four items were rated on a 3-point scale (1 = never, 2 = sometimes, 3 = often) in terms of how often, in the last six months, the child's behavior had been a problem. To account for missing responses in the sum score, scores were averaged across items and then multiplied by the number of items. Scores were converted to a proportion of the maximum (POM) possible score. Scores were then reverse scored so that higher scores reflected greater inhibitory control. Mothers' and fathers' ratings on the Inhibitory Control Composite were significantly correlated ($r[152] = .38$, $p < .001$). Age and sex norm-referenced *T*-scores had a mean of 52.37 ($SD = 10.43$).

Children's Behavior Questionnaire (CBQ). The CBQ assesses children's temperament (i.e., reactivity and regulation). Two versions were used based on the rater type. Parents completed the CBQ (Putnam & Rothbart, 2006). Secondary caregivers completed the CBQ–Teacher Short Form (CBQ–TSF, Teglasi et al., 2015). We used scores from the Inhibitory Control scale (CBQ: 13 items; CBQ–TSF: 6 items). Items were rated on 7-point Likert scale (1 = extremely untrue, 2 = quite untrue, 3 = slightly untrue, 4 = neither true nor untrue, 5 = slightly

true, 6 = quite true, 7 = extremely true). Scores were averaged across items. Scores were converted to a proportion of the maximum possible score. Higher scores reflected greater effortful control. Mothers' ratings on the Inhibitory Control scale were associated with ratings by fathers (r[164] = .46, $p < .001$) and secondary caregivers (r[165] = .31, $p < .001$). Fathers' ratings were associated with ratings by secondary caregivers (r[112] = .35, $p < .001$).

Externalizing Problems

Achenbach System of Empirically Based Assessment. The Achenbach System of Empirically Based Assessment (ASEBA) assesses children's emotional and behavioral problems. Items were rated on a 3-point Likert scale according to how well the item described the child (0 = not true, 1 = somewhat or sometimes true, 2 = very true). Multiple versions were used based on the child's age and rater type. Parents completed the Child Behavior Checklist 1.5–5 (CBCL 1.5–5; Achenbach & Rescorla, 2000) if the child was 3–5 years old or the Child Behavior Checklist 6–18 (CBCL 6–18; Achenbach & Rescorla, 2001) if the child was 6–7 years old. Secondary caregivers completed the Caregiver–Teacher Report Form (C–TRF; Achenbach & Rescorla, 2001) if the child was 3–5 years old or the Teacher's Report Form (TRF; Achenbach & Rescorla, 2001) if the child was 6–7 years old. The ASEBA scales are empirically derived, widely used, and have shown strong reliability (internal consistency, test–retest reliability, and interrater reliability) and validity (content, construct, and criterion-related validity) in large and diverse samples in the U.S. (Sattler, 2022).

Items on the CBCL 1.5–5 and C–TRF were categorized into seven syndrome scales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems (CBCL 1.5–5 only), Attention Problems, and Aggressive Behavior. Items on the CBCL 6–18 and TRF were categorized into eight syndrome scales: Anxious/Depressed, Somatic Complaints,

Withdrawn/Depressed, Social Problems, Thought Problems, Attention Problems, Rule Breaking Behavior, and Aggressive Behavior. Subscales were further categorized into two higher-order factors: internalizing and externalizing. Scores on the Externalizing scale were used. The Externalizing scale consisted of the Attention Problems and Aggressive Behavior syndrome scales for the CBCL 1.5–5 (24 items) and C–TRF (34 items) and the Rule-breaking and Aggressive behavior syndrome scales for the CBCL 6–18 (35 items) and TRF (32 items). To account for missing responses in the sum score, scores were averaged across items and then multiplied by the number of items. As with inhibitory control measures, externalizing problem scores were then converted to a proportion of the maximum possible score to put scores from different ASEBA measures onto a metric with the same possible range. Higher scores reflected more externalizing problems. Mothers' ratings on the Externalizing scale were associated with ratings by fathers ($r[178] = .56$, $p < .001$) and secondary caregivers ($r[174] = .46$, $p < .001$). Fathers' ratings were associated with ratings by secondary caregivers ($r[123] = .44$, $p < .001$). Age and sex norm-referenced *T*-scores had a mean of 46.35 ($SD = 9.73$).

Supplementary Appendix S3. Description of Covariates.

Child age, child sex, and family socioeconomic status were used as covariates. Child age was recorded as the age of the child the date of the first lab visit of each timepoint. Child sex was dummy coded (male = 0, female = 1). Socioeconomic status (SES) was calculated as the average of three z-scored indices: income-to-needs ratio, parent educational attainment, and parent occupational prestige. Income-to-needs ratio was computed based on the ratio of the household's income relative to poverty thresholds from the U.S. Census Bureau given the number of adults and children in the home. Parent educational attainment was scored as the highest level of education completed: 1 = less than 7^{th} grade; 2 = junior high school; 3 = partial high school; 4 = high school graduate; 5 = partial college (at least one year) or specialized training; 6 = standard college or university graduation; 7 = graduate professional training (graduate degree). Parent occupational prestige was scored using the Nam-Powers-Boyd occupational scale (Boyd & Nam, 2015). Given the strong, cross-time rank-order stability of SES ($r = .90$, $p < .001$), we interpolated missing SES values at a given time point by carrying a participant's last observation forward.

Supplementary Appendix S4. Collection and Pre-Processing of Electrophysiological Recordings.

Electrophysiological data were collected using an Electrical Geodesic, Inc (EGI) 128-electrode Hydrocel Geodesic Sensor Net with a Net Amps 400 series amplifier. Electrodes were active electrodes composed of silver chloride (Ag-Cl) plated carbon-embedded plastic. The electrode net was soaked in a saline solution before being placed on the child's head. Net Station Acquisition Software 5.4.2 (Electrical Geodesics, Inc., 2018) was used to collect the continuous EEG data. The recording system's precision was .024 μV/bit and had an analog-to-digital conversion rate of 8000 samples per minute. During recording, electrode impedances were adjusted to be at or below 50 kΩ and continuous EEG data were collected at a sampling rate of 1000 Hz. EEG data were collected during a go/no-go task (i.e., Fish/Sharks), which was administered using E-Prime 2.0.10.356 (Schneider et al., 2012). Stimuli were presented on a computer monitor that was located approximately half of a meter in front of the child. Children were instructed to respond using one hand. Auditory feedback was presented at a volume of 75 decibels (± 2 decibels).

Data were pre-processed in Net Station Tools 5.4.3 (Electrical Geodesics, Inc., 2018). Continuous data were band-pass filtered from 0.1 to 30.0 Hz. Correct go and correct no-go trials were selected and segmented into 1200 ms epochs that began 200 ms prior to the presentation of each stimulus. Epochs were then automatically inspected for artifacts, which included identifying and removing "bad" channels. Eye blinks and eye movements were also identified. Channels were marked bad if they contained a voltage shift greater than 200 μV during a given segment length of 80 ms. Eye blinks were classified as a voltage shift greater than 175 μV (max−min) within a 640 ms moving time window for each trial after running an 80 ms moving-average

smoothing algorithm across the entire trial period. Eye movements were classified as a voltage shift greater than 200 μV (max−min) over a 640 ms window (with an 80 ms moving-average smoothing algorithm). Epochs were marked bad if they contained more than 20 bad channels, an eye blink, or an eye movement. Channels were marked bad across all epochs if 20% or greater of the epochs were marked bad. Channels marked bad across all epochs were removed. Removed channels were interpolated based on the waveforms of surrounding electrodes. If a child did not have at least 10, artifact-free trials in each condition after automatic processing, epochs were manually examined for artifacts. After manually identifying and removing artifacts and bad channels, epochs were subjected to the same automatic inspection procedure described above. Epochs were then averaged within participants, and re-referenced to an average reference (i.e., the average of all scalp electrodes). Finally, epochs were baseline corrected by subtracting the average activity from each epoch's 200 ms baseline.

Supplementary Appendix S5. Description of Temporospatial PCA.

To perform the sequential temporospatial PCA (tsPCA), we followed recommendations by Dien and colleagues (Dien, 2010, 2012; Dien & Frishkoff, 2005). First, we conducted temporal PCA using a promax rotation to identify distinct temporal components. The number of components to retain in each of the PCA analyses was determined using a parallel test, which compares eigenvalues to eigenvalues from randomly simulated data (Horn, 1965). Then, we conducted spatial PCA using an infomax rotation to identify distinct spatial components. Finally, we conducted spatial PCA using an infomax rotation on the temporal components identified in the previous temporal PCA to identify distinct temporospatial components. The tsPCA was conducted across all ages (i.e., 36 to 90 months) but separately for each condition (i.e., go versus no-go trials).

The temporospatial waveform, thought to correspond with the N2 component, was selected based on a priori hypotheses about the latency, topography, and morphology of the component. The selected N2 component was characterized by a frontocentral negativity that peaked at 427 ms in the go condition, and 466 ms in the no-go condition. N2 amplitudes used in analysis were extracted from the peak negative channel (electrode 21) at the peak latency of the selected temporospatial component corresponding to the N2 (Scharf et al., 2022). Full details of the selected N2 component for each condition are described below.

Go trials

The temporal PCA retained 17 temporal components, which collectively explained greater than 98 percent of the variance across timepoints in the waveforms. The spatial PCA retained 17 spatial components, which collectively explained greater than 85 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components

retained 8 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 136 temporospatial components (17 temporal components × 8 spatial components). The selected N2 component explained approximately 6.31 percent of the overall variance in the waveform.

No-Go trials

The temporal PCA retained 20 temporal components, which collectively explained greater than 98 percent of the variance across timepoints in the waveforms. The spatial PCA retained 15 spatial components, which collectively explained greater than 85 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 8 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 160 temporospatial components (20 temporal components × 8 spatial components). The selected N2 component explained approximately 7.43 percent of the overall variance in the waveform.

Age-Specific PCAs

As an additional sensitivity analysis, we conducted separate PCA analyses for younger (36–54 months, n = 81 instances) and older children (63–90 months, n = 77 instances) for both conditions. We conducted age-specific PCAs to be consistent with prior research that has shown that children's capacity for self-regulation, including inhibitory control, rapidly increases between ages 3 and 7 (Hosch et al., 2022). Moreover, there is substantial neural development during the same developmental period (Casey et al., 2005), which may produce different PCA structures in younger versus older children. Thus, to account for these developmental changes,

we conducted separate PCAs for two the age groups.

The N2 component appeared similar in younger and older children. In younger children, the selected N2 component was characterized by frontocentral negativity that peaked at 501 ms. In older children, the component displayed the same frontocentral negativity and peaked slightly earlier at 466 ms. Moreover, the tsPCA-identified no-go N2 amplitudes between younger and older children were strongly correlated ($r[472] = .64$, $p < .001$). Thus, we used the N2 amplitudes from the age-combined PCA to help ensure that we extracted the same ERP component across ages. Details of the age-specific PCAs are below.

Ages 36–54 months

Go trials. The temporal PCA retained 17 temporal components, which collectively explained greater than 98 percent of the variance across timepoints in the waveforms. The spatial PCA retained 16 spatial components, which collectively explained greater than 87 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 7 spatial components, which collectively explained greater than 79 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 119 temporospatial components (17 temporal components × 7 spatial components). The selected N2 component explained approximately 7.28 percent of the overall variance in the waveform.

No-Go trials. The temporal PCA retained 19 temporal components, which collectively explained greater than 98 percent of the variance across timepoints in the waveforms. The spatial PCA retained 15 spatial components, which collectively explained greater than 86 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 7 spatial components, which collectively explained greater than 80 percent of the

variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 133 temporospatial components (19 temporal components × 7 spatial components). The selected N2 component explained approximately 7.95 percent of the overall variance in the waveform.

Ages 63–90 months

Go trials. The temporal PCA retained 15 temporal components, which collectively explained greater than 98 percent of the variance across timepoints in the waveforms. The spatial PCA retained 18 spatial components, which collectively explained greater than 87 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 7 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 105 temporospatial components (15 temporal components × 7 spatial components). The selected N2 component explained approximately 5.40 percent of the overall variance in the waveform.

No-Go trials. The temporal PCA retained 17 temporal components, which collectively explained greater than 98 percent of the variance across timepoints in the waveforms. The spatial PCA retained 15 spatial components, which collectively explained greater than 87 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 7 spatial components, which collectively explained greater than 81 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 119 temporospatial components (17 temporal components × 7 spatial components). The selected N2 component explained approximately 8.50 percent of the overall variance in the waveform.

Supplementary Appendix S6. Exploratory Factor Analysis.

We examined scores from the inhibitory control measures in exploratory factor analysis (EFA). We conducted EFA using the efa() function of the lavaan 0.6-14 package (Rosseel, 2012) in R 4.2.0 (R Core Team, 2022). EFA models were fit with full information maximum likelihood (FIML) with robust standard errors to account for nonnormally distributed data. EFA models used geomin for oblique rotation, to account for the covariation among latent factor dimensions of inhibitory control. To leverage all time points of data for all participants, we stacked the data in long form, so that each combination of participant and timepoint uniquely identified each row. We fit separate EFA models with and without controls for age. A one-factor model accounted for 36.2% of the variance. All but one measures' scores (fathers' ratings on the BRIEF) had a significant factor loading. In a two-factor model, the second factor accounted for 12.2% of the variance. Moreover, all measures that had loadings above .40 on the second factor were questionnaire measures, suggesting that the factor that accounted for the most variance after the primary factor was a method factor. Findings remained consistent when controlling for the child's age. Thus, although inhibitory control measures clearly assessed multiple dimensions, a single factor accounted for considerable variance, and accounted for considerably more variance than the second factor. Based on this evidence, the primary factor appeared to reflect a meaningful operationalization of inhibitory control. Thus, given our goals to examine children's inhibitory control development by aggregating scores from multiple methods, we conducted item response modeling with a single factor. Results of the single factor exploratory factor analysis are in Supplementary Tables S4–S5.

Supplementary Appendix S7. Developmental Scaling Approach.

We used developmental scaling to link scores from the different measures across ages onto the same scale. In this way, we could make meaningful comparisons of scores from different measures across ages and estimate accurate trajectories of inhibitory control and externalizing problems. To perform developmental scaling, we used a two-parameter Bayesian longitudinal item response model in a mixed modeling item response theory (IRT) framework. Such a model allows us to simultaneously account for heterotypic continuity using different measures across time and to model children's trajectories. The model linked scores from measures across all ages in the same model, known as concurrent calibration. Concurrent calibration accounts for within-person dependence of scores across time and results in more precise and stable estimates than two-stage calibration in which separate models across age are fit (Kolen & Brennan, 2014; McArdle et al., 2009). We fit a separate model for inhibitory control and externalizing problems (described below).

The two-parameter item response model estimates two parameters: easiness (ξ; the inverse of difficulty) and discrimination (α). The item's easiness parameter is the expected score on an item at a given level of the construct (Bürkner, 2020). The item's discrimination parameter is how strongly the item is associated with the construct. Easiness and discrimination provide information about the functioning and usefulness of each item—and the whole measurement scheme—at a given age.

A two-parameter logistic IRT model takes the following form:

$$P\left(y_{ij} = 1 \middle| \theta_j, \alpha_i, \xi_i\right) = \frac{e^{\alpha_i\left(\theta_j + \xi_i\right)}}{1 + e^{\alpha_i\left(\theta_j + \xi_i\right)}} \quad (1)$$

where y_{ij} is score for person j on item i, theta (θ_j) is the level on the construct for person j, xi (ξ_i) is the easiness parameter for item i, and alpha (α_i) is the discrimination parameter for item i.

Externalizing Problems. For robust estimates of externalizing problems, we fit multidimensional item response models that included items assessing the three primary dimensions of psychopathology: externalizing problems, internalizing problems, and thought-disordered problems. This allowed borrowing information from each dimension in the estimation of the other, for more accurate estimates given considerable covariation between internalizing, externalizing, and thought-disordered problems (Caspi et al., 2014).

In the present study, externalizing problem items were rated on a three-point scale that ranged from $p_{ij} = 0$–2. There were three possible response options (0, 1, 2), so there were two category boundaries: one boundary between 0 and 1, and one boundary between 1 and 2. Because the response options were ordinal, we fit a graded response model, which allows ordinal responses. For externalizing problem items, we used a cumulative response distribution with a logit link. A two-parameter graded response model takes the following general form of Equation (2):

$$P\left(Y_{ij} = y_{ij} \middle| \theta_j\right) = P^*_{y_{ij}}\left(\theta_j\right) - P^*_{y_{ij}+1}\left(\theta_j\right) \tag{2}$$

where:

$$P^*_{y_{ij}}\left(\theta_j\right) = P\left(Y_{ij} \geq y_{ij} \middle| \theta_j, \alpha_i, \xi_{ic}\right) = \frac{e^{\alpha_i\left(\theta_j + \xi_{ic}\right)}}{1 + e^{\alpha_i\left(\theta_j + \xi_{ic}\right)}} \tag{3}$$

where y_{ij} is score for person j on item i, theta (θ_j) is the level on the construct for person j, xi (ξ_{ic}) is the easiness parameter for item i for category c, and alpha (α_i) is the discrimination parameter for item i.

We estimated the item's easiness parameter (ξ_{ic}) with fixed effects for the child's sex, the role of the rater (mother, father, or secondary caregiver), the psychopathology dimension assessed (externalizing, internalizing, or thought-disordered problems), and linear and quadratic

terms for the child's age. The rater role was dummy coded so that the mother rater was the reference group. The model included a random intercept and random slope for each item. The random slopes for each item were age, role, and an age × role interaction. This allowed each item to differ in its change in easiness over time for each rater type. There was also a random intercept and random slope for each person. The random slopes for each person were age, quadratic age, role, age × role interaction, dimension, age × dimension interaction, dimension × role interaction, and an age × dimension × role interaction. This allowed each person to have a unique trajectory for each psychopathology dimension and rater type.

We estimated the item's discrimination parameter (α_i) with fixed effects for the dimension assessed, the role of the rater, and linear and quadratic terms for the child's age. The model included a random intercept and random slope for each item. The random slopes for each item were age, rater role, and an age × role interaction. This allowed each item to differ in its change in discrimination over time, and for each item to show unique changes in discrimination for each rater type.

In a Bayesian model, the final step is to specify prior distributions for all remaining parameters in the model. We kept the default priors used in the brms package (Bürkner, 2017), which uses vague but proper priors. The prior for the intercept of item discrimination was a normal distribution with a mean of 0 and standard deviation of 1. The intercept for item easiness and all standard deviation parameters were given a half *t*-distribution prior with 3 degrees of freedom, mean 0, and scale parameter 2.5.

Inhibitory Control. For robust estimates of inhibitory control, we fit multidimensional item response models that included measures assessing the various dimensions of related but distinct aspects of self-regulation: inhibitory control, attentional control, emotion regulation, and

delay of gratification. This allowed borrowing information from each dimension in the estimation of the other, for more accurate estimates given considerable covariation between aspects of self-regulation (Espy et al., 2011).

Given the numerous measures of self-regulation that were assessed, the many items, and the varying number of items per measure, we used measure-level (POM) scores (rather than item- and trial-level scores) as the "items" in the item-response model. The self-regulation scores were continuous proportion scores that ranged from $p_{ij} = 0$–1. Because some scores were zero or one (especially one), we used a zero-one-inflated beta distribution for the outcome variable (Ospina & Ferrari, 2012). A traditional beta distribution is a continuous probability distribution that does not allow zeros or ones. A zero-one inflated beta distribution is a mixed continuous-discrete probability distribution, which includes a continuous beta distribution (to capture the continuous distribution of proportion scores) and Bernoulli distributions (to capture zeros and ones). A zero-one-inflated beta response distribution takes the following form:

$$f(p_{ij}) = \begin{cases} \pi_{ij} & \text{if } p_{ij} = 0 \\ (1 - \pi_{ij})\gamma_{ij} & \text{if } p_{ij} = 1 \\ (1 - \pi_{ij})(1 - \gamma_{ij})\text{Beta}(a_{ij}, b_{ij}) & \text{if } p_{ij} \in (0, 1) \end{cases} \quad (4)$$

where pi (π_{ij}) is the probability of $p_{ij} = 0$, gamma (γ_{ij}) is the conditional probability of $p_{ij} = 1$ given that $p_{ij} \neq 0$, and a_{ij} and b_{ij} are the shape parameters of the Beta distribution when $p_{ij} \in (0, 1)$.

The next step in the Bayesian hierarchical model is to put distributions on each of the parameters in Equation 4. We estimated both π_{ij} and γ_{ij} using a logistic mixed model with fixed effects for the child's age and child's sex, along with a random intercept for participant.

Nesting:

Level 1: i = item (i.e., measure of self-regulation)

Level 2: j = person

$$\text{logit}(\pi_{ij}) = \beta_{0\pi} + \beta_{1\pi}\text{age}_{ij} + \beta_{2\pi}\text{sex}_j + \beta_{3\pi}\text{dimension}_d + b_{0\pi j} \quad (5)$$

$$\text{logit}(\gamma_{ij}) = \beta_{0\gamma} + \beta_{1\gamma}\text{age}_{ij} + \beta_{2\gamma}\text{sex}_j + \beta_{3\pi}\text{dimension}_d + b_{0\gamma j}$$

The a_{ij} and b_{ij} parameters in the beta distribution are re-written into the mean $\mu_{ij} = \frac{a_{ij}}{a_{ij}+b_{ij}}$ and the variance $v_{ij} = \frac{a_{ij}\times b_{ij}}{(a_{ij}+b_{ij})^2(a_{ij}+b_{ij}+1)}$. The mean μ_{ij} is a percentage that is given the IRT form of Equation 1 and specified in more detail below in Equation 6.

$$\text{logit}\left(\mu_{ij}\right) = e^{\log(\alpha_i)} \times \eta_{ij} \quad (6)$$

$$\eta_{ij} = \theta_j + \xi_i$$

where mu (μ_{ij}) is the inhibitory control score for person j on item i, alpha (α_i) is the discrimination parameter for item i, eta (η_{ij}) is the sum of the person's level on the construct (θ_j) for person j and the item's easiness (ξ_i) for item i.

We estimated eta (η_{ij}) with fixed effects for the child's sex, the self-regulation dimension assessed (inhibitory control, attentional control, emotion regulation, and delay of gratification), and linear and quadratic terms for the child's age. The model included a random intercept and random slope of age for each task. This allowed each task to differ in its change in easiness over time. There was also a random intercept and random slope for each person. The random slopes for each person were age, quadratic age, dimension, and an age × dimension interaction. This allowed each person to have a unique trajectory for each self-regulation dimension.

We estimated the item's discrimination parameter (α_i) with fixed effects for the dimension assessed and linear and quadratic terms for the child's age. The model included a random intercept and random slope of age for each task. This allowed each task to differ in its change in discrimination over time.

In addition, we took the log of the variance and used a linear mixed model on v_{ij} using an intercept-only log-linear mixed model with a population intercept and a random intercept for participant.

$$\log(v_{ij}) = \beta_{0\gamma} + b_{0\gamma j} \quad (5)$$

We kept the default priors used in the brms package (Bürkner, 2017), which uses vague but proper priors. The priors were logistic (mean 0, scale parameter 1) for the intercept of the probability of having a score of 0 or 1 (zero-one inflation; zoi) and the conditional probability of having a score of 1 given the score is either 0 or 1 (conditional one-inflation; coi). The intercept for precision (phi; i.e., 1/variance) and all standard deviation parameters were given a half *t*-distribution prior with 3 degrees of freedom, mean 0, and scale parameter 2.5.

Developmental Scaling. We performed the developmental scaling and estimation of growth curves in the same model. A given child had up to four time points. Thus, a quadratic was the most complex polynomial of nonlinear growth we could estimate for children's trajectories that still allow measurement error. Because of prior work demonstrating that developmental trajectories in inhibitory control (Montroy et al., 2016) and externalizing problems (Petersen et al., 2015) are nonlinear, we modeled children's growth with a quadratic term. We modeled random intercepts and random linear and quadratic slopes to allow each child to differ in their starting point, form of growth, and curvature. Age in years was centered to set the intercepts at age 3, the youngest age in the sample. We included the child's sex (female = 1, male = 0) and, for psychopathology ratings, the rater role as a predictor of the intercepts and slopes. Bayseian growth curve estimates for inhibitory control and externalizing problems are depicted in Supplmentary Figures S5 and S6.

Our model had no missing data in the predictors (age, sex, and rater); missingness was only in the outcome (scores on psychopathology items and self-regulation measures). Mixed models handle missing data in the outcomes. Mixed models provide valid inferences if the data are missing at random or completely at random (Detry & Ma, 2016). Furthermore, our Bayesian hierarchical mixed model also provides valid inference when data are missing at random or completely at random. Because much of our missingness was due to the coronavirus 2019 (COVID-19) pandemic, and we observed limited patterns of systematic missingness as a function of demographics, predictors, or outcomes with small effect sizes, we felt this modeling approach was appropriate. Moreover, researchers have argued against using multiple imputation in longitudinal designs that use mixed models because multiple imputation can lead to unstable estimates (Twisk et al., 2013).

Developmentally scaled factor scores were estimated from the posterior distribution by averaging model-predicted posterior samples across chains and iterations, within combinations of child-by-measurement occasion (inhibitory control) or within combinations of child-by-measurement occasion-by-rater (externalizing problems). This allowed each child to have a different factor score of inhibitory control at each of their measurement occasions, and for each child to have a different factor score of externalizing problems for each rater at each of their measurement occasions.

We fit the Bayesian longitudinal mixed models using the brm() function of the brms package 2.18 (Bürkner, 2017) in R, which uses the RStan 2.21.7 (Stan Development Team, 2020a) interface to Stan 2.21.0 (Stan Development Team, 2020b) for Bayesian modeling. The models included eight chains and 10,000 iterations.

Supplementary Appendix S8. Sensitivity Analyses.

Cluster-Robust Standard Errors

Effects were somewhat attenuated when including cluster-robust standard errors (as opposed to bootstrapping) to account for the nonindependence of the data. After controlling for the child's age and sex, N2 amplitudes were not related to externalizing problems (i.e., total effect; $\beta = 0.09$, $p = .103$). Similarly, there was not a significant association between N2 amplitudes and externalizing problems after accounting for inhibitory control and covariates (i.e., the direct effect, $\beta = 0.07$, $p = .227$). Moreover, there was no significant indirect effect of inhibitory control ($\beta = 0.02$, $p = .150$). However, effect sizes were similar. Nevertheless, bootstrapping is considered the gold-standard approach to evaluate indirect effects (Hayes, 2009; Shrout & Bolger, 2002), so we interpret these findings with caution.

Additional Covariates

In addition to controlling for child age and sex, we conducted sensitivity analyses controlling for rater type (i.e., mother, father, and secondary caregiver), the number of correct no-go trials kept, and the number of bad electrodes. Effects were somewhat attenuated when including additional covariates. N2 amplitudes were not related to externalizing problems (i.e., total effect; $\beta = 0.07$, $p = .127$). Similarly, there was not a significant association between N2 amplitudes and externalizing problems after accounting for inhibitory control and covariates (i.e., the direct effect, $\beta = 0.05$, $p = .294$). The indirect effect of inhibitory control was marginally significant ($\beta = 0.02$, $p = .090$). However, despite somewhat attenuated effects, the overall effect sizes were similar.

Moderated Mediation by Sex

To examine whether there were sex-related differences in the indirect effect, we

estimated a multigroup model comparing boys and girls. We allowed all parameters to freely vary across groups. There were no significant differences in the indirect effect across groups (β = -.008, p = .799). Thus, the indirect effect appeared to operate similarly for both boys and girls.

Inhibitory Control Estimated from Performance-Based Tasks

Results were largely the same using a latent variable of inhibitory control estimated from performance-based (i.e., behavioral) tasks. After controlling for the child's age and sex, the number of no-go trials kept, and the number of bad channels, N2 amplitudes remained associated with externalizing problems (i.e., total effect; β = 0.06, p = .001). The association between the N2 and externalizing problems remained nonsignificant after accounting for inhibitory control and covariates (i.e., direct effect; β = -0.04, p = .365). Additionally, there was a significant indirect effect of inhibitory control (β = 0.10, p = .001).

Inhibitory Control Estimated from Questionnaires

Effects differed slightly using a latent variable of inhibitory control estimated from informant-based measures (i.e., questionnaires). After controlling for the child's age and sex, the number of no-go trials kept, and the number of bad channels, N2 amplitudes remained associated with externalizing problems (i.e., total effect; β = 0.08, p = .025). There was a significant association between N2 amplitudes and externalizing problems after accounting for inhibitory control and covariates (i.e., the direct effect, β = 0.11, p = .014). Additionally, there was a marginally significant indirect effect of inhibitory control (β = -0.03, p = .05). Unexpectedly, questionnaire ratings of inhibitory control were *positively* associated with externalizing problems in the mediation model (despite having a negative bivariate association: r = -.38). It appears that the unexpected sign of the association between inhibitory control and externalizing problems was likely due to controlling for a variable (age) that was strongly associated with both (age and

inhibitory control: r = .60; age and externalizing problems: r = -.45). After removing age as a covariate, questionnaire ratings of inhibitory control were negatively associated with externalizing problems, as expected. The indirect effect held (β = 0.02, p = .020).

Central Positive-Going ERP

To examine the specificity of the effects of the spatial location of the frontocentral N2, we also examined ERP amplitudes from a different spatial location. We extracted ERP amplitudes from the same temporal component as the frontocentral component but a different spatial component. This component peaked at 466 ms and was characterized by a central positivity—corresponding to a likely dipole of the negative-going N2. Results were largely the same using the central positive-going ERP. After controlling for the child's age and sex, the central positive-going ERP was similarly associated with externalizing problems (i.e., total effect; β = 0.10, p = .027). The association between the central positive-going ERP and externalizing problems was similarly nonsignificant after accounting for inhibitory control and covariates (i.e., direct effect; β = 0.07, p = .128). Additionally, there was a significant indirect effect of inhibitory control (β = 0.03, p = .014).

Earlier Frontocentral Positive-Going ERP

To examine the specificity of the effects of the N2, particularly regarding its timing, we also examined ERP amplitudes from an earlier ERP. We extracted ERP amplitudes from the same spatial component as the N2 but an earlier temporal component. This component peaked at 256 ms and was characterized by a frontocentral positivity. Results differed when using the earlier frontocentral positive-going ERP. After controlling for the child's age and sex, the earlier frontocentral positive-going ERP was marginally associated with externalizing problems (i.e., total effect; β = 0.08, p = .098). After accounting for inhibitory control and covariates, the

association between the earlier frontocentral positive-going ERP and externalizing problems was marginally significant (i.e., direct effect; β = 0.09, *p* = .080). However, the indirect effect of inhibitory control was nonsignificant (β = -0.001, *p* = .553).

N2 Amplitudes Extracted from the Peak Negative Channel

Results were largely the same using N2 amplitudes extracted from the peak negative channel (i.e., electrode 21). After controlling for the child's age and sex, N2 amplitudes were associated with externalizing problems (i.e., total effect; β = 0.10, *p* = .027). After accounting for inhibitory control, the association between the N2 and externalizing problems remained nonsignificant (i.e., direct effect; β = 0.07, *p* = .128). Additionally, there was a significant indirect effect of inhibitory control (β = 0.03, *p* = .014).

N2 Amplitudes Extracted from Electrode Fz

Results were largely the same using N2 amplitudes extracted from electrode Fz (i.e., electrode 11). After controlling for the child's age and sex, N2 amplitudes were associated with externalizing problems (i.e., total effect; β = 0.09, *p* = .049). The association between the N2 and externalizing problems remained nonsignificant after accounting for inhibitory control and covariates (i.e., direct effect; β = 0.07, *p* = .149). Additionally, there was a significant indirect effect of inhibitory control (β = 0.02, *p* = .031).

Supplementary Appendix S9. Power Analysis.

We conducted a post-hoc power analysis to estimate the statistical power of detecting an indirect effect given our sample size and observed effect sizes. We conducted the power analysis using the Monte Carlo simulation (Schoemann et al., 2017) with parameters of 1000 replications, 20,000 Monte Carlo draws per replication, and a 95% confidence level (i.e., alpha = .05). With our sample of 147 children, we had power of .83 to detect significance in the observed indirect effect of $\beta = .02$ or greater.

Supplementary Table S1

Number of Participants with Data for Each Number of Timepoints

Number of Timepoints	N2	IC	EXT
1	60	55	56
2	29	38	38
3	12	16	15
4	1	32	33

Note. N2 = N2 amplitudes for correct no-go trials. IC = Inhibitory Control. EXT = Externalizing problems.

Supplementary Table S2

Number of Participants with Data at Each Wave

Wave (in months)	**N2**	**IC**	**EXT**
36	10	37	37
45	31	57	58
54	40	72	72
63	49	77	77
72	16	35	36
81	10	18	18
90	2	11	11

Note. N2 = N2 amplitudes for correct no-go trials. IC = Inhibitory Control. EXT = Externalizing problems.

Supplementary Table S3

Estimates of Reliability for Study Measures

Task	Inter-rater Reliability	Internal Consistency Reliability	Cross-time 9-month stability
Bear Dragon	ICC[2,*k*] = .99	Mean reliability of all possible split halves: .96	*r* = .47
BRIEF Father	n/a	BRIEF–P: α = .89, ω = .89; BRIEF-2: α = .87, ω_h = .89	*r* = .65
BRIEF Mother	n/a	BRIEF–P: α = .90, ω = .90; BRIEF-2: α = .87, ω_h = .89	*r* = .63
CBQ Father	n/a	α = .80, ω_h = .80	*r* = .33
CBQ Mother	n/a	α = .83, ω_h = .83	*r* = .71
CBQ Secondary	n/a	α = .77, ω_h = .78	*r* = .46
Day/Night	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .94	*r* = .46
Fish/Sharks	n/a	Mean reliability of 1,000,000 split halves split halves: go: .92, no-go: .93	*r* = .41
Grass Snow	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .96	*r* = .47
Hand Game	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .93	*r* = .43
Knock Tap	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .95	*r* = .51
Less is More	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .89	*r* = .35
Peg Tapping	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .93	*r* = .52
Shape Stroop	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .84	*r* = .18
Simon Says	ICC[2,*k*] = 1.00	Mean reliability of all possible split halves: .94	*r* = .69
Stop Signal	n/a	n/a	*r* = .36
CBCL Mother	n/a	CBCL 1.5–5: α = .91, ω_{cat} = .96; CBCL 6–18: α = .88, ω_h = .90	*r* = .56
CBCL Father	n/a	CBCL 1.5–5: α = .90, ω_{cat} = .90; CBCL 6–18: α = .84, ω_h = .85	*r* = .65
TRF	n/a	α = .94, ω_{cat} = .98	*r* = .65
N2	n/a	n/a	*r* = .48

Note. BRIEF = Brief Rating Inventory of Executive Function. CBQ = Children's Behavior Questionnaire. CBCL = Child Behavior Checklist. TRF = Teacher Report Form. N2 = N2 amplitudes for correct no-go trials. Inter-rater reliability was assessed after coding pairs resolved discrepancies. Omega hierarchical values were used for the BRIEF and CBQ. Omega categorical values were used for the CBCL and TRF (due to having only three response categories). However, omega categorical for parents' reports on the CBCL 6–18 was unable to be estimated, so we instead report the omega hierarchical estimate. Estimates of inter-rater reliability are after identifying and resolving discrepancies between coders.

Supplementary Table S4

One-factor Model of Inhibitory Control

Indicator	Factor Loading
Bear Dragon	.817*
BRIEF Mother	.138*
BRIEF Father	.110
CBQ Mother	.326*
CBQ Father	.296*
CBQ Secondary	.170*
Day/Night	.710*
Fish/Sharks	.637*
Grass Snow	.846*
Hand Game	.776*
Knock Tap	.876*
Less is More	.573*
Peg Tapping	.840*
Shape Stroop	.447*
Simon Says	.674*
Stop Signal	.435*

Note. BRIEF = Brief Rating Inventory of Executive Function. CBQ = Children's Behavior Questionnaire.

* significant factor loading

Supplementary Table S5

One-factor Model of Inhibitory Control Controlling for Age

Indicator	Factor Loading
Bear Dragon	.643*
BRIEF Mother	.275*
BRIEF Father	.229*
CBQ Mother	.326*
CBQ Father	.258*
CBQ Secondary	.215*
Day/Night	.522*
Fish/Sharks	.397*
Grass Snow	.696*
Hand Game	.665*
Knock Tap	.812*
Less is More	.351*
Peg Tapping	.726*
Shape Stroop	.231*
Simon Says	.289*
Stop Signal	.212*

Note. BRIEF = Brief Rating Inventory of Executive Function. CBQ = Children's Behavior Questionnaire.

* significant factor loading

Supplementary Table S6

Structure of the Study Data in Long Form

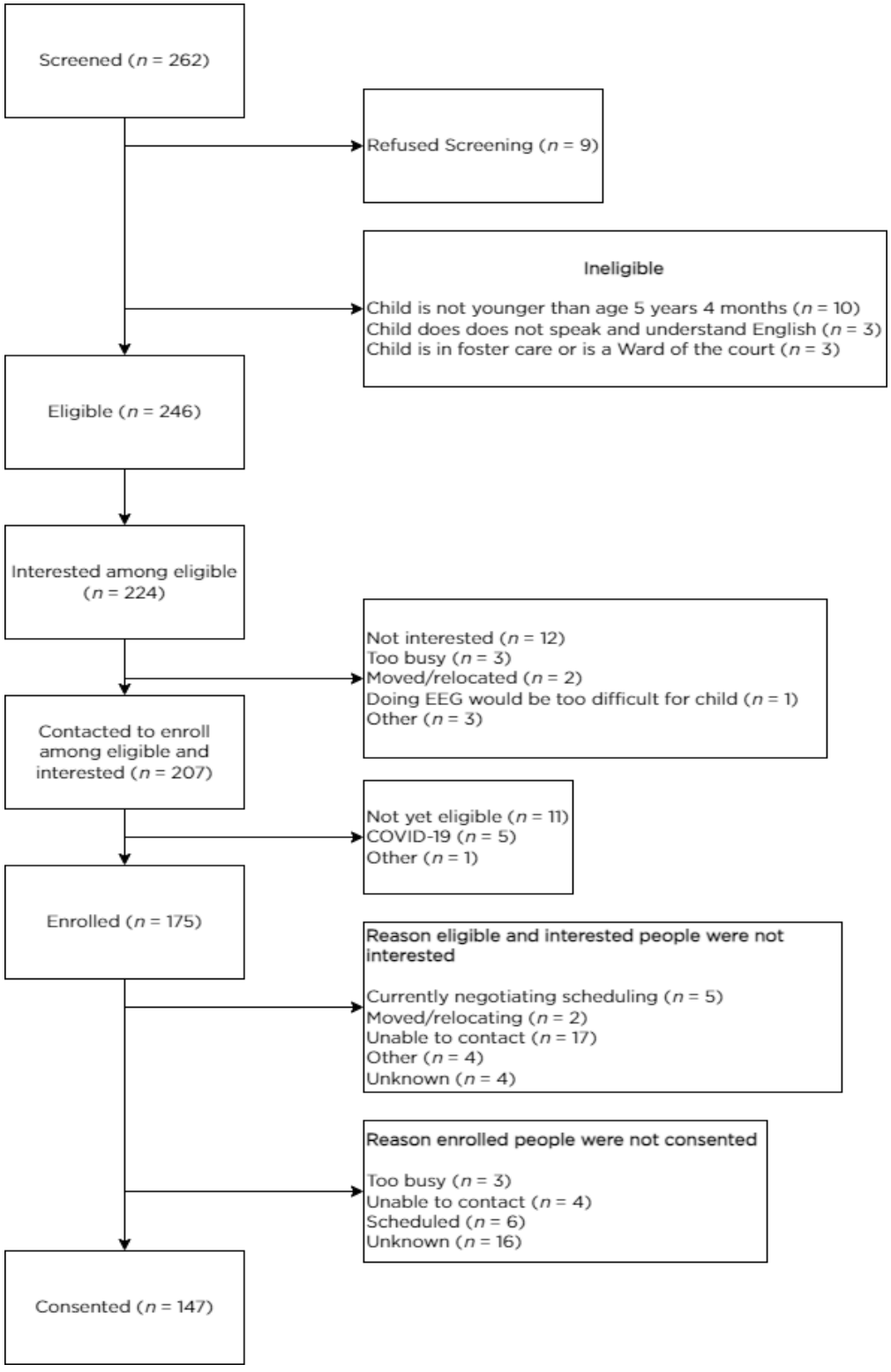
Participant	Wave	Rater	N2	IC	EXT
1001	36	Mother	a	i	q
1001	36	Father	a	i	r
1001	36	Secondary Caregiver	a	i	s
1001	45	Mother	b	j	t
1001	45	Father	b	j	u
1001	45	Secondary Caregiver	b	j	v
1001	54	Mother	c	k	w
1001	54	Father	c	k	x
1001	54	Secondary Caregiver	c	k	y
1001	63	Mother	d	l	z
1001	63	Father	d	l	aa
1001	63	Secondary Caregiver	d	l	ab
1201	45	Mother	e	m	ac
1201	45	Father	e	m	ad
1201	45	Secondary Caregiver	e	m	ae
1201	54	Mother	f	n	af
1201	54	Father	f	n	ag
1201	54	Secondary Caregiver	f	n	ah
1201	63	Mother	g	o	ai
1201	63	Father	g	o	aj
1201	63	Secondary Caregiver	g	o	ak
1201	72	Mother	h	p	al
1201	72	Father	h	p	am
1201	72	Secondary Caregiver	h	p	an
…	…	…	…	…	…

Note. N2 = N2 event-related potential (ERP) amplitudes for correct no-go trials. IC = Inhibitory Control. EXT = Externalizing problems. Wave (i.e., measurement occasion) is in months. The table provides the structure of the data in long form for two example participants. Every row is uniquely identified by the combination of three variables: participant, wave, and rater. The letters in the N2, IC, and EXT columns reflect the common versus unique values in the data structure. The same letter reflects the same value; a different letter reflects a different value. That is, N2 ERP amplitudes and inhibitory control have a unique value for every participant–wave combination. Externalizing problems have a unique value for every participant–wave–rater

combination. The long form structure allowed leveraging all available information without loss of information due to averaging or aggregating scores across raters.

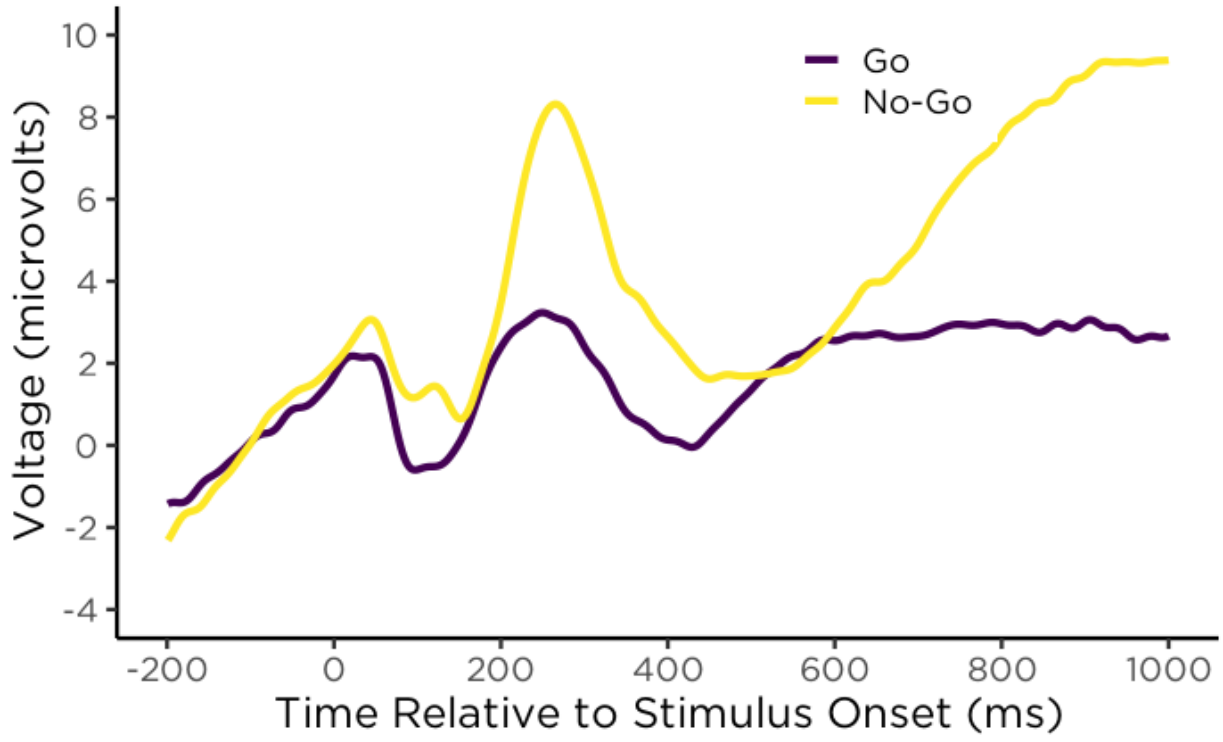
Supplementary Figure S1

Participant Flow Chart



Supplementary Figure S2

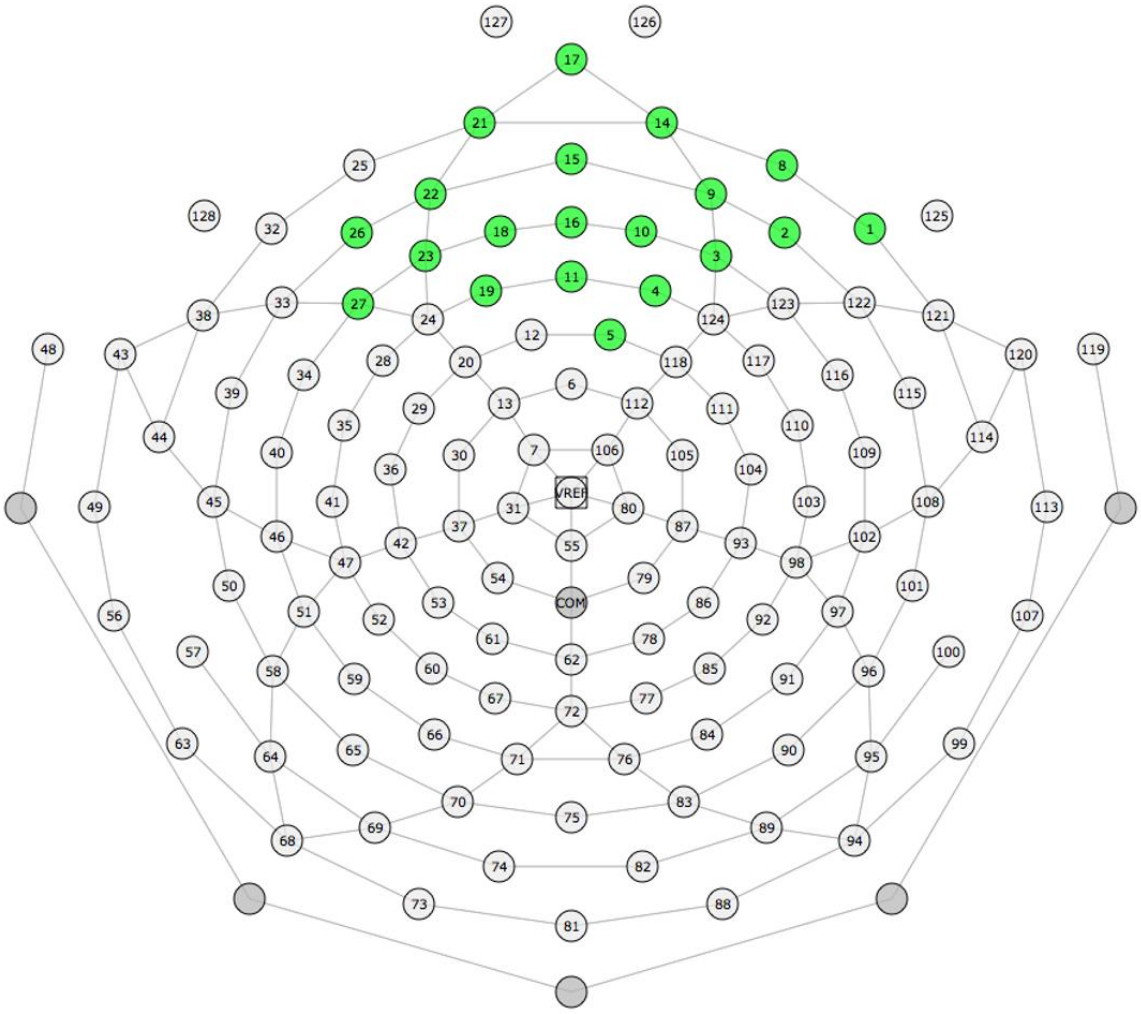
Grand-Averaged Waveform



Note. Grand averaged waveform in both the go, and no-go conditions. For visualization, the waveform was averaged from electrodes with a .5 or greater factor loading on the selected temporospatial component reflecting the N2. The expected latency of the N2 component is 300 to 500 ms post stimulus onset.

Supplementary Figure S3

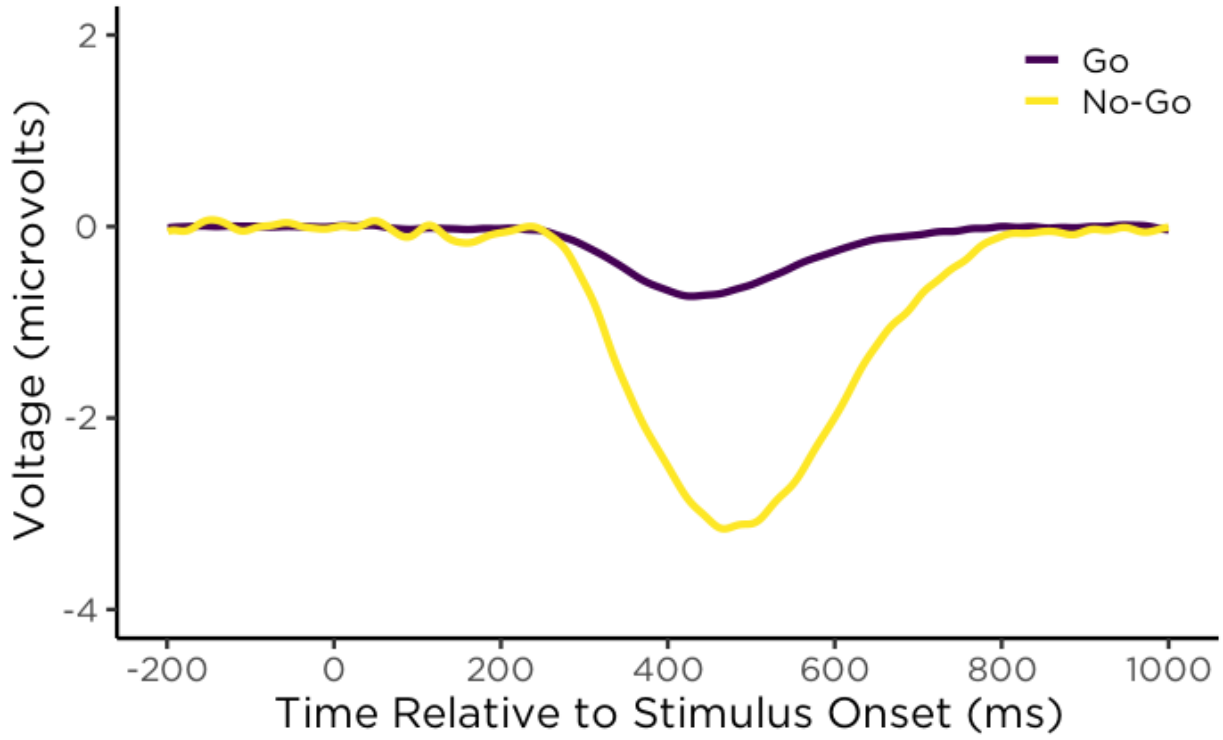
N2 ERP Electrode Cluster



Note. Electrodes highlighted in green correspond to electrodes whose loading on the N2 temporospatial component was .5 or greater.

Supplementary Figure S4

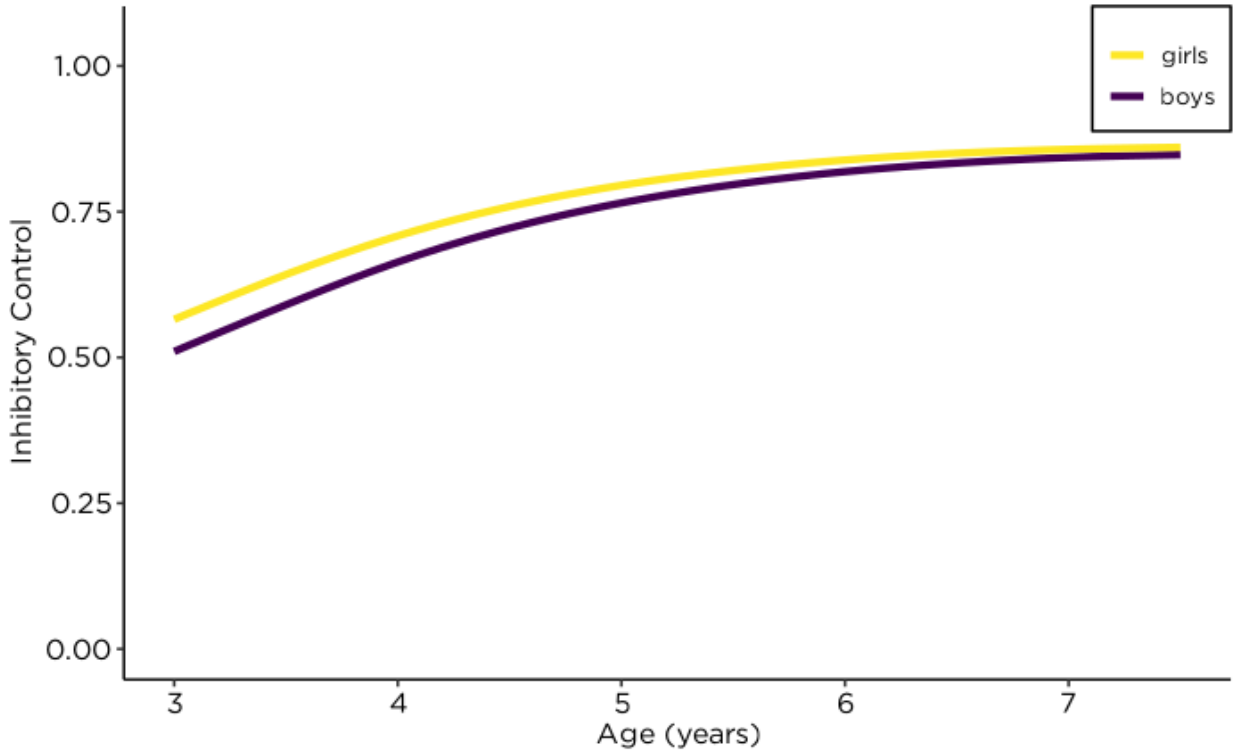
N2 ERP PCA Component Waveform



Note. Temporospatial component reflecting the N2 component in both the go, and no-go conditions. For visualization, the waveform was averaged from electrodes with a .5 or greater factor loading on the selected temporospatial component reflecting the N2. The expected latency of the N2 component is 300 to 500 ms post stimulus onset.

Supplementary Figure S5

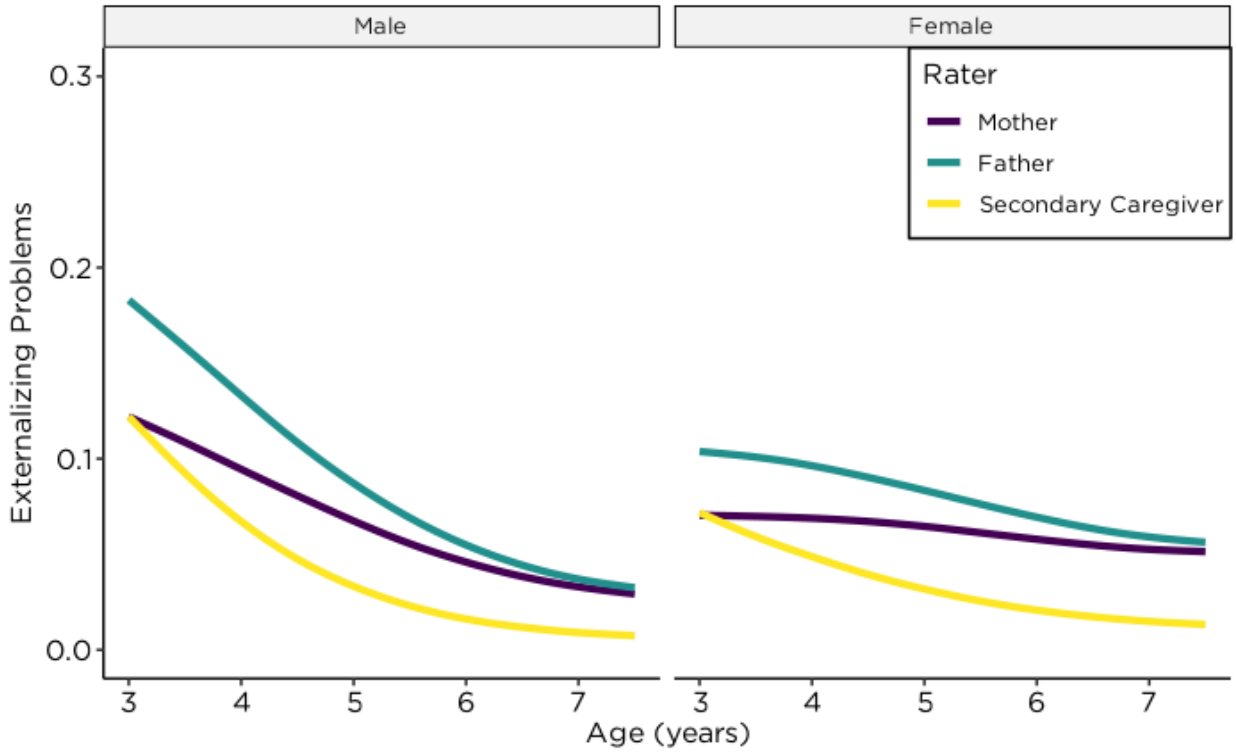
Bayesian Growth Curve Estimates for Inhibitory Control



Note. Bayesian growth curve estimates for inhibitory control separated by child sex.

Supplementary Figure S6

Bayesian Growth Curve Estimates for Externalizing Problems



Note. Bayesian growth curve estimates for externalizing problems separated by rater and child sex.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms and Profiles: An integrated system of multi-informant assessment.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont, Department of Psychiatry.

Berger, A., Alyagon, U., Hadaya, H., Atzaba-Poria, N., & Auerbach, J. G. (2013). Response inhibition in preschoolers at familial risk for attention deficit hyperactivity disorder: A behavioral and electrophysiological stop-signal study. *Child Development*, *84*(5), 1616–1632. https://doi.org/10.1111/cdev.12072

Boyd, M., & Nam, C. B. (2015). *The newest Nam-Powers-Boyd occupational scale: Development and insights*. Southern Demographic Association annual meeting.

Bürkner, P.-C. (2017). brms: An R Package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), Article 1. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2020). Bayesian item response modeling in R with brms and Stan. *ArXiv:1905.09501*. https://doi.org/10.48550/arXiv.1905.09501

Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science*, *16*(8), 609–616. https://doi.org/10.1111/j.1467-9280.2005.01583.x

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain:

What have we learned about cognitive development? *Trends in Cognitive Sciences*, *9*(3), 104–110. https://doi.org/10.1016/j.tics.2005.01.011

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137.

Dien, J. (2010). Evaluating two-step PCA of ERP data with geomin, infomax, oblimin, promax, and varimax rotations. *Psychophysiology*, *47*(1), 170–183. https://doi.org/10.1111/j.1469-8986.2009.00885.x

Dien, J. (2012). Applying principal components analysis to event-related potentials: A tutorial. *Developmental Neuropsychology*, *37*(6), 497–517. https://doi.org/10.1080/87565641.2012.697503

Dien, J., & Frishkoff, G. A. (2005). Introduction to principal components analysis of event-related potentials. In T. C. Handy (Ed.), *Event related potentials: A methods handbook* (pp. 189–207). MIT Press.

Eisenberg, N., Edwards, A., Spinrad, T. L., Sallquist, J., Eggum, N. D., & Reiser, M. (2013). Are effortful and reactive control unique constructs in young children? *Developmental Psychology*, *49*(11), 2082–2094. https://doi.org/10.1037/a0031745

Electrical Geodesics, Inc. (2018). *Net Station 5.4 EEG software* [Computer software].

Espy, K. A., Sheffield, T. D., Wiebe, S. A., Clark, C. A. C., & Moehr, M. (2011). Executive control and dimensions of problem behaviors in preschool children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(1), 33–46. https://doi.org/10.1111/j.1469-7610.2010.02265.x

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition*, *53*(2), 129–153. https://doi.org/10.1016/0010-0277(94)90068-x

Gioia, G. A., Andrwes, K., & Isquith, P. K. (1996). *Behavior Rating Inventory of Executive Function: Preschool version (BRIEF-P) [Measurement Instrument]*. Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function: Second edition (BRIEF2) [Measurement Instrument]*. Odessa, FL: Psychological Assessment Resources.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hosch, A., Oleson, J. J., Harris, J. L., Goeltz, M. T., Neumann, T., LeBeau, B., Hazeltine, E., & Petersen, I. T. (2022). Studying children's growth in self-regulation using changing measures to account for heterotypic continuity: A Bayesian approach to developmental scaling. *Developmental Science*, *25*, e13280. https://doi.org/10.1111/desc.13280

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology*, *20*(1), 407–428. https://doi.org/10.1207/S15326942DN2001_6

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development*, *68*(2),

263–277. https://doi.org/10.2307/1131849

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*(2), 490–507.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220–232. https://doi.org/10.1037/0012-1649.36.2.220

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. Springer Science & Business Media.

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.

Luria, A. R., Pribram, K. H., & Homskaya, E. D. (1964). An experimental analysis of the behavioral disturbance produced by a left frontal arachnoidal endothelioma (meningioma). *Neuropsychologia*, *2*(4), 257–280. https://doi.org/10.1016/0028-3932(64)90034-X

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*(2), 126–149. https://doi.org/10.1037/a0015857

Moeller, J. (2015). A word on standardization in longitudinal studies: Don't. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01389

Montroy, J. J., Bowles, R. P., Skibbe, L. E., McClelland, M. M., & Morrison, F. J. (2016). The development of self-regulation across early childhood. *Developmental Psychology*, *52*(11), 1744–1762. https://doi.org/10.1037/dev0000159

Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis, 56*(6), 1609–1623. https://doi.org/10.1016/j.csda.2011.10.005

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology*, *27*(3), 791–818. https://doi.org/10.1017/S0954579414000789

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, *87*(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09

Sattler, J. M. (2022). *Foundations of behavioral, social, and clinical assessment of children (7th ed.)*. Jerome M. Sattler, Publisher, Inc.

Scharf, F., Widmann, A., Bonmassar, C., & Wetzel, N. (2022). A tutorial on the use of temporal principal component analysis in developmental ERP research – Opportunities and challenges. *Developmental Cognitive Neuroscience*, *54*, 101072. https://doi.org/10.1016/j.dcn.2022.101072

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, *8*(4), 379–386. https://doi.org/10.1177/1948550617715068

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445.

https://doi.org/10.1037/1082-989X.7.4.422

Stan Development Team. (2020b). *RStan: The R interface to Stan. R package version 2.21.0.* http://mc-stan.org/

Stan Development Team. (2020a). *RStan: The R interface to Stan. R package version 2.21.3.* http://mc-stan.org/

Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon says." *Child Development*, *44*(4), 849–853. https://doi.org/10.2307/1127737

Teglasi, H., Schussler, L., Gifford, K., Annotti, L. A., Sanders, C., & Liu, H. (2015). Child behavior questionnaire–short form for teachers: Informant correspondences and divergences. *Assessment*, *22*(6), 730–748. https://doi.org/10.1177/1073191114562828

Twisk, J., de Boer, M., de Vente, W., & Heymans, M. (2013). Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *Journal of Clinical Epidemiology*, *66*(9), 1022–1028. https://doi.org/10.1016/j.jclinepi.2013.03.017

Verbruggen, F., Aron, A. R., Band, G. P., Beste, C., Bissett, P. G., Brockett, A. T., Brown, J. W., Chamberlain, S. R., Chambers, C. D., Colonius, H., Colzato, L. S., Corneil, B. D., Coxon, J. P., Dupuis, A., Eagle, D. M., Garavan, H., Greenhouse, I., Heathcote, A., Huster, R. J., … Boehler, C. N. (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *ELife*, *8*, e46323. https://doi.org/10.7554/eLife.46323

Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development*, *83*(4), 1245–1261. https://doi.org/10.1111/j.1467-8624.2012.01765.x