AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

© 2021 American Psychological Association
ISSN: 0012-1649

2021, Vol. 57, No. 11, 1755–1771
https://doi.org/10.1037/dev0001025

Heterotypic Continuity of Inhibitory Control in Early Childhood: Evidence From Four Widely Used Measures

Isaac T. Petersen[1], John E. Bates[2], Maureen E. McQuillan[3], Caroline P. Hoyniak[4], Angela D. Staples[5], Kathleen M. Rudasill[6], Dennis L. Molfese[7], and Victoria J. Molfese[8]

[1] Department of Psychological and Brain Sciences, University of Iowa
[2] Department of Psychological and Brain Sciences, Indiana University
[3] Department of Pediatrics, Indiana University School of Medicine
[4] Department of Psychiatry, Washington University School of Medicine
[5] Psychology Department, Eastern Michigan University
[6] School of Education, Virginia Commonwealth University
[7] Department of Psychology, University of Nebraska
[8] Department of Child, Youth, and Family Studies, University of Nebraska

Inhibitory control has been widely studied in association with social and academic adjustment. However, prior studies have generally overlooked the potential heterotypic continuity of inhibitory control and how this could affect assessment and understanding of its development. In the present study, we systematically considered heterotypic continuity in four well-established measures of inhibitory control, testing two competing hypotheses: (a) the manifestation of inhibitory control coheres within and across time in consistent, relatively simple ways, consistent with homotypic continuity. Alternatively, (b) with developmental growth, inhibitory control manifests in more complex ways with changes across development, consistent with heterotypic continuity. We also explored differences in inhibitory control as a function of the child's sex, language ability, and the family's socioeconomic status. Children (N = 513) were studied longitudinally at 30, 36, and 42 months of age. Changes in the patterns of associations within and among inhibitory control measures across ages suggest that the measures' meanings change with age, the construct manifests differently across development, and, therefore, that the construct shows heterotypic continuity. We argue that the heterotypic continuity of inhibitory control motivates the use of different combinations of inhibitory control indexes at different points in development in future research to improve validity. Confirmatory factors and growth curves also suggest that individual differences in inhibitory control endure, with convergence among inhibitory control measures by 36 months of age.

Keywords: construct validity invariance, go/no-go, heterotypic continuity, inhibitory control, longitudinal

Supplemental materials: https://doi.org/10.1037/dev0001025.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Inhibitory control, "the ability to inhibit responses to irrelevant stimuli while pursuing a cognitively represented goal" (Carlson & Moses, 2001, p. 1033), is a key construct associated with important outcomes, both academic (Allan et al., 2014) and behavioral (Lipszyc & Schachar, 2010; Wright et al., 2014). Despite the developmental importance of inhibitory control and the extensive research on the construct, the measurement of inhibitory control still presents challenges because facets of the construct appear to develop at different ages (Petersen et al., 2016). Increased precision in the measurement of inhibitory control could facilitate an improved understanding of its developmental course and its role in predicting negative developmental consequences, such as poor academic performance and behavior problems (Blair et al., 2005).

Measurement of Inhibitory Control Across Development

Numerous behavioral tasks have been used to assess individual differences in inhibitory control across childhood. For example, Carlson (2005) tested a total of 24 inhibitory control tasks across

Isaac T. Petersen https://orcid.org/0000-0003-3072-6673
Maureen E. McQuillan https://orcid.org/0000-0002-0109-8500
Caroline P. Hoyniak https://orcid.org/0000-0003-4151-0376
Kathleen M. Rudasill https://orcid.org/0000-0002-9415-8575

This study was not preregistered. A data dictionary of the analysis variables is published at: https://osf.io/a52j4 (Petersen, Bates, et al., 2021). We have no conflicts of interest to disclose. The Toddler Development Study was funded by Indiana University and Grant HD073202 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (National Institute of Child Health and Human Development). Isaac T. Petersen was funded by Grant HD098235 from National Institute of Child Health and Human Development.

Correspondence concerning this article should be addressed to Isaac T. Petersen, Department of Psychological and Brain Sciences, University of Iowa, 175 Psychological and Brain Sciences Building, Iowa City, IA 52242, United States. Email: isaac-t-petersen@uiowa.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ages 2–6 years, all of which required children to inhibit a prepotent response in favor of producing a correct response. These inhibitory control tasks have often been aggregated to form a unified inhibitory control composite to use across ages (e.g., Carlson & Moses, 2001) or to form separate inhibitory control composites to use at different time points, such as during toddlerhood, preschool, and school-age periods (Carlson, 2005; Kochanska et al., 1997). Based on their theoretical similarities, these tasks would be expected to consistently converge and reflect inhibitory control at each age, but empirical tests of that notion have rarely been reported. For example, Carlson and Moses (2001) reported that correlations among ten inhibitory control tasks ranged from r = .04 to .49 (M_r = .28), but they did not analyze *changes* in the longitudinal associations between or within tasks.

Although inhibitory control measures might be expected to correlate in consistent ways across ages, it is also possible that inhibitory control changes in its manifestation across development, with observed changes in the associations among inhibition control measures. The meaning of a measure is defined as the processes that a measure assesses, and a measure's meaning is compared with the construct (or concept) of interest to evaluate the construct validity of the measure (Cronbach & Meehl, 1955). A measure's meaning is inferred from both theoretical and empirical considerations. Theoretical considerations include the processes that the measure likely assesses or was designed to assess, based on theory. Empirical considerations include the degree of change (or stability) of its scores across ages (as compared with what would be expected based on theory) and its association with other measures, that is, its convergent and discriminant validity. Convergent and discriminant validity help to inform what the measure means in association with the construct of interest (Campbell & Fiske, 1959). Changes in the associations among a set of measures across time can provide evidence that one or more of the measures changed in meaning and therefore changed in degree of construct validity.

A given measure intended to assess inhibitory control, such as a child's ability to avoid saying "night" when shown a picture of the night sky, may be meaningfully linked to the construct of inhibitory control during only specific developmental periods, such as the preschool years in this example. However, this task may not be meaningfully linked to inhibitory control in infancy when the child may refrain from saying "night" simply because they have not developed the expressive language skills to do so, and not in the elementary years, because almost all children can easily do this. Early in development, inhibitory control is effortful. However, with practice and development, these effortful processes may become more efficient through the recruitment and integration of different cognitive and neural systems (Marek et al., 2015). In this case, the child's ability to correctly inhibit saying "night" may be correlated with other inhibitory control measures at preschool age but not at other ages, and this change in the associations among measures would be crucial for understanding the manifestation and development of inhibitory control.

Performance on behavioral measures of inhibitory control develops across time. The age at which performance reaches adult levels depends largely on task complexity and difficulty. Performance on some basic inhibitory control tasks reaches adult levels in early childhood, such as on the A-not-B task, in which participants repeatedly find an object in box A, and then are asked to find the object after they see it being moved from box A to box B, and tasks involving basic spatial conflict (e.g., pushing a button on the contralateral side of a target stimulus), spatial reversal (i.e., tasks with similar procedures to the A-not-B task, but the object is moved to box B when out of sight of the child), or reverse categorization (e.g., sorting big blocks into a little bucket and little blocks into a big bucket). By contrast, performance on tasks requiring the integration of multiple executive functions (e.g., go/no-go tasks) continues to improve through adolescence (for a review, see Garon et al., 2008). For instance, accuracy on the go/no-go task, which involves the integration of inhibitory control and working memory to actively respond to a subset of "go" stimuli and inhibit a response to a subset of "no-go" stimuli, has been shown to improve from adolescence to adulthood (Eigsti et al., 2006).

In addition, different facets of inhibitory control have been identified and shown to develop at different ages (Petersen et al., 2016). Perceptual inhibition involves the inhibition of automatic responses to perceptual information, such as in the Shape Stroop task (Kochanska et al., 1997) in which children are instructed to point to a smaller shape while inhibiting a prepotent response to point to the larger shape within which smaller shapes are embedded. Performance inhibition involves the inhibition of a behavioral response to a cue, such as in the Bear/Dragon task (Kochanska et al., 1996) in which children are instructed to respond to prompts from a Bear puppet (i.e., activating on go trials) and inhibit responses to prompts from a Dragon puppet (i.e., inhibiting on no-go trials). Association inhibition involves the inhibition of a dominant, prepotent response to generate a competing response, such as in the Grass/Snow task (Carlson & Moses, 2001) in which children are instructed to point to a white square (rather than the green square) upon hearing the word "grass," inhibiting their prepotent response to point to the green square owing to the strong word–color association for that pairing. Motivational inhibition involves the inhibition of a motivational, affective, or "hot" process, such as in the Snack Delay task (Kochanska et al., 2000) in which children are instructed to wait to eat a snack placed in front of them. Perceptual inhibition has been shown to develop earlier than performance inhibition, association inhibition, and motivational inhibition (Petersen et al., 2016). These developmental differences among the different facets of inhibitory control allow for the possibility that inhibitory control shows heterotypic continuity. This would mean that individual differences in inhibitory control endure, but its specific manifestations change with development. To date, the necessary analyses have not been conducted to evaluate the nature of the associations among various inhibitory control measures over a period of development.

Heterotypic Continuity

Heterotypic continuity refers to the persistence of an underlying construct or process despite behavioral manifestations that change over the course of development (e.g., Caspi & Shiner, 2006; Cicchetti & Rogosch, 2002). An example of a construct that shows heterotypic continuity is externalizing behavior, which reflects aggression, impulsivity, and other "problems that mainly involve conflicts with other people and with their expectations for the child" (Achenbach & Rescorla, 2001, p. 24). Individual differences in externalizing behavior are quite stable across development (Olweus, 1979), but the particular manifestations of the disposition

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

to such behavior change with age (Patterson, 1993). In early childhood, externalizing behavior is often expressed overtly, such as with defiance and physical aggression, but in adolescence, externalizing behavior is often expressed covertly, such as with indirect or relational forms of aggression, rule-breaking, or illicit drug use (Miller et al., 2009). The specific externalizing behaviors change across development, but the essence of the construct endures, demonstrating heterotypic continuity. Heterotypic continuity can be contrasted with homotypic continuity, when the manifestation of a construct remains stable across development and can thus be measured the same way across time, for example, physical growth measured in height and weight.

To date, no studies have fully tested whether inhibitory control shows heterotypic continuity. This is surprising because inhibitory control is considered an underlying phenotype of externalizing psychopathology (Young et al., 2009), which is known to show heterotypic continuity. To test whether inhibitory control shows heterotypic continuity, it would be helpful to use a set of developmentally appropriate inhibitory control measures repeatedly in a longitudinal design to examine patterns of intra- and intermeasure associations, thus testing whether inhibitory control persists in some form while also changing in its behavioral manifestation across time. Using this approach could advance theoretical understanding of how inhibitory control develops and how individual differences in its development relate to children's adjustment. At the same time, findings from this approach could also advance methodology by showing which measures maximize construct validity at various ages, which would allow for better developmental inferences.

If inhibitory control does display heterotypic continuity, this would pose a conceptual and methodological challenge to developmental research. Longitudinal studies frequently use the same measure across ages, which has statistical and measurement advantages. However, if the aim is to describe growth across time, it is important that the measure validly assesses the same construct across the time period of the study. This is clearly possible for measures of physical growth. However, if the construct of interest shows heterotypic continuity, differing in its manifestations across development, different measures may be needed to accurately index that construct across time. Our aim is to more accurately index the developmental construct of inhibitory control in early childhood. Accuracy of indexing can be increased by accounting for heterotypic continuity when studying growth in inhibitory control (Petersen et al., 2020). If inhibitory control changes in manifestation across development and the selected measures in any given study do not align with these changes, the measures will lack construct validity invariance, which may lead to inaccurate inferences about development of inhibitory control. Understanding the manifestation of inhibitory control at different ages will lead to better understanding and measurement of how it develops.

Constructs similar to inhibitory control have been previously considered in ways relevant to the notion of heterotypic continuity. For instance, Chang and colleagues (2015) examined indexes of emotional and behavioral control, which are part of a broader self-regulation construct that also encompasses inhibitory control. The authors found that negative emotionality at 18 months predicted more oppositionality and aggression at 24 months, which predicted less frustration tolerance at 42 months, which predicted poorer interpersonal regulation at 60 months. This hints at the notion of heterotypic continuity, because early negative emotionality predicted later constructs conceptually associated with negative emotionality as well as constructs that more broadly reflect the adaptive implications of negative emotionality. However, it does not convincingly demonstrate heterotypic continuity, because the study did not consider associations among the measures at each age. The question remains: Do individual differences in the construct of inhibitory control show both (a) coherence in the nature of adaption across development and (b) change across time in the construct's manifestations?

The Present Study

We used four well-established measures of inhibitory control to determine whether the manifestation of the construct coheres within and across time in (a) relatively simple, consistent ways or in (b) more complex ways reflecting changes in the manifestation of inhibitory control across time. Evidence of stable manifestation across development would be consistent with homotypic continuity. Evidence of more complex, changing coherence of manifestation would be consistent with heterotypic continuity. To determine whether the manifestation of inhibitory control changes across time, we examined intra- and intermeasure associations across a year of early childhood at 6-month intervals because early childhood is characterized by substantial improvements in inhibitory control (Goswami, 2011) that are supported by neural development in the prefrontal cortex (Diamond, 2002).

Examining changes in associations within and among measures across time is consistent with prior research on other constructs that show heterotypic continuity, such as temperamental emotionality (Durbin et al., 2007). If the pattern of intra- and intermeasure correlations changes across time, this would reflect changes in convergent and discriminant validity and changes in the meaning of measures. If the measures differ in meaning at different ages, that is, differ in their connection to the construct of inhibitory control, this could suggest that (a) inhibitory control differs in its manifestation across development, and (b) the broad concept of inhibitory control may be best assessed via different measures at different time points. In addition to normative development of inhibitory control, there are also enduring individual differences in inhibitory control, which are often reflected as rank-order stability (i.e., children who show advanced inhibitory control relative to their peers tend to remain relatively advanced; Eigsti et al., 2006). The empirically informed meanings of measures at different ages can therefore be interpreted in terms of rank-order stability of indexes and the pattern of correlations among them. To our knowledge, this is the first study to examine whether inhibitory control shows homotypic or heterotypic continuity by examining stability and changes in meaning of four widely used measures of inhibitory control.

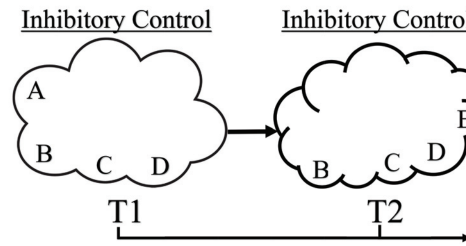
We examined developmental changes in the meaning of inhibitory control measures using reports from multiple informants on the widely used Inhibitory Control scale of the Children's Behavior Questionnaire–Short Form (CBQ; Putnam & Rothbart, 2006) as well as three lab tasks: the Bird/Alligator task (a variant of the Bear/Dragon task; Kochanska et al., 1996), the Shape Stroop task (Kochanska et al., 1997), and the Grass/Snow task (Carlson & Moses, 2001). Meta-analytic evidence suggests that all four of these measures are useful for specifying individual differences (i.e., mean proportion accuracy scores are between .2 and .8) during the age span of the present study (2½ to 3½ years of age; Petersen et al., 2016).

Inhibitory control is operationally defined not only with lab tasks but also with ratings by adults who know the child well. Ratings of children's inhibitory control on the CBQ (Putnam & Rothbart, 2006) assess parent and childcare providers' observations of a child's typical behavior in a broad range of real-world situations, whereas performance-based measures tend to assess a child's optimal performance in specific tasks reflecting narrow abilities (Acar et al., 2019; Toplak et al., 2013). Both of these information sources are relevant for the construct of inhibitory control. CBQ ratings of inhibitory control (CBQ–IC) are useful because they offer ecological validity (Barkley, 2012) as well as long-term utility across at least four years of development (Petersen et al., 2016). The ratings also enable a multimethod approach to reduce method bias, which would not be possible if using only behavioral tasks.

The selected lab tasks have been shown to have overlapping developmental utility during this particular period of development (age 2½ to 3½), with Shape Stroop also being useful at earlier ages, and Grass/Snow also being useful at later ages (Petersen et al., 2016). Thus, this selected set of measures is not only well suited for this particular period of development but is also connected to earlier and later points in development. Moreover, these measures are widely used in early childhood and were extensively pilot tested for the present study to ensure that young children (i.e., aged 30–42 months) can complete the tasks. In each task, children have to hold a rule in mind, respond according to the rule, and inhibit a dominant response (Garon et al., 2008). We collected each measure at 30, 36, and 42 months of age, and we expected that the four measures would either converge consistently at each time point, consistent with homotypic continuity, or change in meaning over time, consistent with heterotypic continuity. If the latter pattern were observed, we would expect to see changes in the intrameasure and intermeasure associations across time, as depicted in Figure 1.

Figure 1

Example Depicting the Content (i.e., Facets) of a Construct (Inhibitory Control) at Two Time Points



Note. The construct is thought to change in its behavioral manifestation across time (i.e., shows heterotypic continuity) but retains an enduring essence, as indicated by the arrow connecting the two clouds, such as might be found in a cross-time correlation coefficient. Different measures may assess different content facets. The construct is thought to include different content across time. For instance, the figure visualizes the construct as including content A, B, C, and D at the first time point (T1), whereas the construct includes content B, C, D, and E at a later time point (T2). The age-differing content (A and E) change in meaning with respect to the construct across time. For instance, content A reflects the construct at T1 but not at T2. The example illustrates that if a given content changes in meaning in connection with a construct across time, the construct that encompasses that content could change in its manifestation across time (i.e., heterotypic continuity). For instance, if performance inhibition (e.g., content E) changes over time in its meaning with respect to the construct of inhibitory control, it would provide evidence that inhibitory control shows heterotypic continuity.

Method

Participants

Children ($N = 534$; 46% girls) were recruited from the Bloomington, IN, and Lincoln, NE, areas to participate in a study that included assessments conducted within two weeks of the ages 30, 36, and 42 months. Among the primary caregivers (96% mothers), 88% were non-Hispanic White, 4% were Hispanic, 3% were African American, 2% were Asian American, 1% were of mixed race, < 1% were American Indian, and 1% were of "other" ethnicity. Primary caregivers ranged from 19–53 years of age ($M = 32.98$, $SD = 4.93$). The Hollingshead index of social class (Hollingshead, 1975) was used as the measure of socioeconomic status (SES). Scores on the Hollingshead index ranged from 12.5 to 66 ($M = 48.18$, $SD = 13.12$), suggesting a sample with some variation in SES, but with a solid middle-class core. Parents' educational attainment included 8th grade or less (< 1% of the sample), some high school (1%), GED (< 1%), high school diploma (4%), some college (13%), college degree (49%), master's degree (21%), and doctoral degree (11%). Primary caregivers' marital status included single (8% of the sample), married (87%), separated (1%), divorced (3%), and remarried (1%).

Some data were missing because of children's inability or refusal to play the lab tasks, families moving, or families being unable to be contacted. To ensure that our inferences were informed by objective, performance-based measures, children were included in the analyses for the present report if they had scores for the Bird/Alligator, Grass/Snow, or Shape Stroop inhibitory control tasks at one or more ages, resulting in a final sample of 513 children (232 girls, 45%). Participants were recruited using a database of county birth records, community outreach efforts (e.g., the local Head Start agency and the Housing Authority), and advertisements (e.g., postcards). Extent of missingness, tests of systematic missingness, and descriptions of missing data handling are in Supplementary Materials S1 in the online supplemental materials. We observed some systematic missingness. Inhibitory control scores were more likely to be missing for children whose primary caregiver was Hispanic or African American and children who were from lower SES families. The Institutional Review Boards at Indiana University and the University of Nebraska approved all procedures for the study, entitled, "Toddler Development Study" (protocol #: 0811000120). Trained research assistants obtained informed consent in person from the legal guardian(s) of all child participants prior to their participation.

Measures

A data dictionary of the analysis variables (not all study variables) is published at: https://osf.io/a52j4. Descriptive statistics for the analysis variables are provided in Table S1 in the online

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

supplemental materials. Descriptive statistics of language ability and SES among those who have inhibitory control scores (separated by measure) are provided in Table S2 in the online supplemental materials. Percent of those with scores (out of those who had a laboratory visit), and the sex distribution by measure are provided in Table S3 in the online supplemental materials.

Inhibitory Control

We used four inhibitory control measures at each age. Interrater reliability was strong, as reported in Table S4 in the online supplemental materials. For the three lab tasks, scores were averaged across coders.

Bird/Alligator

In Bird/Alligator, a go/no-go task (adapted from the Bear/Dragon task; Kochanska et al., 1996), the child is instructed to follow directions from a bird puppet, but ignore directions from an alligator puppet. The children completed several practice trials and were then presented with 12 test trials, including six go (i.e., bird) trials and six no-go (i.e., alligator) trials in pseudorandom order with no more than three trials in a row of the same type (go or no-go). After six trials, children were reminded of the rules. Each no-go trial was scored from 0 to 3 (0 = full commanded movement, 1 = partial movement, 2 = wrong movement, and 3 = no movement), consistent with Carlson and Moses (2001). Scoring was reversed for go trials. The final Bird/Alligator go and no-go scores were the child's average scores on all the go and all the no-go trials, respectively (0–3).

Shape Stroop

In Shape Stroop (Kochanska et al., 1997), the child is instructed to point to pictures of small fruit embedded within pictures of different, larger fruit. The child was presented with three pictures, each containing a small fruit embedded within a larger fruit. In three of the trials, the child was asked to point to a large fruit (e.g., the large banana). After the three large fruit trials, the child was asked to point to a small fruit (e.g., the small apple) in three more trials. Each small fruit trial was scored from 0 to 2 (0 = incorrect, 1 = initially incorrect, but changed response to correct, 2 = correct), consistent with Kochanska et al. (2000). The final Shape Stroop score was the average score on the small fruit trials (0–2).

Grass/Snow

In Grass/Snow (Carlson & Moses, 2001), the child is instructed to touch a white square when they hear the word "grass" and a green square when they hear the word "snow." Following several practice trials, the child was presented with 12 trials, six of each word in a fixed, quasi-random order, and each trial is scored either correct (1) or incorrect (0), consistent with Carlson and Moses (2001).

Children's Behavior Questionnaire

Primary caregivers and, as applicable, their parenting partner and a secondary caregiver (e.g., daycare teacher or babysitter) rated the child's inhibitory control on the Children's Behavior Questionnaire–Short Form (CBQ; Putnam & Rothbart, 2006). We used the Inhibitory Control subscale of the CBQ (CBQ–IC), which includes six items (e.g., "Can easily stop an activity when s/he is told 'no.'") rated on a Likert-type scale, ranging from 1 = *extremely untrue of your child* to 7 = *extremely true of your child*. The six items assess the child's waiting, preparation and planning, stopping of ongoing activities, sitting still, following instructions, and approaching dangerous places slowly and cautiously. The Inhibitory Control scale had an internal consistency of $\alpha = .64$ for primary caregivers, .64 for parenting partners, and .77 for secondary caregivers in the present study. Primary caregiver ratings numbered 488 at 30 months, 409 at 36 months, and 393 at 42 months. Parenting partner ratings numbered 209 at 30 months, 150 at 36 months, and 136 at 42 months. Secondary caregiver ratings numbered 232 at 30 months, 205 at 36 months, and 213 at 42 months. Correlations of primary caregivers' ratings with parenting partners' and secondary caregivers' ratings were $r = .32$ and $r = .25$ ($ps < .001$), respectively. The correlation between parenting partners' and secondary caregivers' ratings was $r = .06$ ($p = .294$). Of 1,539 cases (513 children $\times$ 3 measurement occasions) in the final sample, 17% had three raters, 38% had two raters, 32% had one rater, and 13% had no raters. We averaged a child's score across parents' ratings (i.e., primary caregiver and their parenting partner, as applicable), resulting in 499 scores at 30 months, 417 scores at 36 months, and 399 scores at 42 months, with higher scores reflecting greater inhibitory control. Given the nonsignificant association between parenting partners and secondary caregivers, and given the likely differences in settings and informant perspective in ratings of the child's behavior, we examined secondary caregivers' ratings separately from parents' ratings, consistent with prior work (Rudasill et al., 2014).

Language Ability

Language ability was examined as a covariate. Child language ability was assessed using the Differential Ability Scales, using one version (Elliott, 1997) in the early phase of the study and another version in the later phase (the Differential Ability Scales-II; Elliott, 2007). Language ability for both versions was assessed as the average of the *T*-scores on two language subtests, Verbal Comprehension (receptive language) and Naming Vocabulary (expressive language). *T*-scores were used to ensure comparability of scores across versions.

Socioeconomic Status

SES was examined as a covariate. Given the rank-order stability of SES ($rs > .85$), we averaged the SES scores for a child's family across time. Of children in the final sample, 98% had scores for SES.

Statistical Analysis

We evaluated the role of missing data by performing multiple imputation as a sensitivity analysis. The substantive findings were unchanged when using multiple imputation; thus, the results from the raw data are presented.

Intrameasure Associations

We examined intrameasure associations by conducting Pearson correlations for each inhibitory control measure separately. This included, (a) correlations between Bird/Alligator inhibition (no-go) and activation (go) scores at each age and (b) rank-order stability correlations for Shape Stroop, Bird/Alligator go and no-go, Grass/Snow, and CBQ–IC scores across 30, 36, and 42 months. These analyses were used to determine whether the meaning of each measure changed across time, which would be indicated by changes in the bivariate correlations between subindexes from the same measure across time and low rank-order stability in the measure across time.

As subindexes of the same task, inhibition and activation from Bird/Alligator were expected to be associated with one another across individuals. This is also consistent with prior evidence demonstrating that inhibition and activation were modestly positively associated in children (Muris et al., 2005). However, because inhibition and activation depend on distinguishable subsystems (Gray, 1990), these subindexes were also expected to be distinguishable. If the manifestation of inhibitory control were stable across childhood, we would expect to see a consistent pattern of associations (or nonassociations) between an inhibition measure and measures of other constructs (e.g., activation), as part of the nomological network of inhibitory control, at least across relatively short spans. Alternatively, if inhibitory control shows heterotypic continuity, we would expect to see changes in associations between inhibition and activation across development. In particular, we expected the intrameasure association between inhibition and activation from Bird/Alligator to strengthen over time as children develop more coherent behavioral strategies.

For the rank-order stability of the four inhibitory control measures' inhibition scores across time, we expected somewhat low rank-order stability. CBQ–IC scores were expected to show the strongest rank-order stability because of expected stability in molar behavioral patterns of the child and the adult informant's stability in both relationship with the child and questionnaire response styles (Weijters et al., 2010). Thus, the CBQ–IC was expected to serve as a relatively stable metric of inhibitory control, allowing us to interpret developmental changes in the other indexes. Nevertheless, even items on the CBQ–IC have shown change in meaning over time, consistent with heterotypic continuity (Geeraerts et al., 2021).

Intermeasure Associations

Next, we examined intermeasure associations using Pearson correlations. We also compared the magnitude of the associations across time using Fisher's r-to-z tests. We expected to find moderate associations among Bird/Alligator no-go, Shape Stroop, and Grass/Snow scores based on prior work suggesting that performance inhibition (indexed by Bird/Alligator no-go), perceptual inhibition (indexed by Shape Stroop), and association inhibition (indexed by Grass/Snow) may be related, even if distinct, facets of inhibitory control (Petersen et al., 2016). We also expected that the bivariate correlations among the measures would change with age. Meta-analytic evidence suggests that perceptual inhibition develops earlier than performance inhibition and association inhibition (Petersen et al., 2016). As children get older and approach school-age, they appear to develop more advanced inhibition skills. With such advances, inhibition skills may also become more coherent and integrated, allowing the child to inhibit responses not only to perceptual information, but also to highly salient behavioral commands. Given this theorized increase in the coherence of the manifestation of inhibitory control, we hypothesized that the association between measures of perceptual inhibition (i.e., Shape Stroop scores) and performance inhibition (i.e., Bird/Alligator no-go scores) would strengthen with age. Likewise, we hypothesized that the association between measures of perceptual inhibition (i.e., Shape Stroop scores) and association inhibition (i.e., Grass/Snow scores) would strengthen with age. We had no specific predictions about changes in the association between Bird/Alligator no-go scores and Grass/Snow scores across time because performance inhibition and association inhibition appear to develop around the same time (Petersen et al., 2016). We also had no specific predictions about changes in the association between CBQ–IC scores and other measures across time because of the expected rank-order stability of CBQ–IC scores.

Sensitivity Analyses: Spearman's Rho, Covariates, and Exclusion of Scores at Ceiling or Floor

We conducted several sensitivity analyses of the bivariate intra- and intermeasure associations. As a sensitivity analysis to supplement the Pearson correlations for the intrameasure and intermeasure associations described above, we examined Spearman's rho to determine the extent to which our findings may have been driven by extreme values. Spearman's rho is less influenced by extreme values compared with Pearson correlations (Caruso & Cliff, 1997). We also examined whether the intrameasure and intermeasure associations differed when excluding scores that reflected potential ceiling effects (maximum possible score) or floor effects (minimum possible score).

In addition, we examined whether the intrameasure and intermeasure associations differed when controlling for children's language ability, SES, or sex using partial correlations. Sensitivity analyses are presented in Supplementary Materials S2. Language development may explain part of the developmental pattern of associations among the lab task indexes of inhibitory control as language ability has predicted inhibitory control on these tasks in previous research (Petersen, Bates, & Staples, 2015), and individual differences in performance on tasks that are intended to assess inhibitory control may reflect differences in the comprehension of task rules. In addition, we examined these associations while controlling for SES, considering the positive association between SES and inhibitory control (Sarsour et al., 2011). Lastly, we examined sex as a covariate and whether there were sex-related differences in inhibitory control, considering that girls tend to demonstrate better inhibitory control than boys (e.g., Kochanska et al., 1997). For any observed differences between boys and girls in inhibitory control, we examined whether the differences held when controlling for differences in language ability given the commonly observed sex-related differences in language ability (Zambrana et al., 2012). This additional test could identify mechanisms involved in girls' apparent advantages over boys' inhibitory control development. Analyses examining sex-related differences are described in Supplementary Materials S3.

Growth Curve Analyses

To examine mean-level growth in inhibitory control, and intra- and intermeasure associations of intercepts and slopes, we examined growth curve models in hierarchical linear modeling. Prior

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

work has established that there are individual differences in level and rates of inhibitory control development (e.g., Moilanen et al., 2010). We hypothesized rapid growth in inhibitory control from 30 to 42 months. We also expected that children's individual differences in levels and rates of change of inhibitory control would be correlated within and across measures in theoretically meaningful ways consistent with convergent and discriminant validity. Growth curve models and results are described in Supplementary Materials S4 and summarized in the results below.

Latent Construct Analyses

To determine whether the measures demonstrated longitudinal factorial invariance, we examined a latent inhibitory control construct using structural equation modeling. Consistent with the overarching notion of heterotypic continuity, we expected that there would be continuity in the construct of inhibitory control, as assessed with a latent factor, with simultaneous changes in the behavioral manifestations of the construct (see Figure 1). The latent construct analyses are described in detail in Supplementary Materials S5 and summarized in the results below.

Results

Intrameasure Associations Across Time

Intrameasure correlations are provided in Table 1. Fisher's *r*-to-*z* tests are in Table 2.

Bird/Alligator Task

Scatterplots of the association between go and no-go scores from the Bird/Alligator task at each age are shown in Figure 2. No-go scores are the most-often-used index of inhibitory control, and we examined go scores as a complement to no-go scores. Go and no-go scores were negatively associated at 30 and 36 months, suggesting go and no-go scores assessed different processes (i.e., activation and inhibition) that may be in conflict at this point in development. At 42 months, go and no-go scores were positively, but not significantly, associated using Pearson correlation, and were positively and significantly associated using Spearman's rho, suggesting that go and no-go scores may assess processes that support each other at age 42 months. The association between go and no-go scores was significantly greater (more negative) at 30 than at 36 months, with a medium effect size. The association between go and no-go scores at 42 months was also significantly greater (more positive) at 42 months than at 30 and 36 months, with a large and small-to-medium effect size, respectively.

We also observed a developmental shift in the rank-order stability of Bird/Alligator no-go scores across time, with stronger rank-order stability from 36 to 42 months, compared with 30 to 36 months, with a medium effect size. Notably, no-go scores at 30 months were not significantly associated with no-go scores at 42 months. Go scores showed nonsignificantly stronger rank-order stability from 36 to 42 months, compared with 30 to 36 months, with a small effect size. Similar to no-go scores, go scores at 30 months were not significantly associated with go scores at 42 months.

Shape Stroop Task

Shape Stroop scores showed modest rank-order stability from 30 to 42 months. There was no significant difference in the rank-order stability of Shape Stroop scores from 30 to 36 months compared with 36 to 42 months, although the correlations were in the direction of stronger rank-order stability with age, with a small effect size.

Grass/Snow Task

Like Bird/Alligator no-go scores, Grass/Snow scores showed no rank-order stability from 30 to 42 months. There was a trend of stronger rank-order stability of Grass/Snow scores from 36 to 42 months compared with 30 to 36 months, with a small effect size.

Children's Behavior Questionnaire–Inhibitory Control Ratings

Parent- and secondary caregiver-reported CBQ–IC scores showed moderate rank-order stability from 30 to 42 months. There was no significant difference in the rank-order stability of parent- or secondary caregiver-reported CBQ–IC scores from 30 to 36 months compared with 36 to 42 months, although the correlations were in the direction of stronger rank-order stability with age, with a small effect size.

Intermeasure Associations Across Time

Intermeasure correlations are provided in Table 1. We examined concurrent correlations between Bird/Alligator no-go scores and scores on the Shape Stroop task, at each age. Bird/Alligator no-go scores were not associated with Shape Stroop at 30 months, but they were positively associated at 36 months and even more strongly positively associated at 42 months, suggesting that convergent validity among these inhibitory control measures increased with age. The association significantly strengthened from 30 to 36 months with a small-to-medium effect size. However, the association did not significantly strengthen from 36 to 42 months, although the correlation was nonsignificantly larger with a small effect size.

Next, we examined correlations between Bird/Alligator no-go scores and scores on the Grass/Snow task at each age. Bird/Alligator no-go scores were positively associated with Grass/Snow at 30, 36, and 42 months. Although the correlation at 36 months appears larger than the correlation at 30 months, the difference between these two correlations was nonsignificant. There was also no significant difference from 36 to 42 months in the strength of the association.

Next, we examined correlations between Bird/Alligator no-go scores and CBQ–IC scores, at each age. Bird/Alligator no-go scores were not associated with parent- or secondary caregiver-reported CBQ–IC scores at 30 months but were positively associated with parent- and secondary caregiver-reported CBQ–IC scores at 36 and 42 months (the association with secondary caregivers' ratings at 36 months was at a trend level). For parents' ratings, the association was greater at a trend level from 30 to 36 months and from 36 to 42 months, with a small effect size. For secondary caregivers' ratings, the association did not significantly strengthen from 30 to 36 months, or from 36 to 42 months, although the correlations were in the direction of stronger convergent validity with age, with a small effect size.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Pearson Correlation Matrix of Study Variables

			30 months							36 months							42 months						
Variable	Sex	SES	Lang	BA Go	BA No-Go	SS	GS	PR	SCR	Lang	BA Go	BA No-Go	SS	GS	PR	SCR	Lang	BA Go	BA No-Go	SS	GS	PR	SCR
Sex	—																						
SES	−.01	—																					
Lang 30	.08†	.27***	—																				
BA Go 30	.20***	.07	.27***	—																			
BA No-Go 30	−.15*	.02	−.07	−.44***	—																		
SS 30	.09†	.02	.37***	.14*	−.02	—																	
GS 30	.04	−.03	−.04	−.01	.16*	−.04	—																
PR 30	.22***	.04	.14***	.07	.00	−.02	.05	—															
SCR 30	.21***	.12†	.14*	.08	.01	.02	−.02	.15*	—														
Lang 36	.09†	.19***	.65***	.26***	−.01	.29***	−.13*	.07	.23***	—													
BA Go 36	.10†	.03	.20***	.07	−.03	.04	−.05	.07	.03	.18***	—												
BA No-Go 36	.09†	.07	.22***	.11†	.16*	.17***	.07	.14*	.19*	.28***	−.11*	—											
SS 36	.10*	.11*	.37***	.10	−.02	.23***	−.11†	.01	.25***	.39***	.15***	.14*	—										
GS 36	.05	.00	−.01	.05	.06	.02	.08	.00	.08	.03	.01	.20***	−.02	—									
PR 36	.26***	.07	.10†	.14*	−.03	−.01	.06	.62***	.19*	.07	.00	.13*	.05	.03	—								
SCR 36	.23***	.04	.14†	.21*	.08	.08	.11	.24***	.50***	.15*	.02	.14†	.18*	.16*	.34***	—							
Lang 42	.07	.17***	.55***	.24***	−.15*	.29***	−.09	.07	.17*	.67***	.22***	.20***	.38***	−.05	.05	.06	—						
BA Go 42	.10†	−.01	.11*	.09	−.04	.00	−.07	.04	.11	.16***	.20***	.10†	.17***	.07	.03	.11	.12*	—					
BA No-Go 42	.11*	.11*	.33***	.22***	−.01	.17***	−.01	.10†	.18*	.41***	.14*	.40***	.32***	.08	.13*	.16*	.41***	.07	—				
SS 42	.05	−.01	.22***	.21***	.01	.24***	−.05	−.03	.14†	.26***	.12*	.05	.28***	−.03	−.07	.03	.27***	.13*	.20***	—			
GS 42	.06	.13*	.15***	.17***	−.01	.16***	−.01	−.08	.11	.22***	−.07	.19***	.15*	.21***	−.01	.10	.20***	.13*	.17***	.09†	—		
PR 42	.22***	.07	.10†	.03	.01	−.02	.06	.55***	.13†	.10†	−.01	.12*	.08	.09	.68***	.33***	.09†	.01	.25***	.03	−.03	—	
SCR 42	.33***	.07	.09	.18*	.06	.00	.10	.23***	.45***	.16*	−.04	.23***	.09	.07	.23***	.55***	.19*	.21***	.27***	.09	.15*	.19*	—

Note. SES = socioeconomic status; Lang = language ability; BA = Bird/Alligator; NG = no-go; SS = Shape Stroop; GS = Grass/Snow. PR and SCR are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form. Sex is coded with male = 0 and female = 1.
† $p < .10$. * $p < .05$. *** $p < .001$; all *p*s two-tailed.

Table 2
Fisher's r-to-z Tests and Effect Sizes (Cohen's q)

Type	Association 1	r_1	Association 2	r_2	z	p	Cohen's q
Cross-time stability	BA Go 30–BA Go 36	.07	BA Go 36–BA Go 42	.20	1.60	.111	0.13
Cross-time stability	BA No-Go 30–BA No-Go 36	.16	BA No-Go 36–BA No-Go 42	.40	3.09	.002	0.24
Cross-time stability	SS 30–SS 36	.23	SS 36–SS 42	.28	0.71	.475	0.05
Cross-time stability	GS 30–GS 36	.08	GS 36–GS 42	.21	1.70	.088	0.14
Cross-time stability	PR 30–PR 36	.62	PR 36–PR 42	.68	1.56	.118	0.11
Cross-time stability	SCR 30–SCR 36	.50	SCR 36–SCR 42	.55	0.65	.519	0.08
Intrameasure association	B/A Go 30–B/A No-Go 30	−.44	B/A Go 36–B/A No-Go 36	−.11	4.76	< .001	0.35
Intrameasure association	B/A Go 36–B/A No-Go 36	−.11	B/A Go 42–B/A No-Go 42	.07	2.47	.014	0.18
Intrameasure association	B/A Go 30–B/A No-Go 30	−.44	B/A Go 42–B/A No-Go 42	.07	7.29	< .001	0.53
Intermeasure association	B/A No-Go 30–SS 30	−.02	B/A No-Go 36–SS 36	.14	2.10	.036	0.16
Intermeasure association	B/A No-Go 36–SS 36	.14	B/A No-Go 42–SS 42	.20	0.88	.379	0.06
Intermeasure association	B/A No-Go 30–GS 30	.16	B/A No-Go 36–GS 36	.20	0.52	.602	0.04
Intermeasure association	B/A No-Go 36–GS 36	.20	B/A No-Go 42–GS 42	.17	0.40	.691	0.03
Intermeasure association	B/A No-Go 30–PR 30	.00	B/A No-Go 36–PR 36	.13	1.70	.090	0.13
Intermeasure association	B/A No-Go 36–PR 36	.13	B/A No-Go 42–PR 42	.25	1.71	.087	0.13
Intermeasure association	B/A No-Go 30–SCR 30	.01	B/A No-Go 36–SCR 36	.14	1.21	.225	0.13
Intermeasure association	B/A No-Go 36–SCR 36	.14	B/A No-Go 42–SCR 42	.27	1.23	.217	0.13
Intermeasure association	SS 30–GS 30	−.04	SS 42–GS 42	.09	1.79	.073	0.13
Intermeasure association	SS 30–SCR 30	.02	SS 36–SCR 36	.18	1.68	.092	0.17
Intermeasure association	SS 36–SCR 36	.18	SS 42–SCR 42	.09	0.94	.349	0.09
Intermeasure association	GS 30–SCR 30	−.02	GS 36–SCR 36	.16	1.72	.085	0.19
Intermeasure association	GS 36–SCR 36	.16	GS 42–SCR 42	.15	0.01	.895	0.01

Note. BA = Bird/Alligator; SS = Shape Stroop; GS = Grass/Snow. PR and SCR are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form. Fisher's r-to-z tests compare the magnitude of two associations. The power to detect differences in the magnitude of correlations is considerably lower than the power to detect bivariate associations, so we also provide effect sizes of Fisher's r-to-z tests with Cohen's q, which reflects the magnitude of the difference between two correlation coefficients (Cohen, 1988). For simplicity, some Fisher's r-to-z tests are not shown because of nonsignificant bivariate correlations, for which the associations would not be expected to have significantly different magnitude.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Next, we examined Shape Stroop scores in relation to concurrent Grass/Snow scores. There were no significant correlations at any age, although there was a trend-level positive association at 42 months, and the association became nonsignificantly larger from 30 to 42 months of age (with a small effect size), as with the other intermeasure correlations.

Neither Shape Stroop scores nor Grass/Snow scores were concurrently associated with parent-reported CBQ–IC scores at any age. However, Grass/Snow scores were positively associated with secondary caregiver-reported CBQ–IC scores at 36 and 42 months. The association was greater, at a trend level, from 30 to 36 months, with a small-to-medium effect size. Shape Stroop scores were positively associated with secondary caregiver-reported CBQ–IC scores at 36 months. The association was greater at a trend level from 30 to 36 months, with a small-to-medium effect size.

Covariates

We controlled for child language ability, SES, and sex in a set of partial correlations, and we found that the pattern of associations within and between measures was not changed, with one exception: Bird/Alligator no-go scores were not associated with Shape Stroop scores at 36 months when controlling for language ability, $r(356) = .03$, $p = .572$. Additional detail about covariate analyses is provided in Supplementary Materials S2.

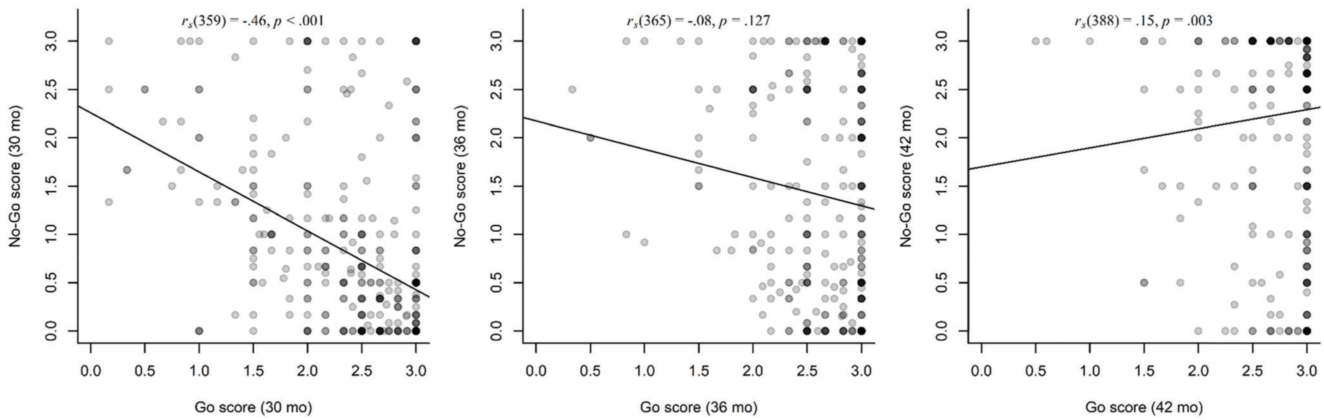
Growth Curves

We next modeled the scores over time with growth curves. We examined intra- and intermeasure associations of the intercept and slope parameters (see Table 3). The intercept and slope within each lab task were strongly correlated ($|rs| > .70$), and the correlation between the intercept and slope for parent-reported CBQ–IC scores was at a trend level and small in magnitude. The direction of the correlations between intercepts and slopes differed by measure. For the Bird/Alligator go, Shape Stroop, and secondary caregiver-reported CBQ–IC scores, the correlations between intercepts and slopes were negative, indicating that children who scored higher initially showed lower rates of growth in scores. For the Bird/Alligator no-go, Grass/Snow, and parent-reported CBQ–IC scores, by contrast, the correlations between intercepts and slopes were positive, indicating that children who started at higher levels showed steeper growth. For additional details, see Supplementary Materials S4.

Across measures, the intercepts of Bird/Alligator no-go scores were positively associated with the intercepts of Shape Stroop, Grass/Snow, and parent- and secondary caregiver-reported CBQ–IC scores but were not associated with intercepts of Bird/Alligator go scores. The slopes of Bird/Alligator no-go scores were positively associated with the slopes of Grass/Snow and parent-reported CBQ–IC scores, demonstrating convergent validity of these measures in a developmental way; however, slopes of Bird/Alligator no-go scores were negatively associated with slopes of Shape Stroop and secondary caregiver-reported CBQ–IC scores. Intercepts and slopes of Bird/Alligator go scores, on the other hand, were less strongly associated with the intercepts and slopes of the other measures, demonstrating discriminant validity of the inhibitory control measures in relation to the Bird/Alligator go scores. The intercepts of Shape Stroop, Grass/Snow, and parent-

Figure 2

Association Between Bird/Alligator Go (Activation) and No-Go (Inhibition) Scores at 30, 36, and 42 Months of Age



Note. Correlations represent Spearman's rho to account for extreme values (see Table S5). Pearson correlations are presented in Table 1. To address the problem of hidden points in the scatterplot because of multiple observed scores occupying the same coordinate location (i.e., overplotting), lighter gray points represent fewer observed scores, whereas darker points represent more observed scores.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

reported CBQ–IC scores were positively associated with intercepts of secondary caregiver-reported CBQ–IC scores. The slopes of Shape Stroop scores were positively associated with slopes of secondary caregiver-reported CBQ–IC scores. For additional details, see Supplementary Materials S4.

Latent Construct

We attempted to fit a confirmatory factor analysis (CFA) model using structural equation modeling (SEM) to examine inhibitory control as a latent construct, composed of the common variance among the Bird/Alligator no-go, Shape Stroop, Grass/Snow, and CBQ–IC scores. We were unable to fit a converging model using longitudinal CFA, perhaps because of the age-dependent pattern of association between performance on Bird/Alligator, Shape Stroop, Grass/Snow, and CBQ–IC and weak intermeasure associations at 30 months. Thus, we were unable to test longitudinal factorial invariance of the latent inhibitory control construct. Because we were unable to fit a longitudinal CFA model, we attempted to fit separate CFA models at each age. At 30 months, the CFA model failed to converge because the Bird/Alligator no-go scores had a negative residual variance. At 36 months and 42 months, the CFA converged and fit well, with positive loadings for Bird/Alligator no-go scores, Shape Stroop scores, Grass/Snow scores, parent-reported CBQ–IC scores, and secondary caregiver-reported CBQ–IC scores (see Figure 3). Bird/Alligator go scores were not significantly associated with the latent inhibitory control construct at either age; however, there was a positive trend-level association at 42 months.

Measures' intercepts and factor loadings appeared to change from 36 to 42 months (see Table 4). Factor loadings of the performance-based tasks increased from 36 to 42 months. By contrast, factor loadings of the parent- and secondary caregiver-reported CBQ–IC scores decreased from 36 to 42 months. In addition, intercepts of the performance-based tasks increased from 36 to 42 months. Additional details are summarized in Supplementary Materials S5.

Discussion

Inhibitory control is associated with important academic and behavioral outcomes, but measurement of inhibitory control presents a challenge: Facets of it appear to develop at different ages (Petersen et al., 2016). The present study responds to this challenge, for the first time we are aware of, by using a set of four well-established inhibitory control measures in a longitudinal design to examine their intra- and intermeasure associations across a year of early childhood. This approach allowed us to determine whether and how these measures of inhibitory control change in meaning across time, as evidenced empirically by changes in associations within and among measures, that is, changes in convergent and discriminant validity. By considering the continuity and change in meanings, we could ask whether inhibitory control demonstrates homotypic or heterotypic continuity, which could have important methodological and developmental implications.

Summary and Interpretation of Findings

Intrameasure Associations

At 30 and 36 months of age, we found a negative association between activation and inhibition, such that children who complied with more of the go commands in the Bird/Alligator task were *less* likely to inhibit on no-go trials at 30 and 36 months of age. At 42 months, the direction of this association changed, suggesting that the meaning of the measure changed over time. At 42 months, the association between activation and inhibition was positive, such that children who complied with more of the go commands in the Bird/Alligator task were *more* likely to inhibit at 42 months. Our findings support the notion that inhibitory control shows heterotypic continuity and changes in manifestation across time. At the early ages, activation and inhibition appeared to be in conflict with each other and did not cohere in expected ways until later in development when they appear to support each other.

Earlier in the year of toddlerhood we observed, those who were less able to inhibit responses to Alligator commands were better at

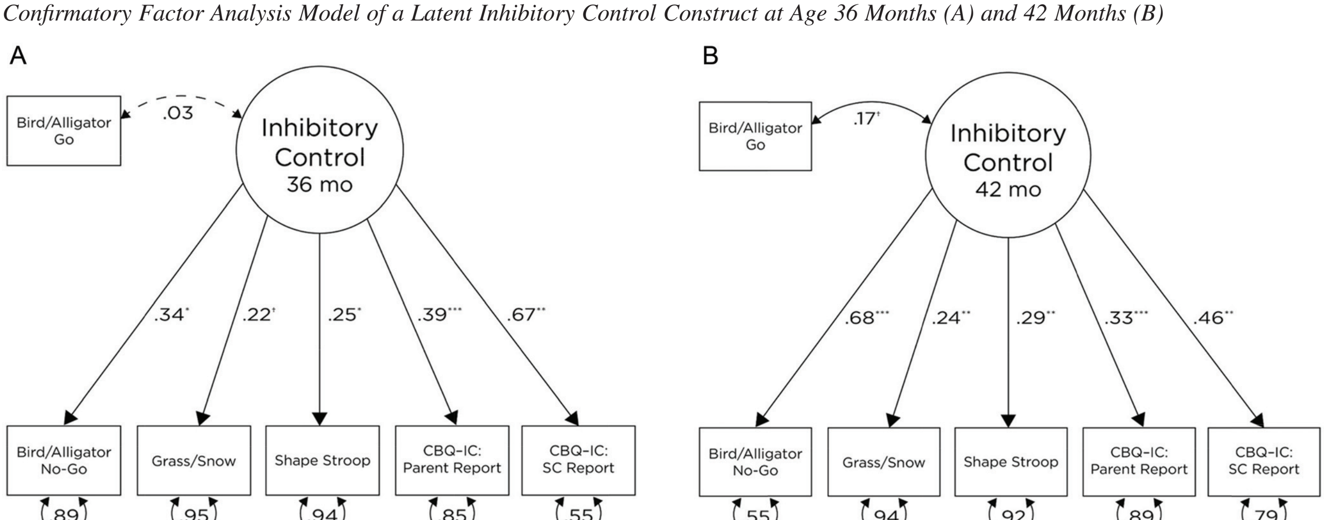
Table 3
Pearson Correlation Matrix of Growth Curve Parameters

Measure	BA Go Intercept	BA Go Slope	BA No-Go Intercept	BA No-Go Slope	SS Intercept	SS Slope	GS Intercept	GS Slope	PR Intercept	PR Slope	SCR Intercept	SCR Slope
BA Go intercept	—											
BA Go slope	−.9997***	—										
BA No-Go intercept	.011	−.010	—									
BA No-Go slope	.077†	−.076†	.987***	—								
SS intercept	.152**	−.151**	.237***	.254***	—							
SS slope	−.149**	.148**	−.235***	−.251***	−.9998***	—						
GS intercept	.055	−.054	.207***	.195***	.054	−.054	—					
GS slope	.087†	−.086†	.206***	.211***	.154**	−.154**	.711***	—				
PR intercept	.076†	−.076†	.154**	.163***	.011	−.011	.026	−.028	—			
PR slope	−.048	.047	.126**	.141**	.041	−.040	.069	.059	.074†	—		
SCR intercept	.121*	−.119*	.218***	.218***	.134*	−.132*	.139*	.118*	.269***	.046	—	
SCR slope	−.087	.087	−.127*	−.129*	−.119*	.117*	−.071	−.056	−.175**	−.021	−.843***	—

Note. BA = Bird/Alligator; SS = Shape Stroop; GS = Grass/Snow. PR and SCR are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form. Some correlation coefficients are presented to four decimal places to distinguish between those with large coefficients ($rs > .99$).
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$; all ps two-tailed.

appropriately following the Bird commands. In other words, children who did better at following correct go responses did worse at inhibiting on the no-go trials. We speculate that at 30 months, the Bird/Alligator no-go scores likely reflect noninhibitory control processes—possibly processes related to child lack of interest in the task or to affective inhibition, such as slightly fearful reaction to the novelty of the task or to the corrections given by the experimenter during training on the task. Consistent with this interpretation, we observed more inhibition on go trials at 30 months than at any other age. This interpretation is also supported by prior findings that infants' positive affect positively predicts, and negative affect inversely predicts, later inhibitory control in toddlerhood (Putnam et al., 2008). Go/no-go tasks (such as Bird/Alligator) are widely used measures of inhibitory control, but in early childhood performance on this type of task might be more strongly associated with children's language, motivation, or attention skills and less strongly associated with a core, separate construct of inhibitory control (as shown by Espy, 2016). However, at later points in development, as language and attention skills normatively become more efficient and automatic, individual differences in performance on a go/no-go task might be more strongly associated with individual differences in inhibitory control. A task's demands on the surface are the same at each age, but how children respond to those task demands likely depends on their underlying skills. Thus, a task's meaning in relation to the construct of inhibitory control likely changes with developmental changes in skills. Future research is necessary to replicate this result and determine the precise functional interpretation of no-go trials at 30 months. At later ages, the no-go scores appeared, in bivariate and multivariate (confirmatory factor) analyses, to assess individual differences

Figure 3
Confirmatory Factor Analysis Model of a Latent Inhibitory Control Construct at Age 36 Months (A) and 42 Months (B)



Note. Factor loadings, residual variances, and covariances are standardized estimates. Dashed line represents a nonsignificant path. CBQ–IC = Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form; SC = secondary caregiver. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4
Factor Loadings and Intercepts of Inhibitory Control Measures at Each Age

Measure	Age (Months)	Unstandardized		Standardized	
		Intercept	Factor Loading	Intercept	Factor Loading
Bird/Alligator No-Go	36	1.37	0.41	1.13	0.34
Shape Stroop	36	1.58	0.14	2.73	0.25
Grass/Snow	36	0.44	0.07	1.38	0.22
CBQ IC: Parent	36	4.41	0.31	5.66	0.39
CBQ IC: Secondary caregiver	36	4.66	0.72	4.33	0.67
Bird/Alligator No-Go	42	2.26	0.74	2.07	0.68
Shape Stroop	42	1.80	0.12	4.32	0.29
Grass/Snow	42	0.58	0.08	1.73	0.24
CBQ IC: Parent	42	4.59	0.26	5.75	0.33
CBQ IC: Secondary caregiver	42	4.77	0.47	4.68	0.46

Note. CBQ IC = Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

in inhibitory control in a more generalizable and convergent way, and thus more similar to patterns of general inhibitory control (Joyce et al., 2016).

We also examined rank-order stability for each of the inhibitory control measures. Bird/Alligator no-go scores and Grass/Snow scores showed some rank-order stability across short-term six-month spans, but they did not show rank-order stability across the 12-month span from 30 to 42 months of age. Shape Stroop scores showed only modest rank-order stability. The limited rank-order stability across these ages and measures further suggests that the inhibitory control measures change in meaning across time, consistent with heterotypic continuity. As expected, CBQ–IC ratings showed stronger rank-order stability than scores on the lab tasks.

Intermeasure Associations

Patterns of association among the measures of inhibitory control also changed over time. At 30 months, Bird/Alligator no-go scores were positively associated with Grass/Snow, but they were not associated with Shape Stroop or CBQ–IC ratings. At 36 months, Bird/Alligator no-go scores were positively associated with Shape Stroop, Grass/Snow, and CBQ–IC scores, and at 42 months they were most strongly associated with the other measures. By contrast, some measures showed weak and nonsignificant associations. Shape Stroop, Grass/Snow, and parent-reported CBQ–IC scores were not significantly related at any age.

Interestingly, parent-reported CBQ–IC scores were correlated with only Bird/Alligator no-go scores at 36 and 42 months, but parent-reported CBQ–IC scores were not correlated with no-go scores at 30 months or with any of the other inhibitory control measures at any age. Perhaps the Bird/Alligator task reflects behaviors that are more similar to those observed by parents in everyday life—a child complying with some commands while also inhibiting a response to competing distractors. For instance, a child may comply with a command not to disturb their parent, whose attention is still desired by the child, when the parent is using the telephone (i.e., the no-go stimulus). In comparison, the behaviors assessed in the Grass Snow task and the Shape Stroop task may be more cognitively abstract and less frequently a part of parent–child interactions. It is not an everyday activity of toddlers to play a game in which the goal is to inhibit a response to a prepotent color–word association or to inhibit a response to competing perceptual information. Sometimes such games are played, perhaps, but not as frequently as there are opportunities for inhibition to a no-go stimulus. This interpretation is speculative and requires replication and further probing in future research. By contrast, *secondary caregiver*–reported CBQ–IC scores showed modest associations with Grass/Snow scores at 36 and 42 months and with Shape Stroop scores at 36 months. Although the convergence of the CBQ–IC with the other inhibitory control measures was limited, we still recognize the utility of the CBQ–IC as a complementary assessment of functional behavior in everyday life, and we discuss the utility of secondary caregiver ratings in more detail below.

In sum, these intermeasure associations, in combination with the intrameasure associations described above, suggest that the measures change in meaning across time, and appear to have greater construct validity for inhibitory control at 36 and 42 months of age than at 30 months, which in turn suggests that the construct of inhibitory control changes in manifestation and reflects heterotypic continuity across this year of toddlerhood.

Covariates: Socioeconomic Status, Sex, and Language Ability

The change in patterns of association could not be explained by differences in SES or sex. The change in patterns of association also could also not be explained by children's differences in language ability, which suggests that the different associations over time could not be attributed to differences in the comprehension of task rules.

Growth Curve Analyses

Although some of the intercept and slope parameters were not associated in expected ways, the overall pattern of intermeasure associations of the intercept and slope parameters across measures generally supported the pattern of convergent and discriminant validity of the measures extracted from the main analyses.

Latent Inhibitory Control Construct

The pattern of different associations by age was confirmed when we considered latent inhibitory control constructs. The measures appear to more consistently and coherently assess the construct of inhibitory control by 36 months of age. Thus, we see all four measures of inhibitory control as distinct but complementary ways of assessing the overarching inhibitory control construct. Although the present results suggest that the meanings of these inhibitory control measures change over time, they do appear to be part of an enduring inhibitory control construct, at least by the older ages (36 and 42 months).

Practical Implications and Recommendations for Future Research

Inhibitory control appeared to change in manifestation with age, consistent with Figure 1. At 30 months, inhibitory control may manifest in more simple ways, such as with inhibition of perceptual information as shown in the Shape Stroop task, which likely requires less advanced or efficient attention skills. At this early age, children's inhibitory control may be inconsistent across tasks and situations; inhibitory difficulties in one situation may not indicate difficulties in other situations. Once inhibitory control skills become more efficient and automatic, children's manifestation of inhibitory control can be more complex, and children may show greater consistency in inhibitory control across situations. The growing complexity of inhibitory control may be observed with inhibition and activation on the Bird Alligator task or with inhibition of prepotent associations on the Grass Snow task by later ages. These changes in the manifestation of inhibitory control may be related to brain development in the prefrontal cortex that supports inhibitory control (Diamond, 2002; Moriguchi & Hiraki, 2013). Few studies have examined the neural basis of inhibitory control in toddlerhood, so this is an important area for future research.

It is widely agreed by developmental researchers that inhibitory control is foundational for children's social development outcomes (Lipszyc & Schachar, 2010; Wright et al., 2014). The present study makes a novel and important contribution to the field of child development by offering an empirical demonstration of how the construct of inhibitory control shows heterotypic continuity. Some prior research suggested that inhibitory control may change in its behavioral manifestation with development (e.g., Chang et al., 2015; Petersen et al., 2016), but no prior studies have fully tested whether inhibitory control shows heterotypic continuity because changes in the associations among inhibitory control measures have not been examined. The present study is the first to empirically test and provide evidence suggesting that inhibitory control shows heterotypic continuity. This demonstration allows recommendations for future research. Previous studies have often used the same inhibitory control measure(s) across ages to examine growth curves of inhibitory control. If the construct changes in manifestation (as our study suggests) and the selected measures do not align with these changes, the measures will lack construct validity invariance, which may lead to inaccurate inferences about development.

These important theoretical and conceptual arguments are not limited to inhibitory control. Many constructs likely change in manifestation with development, but so far, very few studies of constructs have methodologically and statistically accounted for their heterotypic continuity when examining development (Petersen et al., 2020). Thus, the findings from the present study may help lead the field to more closely align methodological approaches for studying constructs with the field's theoretical understanding of them.

Researchers using no-go scores at 30 months to assess inhibition may reach invalid developmental conclusions about inhibitory control at this age and may be better served by using more valid indexes of inhibitory control, such as Shape Stroop, at this early point in development. The finding that no-go inhibition performance at 30 months does not reflect self-regulatory inhibition may also be important for interpreting findings from prior studies that have used the Bird/Alligator task (or other go/no-go tasks) at 30 months of age or earlier. Other than our own work with children as young as 30 months of age from a subset of the present sample (Petersen, Bates, & Staples, 2015), we are aware of at least one study that has used a comparable variant of the Bird/Alligator task in children as young as 25 months of age (Kraybill, 2013).

The present study offers possibilities that the construct of inhibitory control can be assessed in future research to sensitively detect both individual differences and normative developmental changes, such as in a study of how parenting or teaching might influence self-regulation development. Future research will either need to (a) modify the task to assess the same construct over the target age span (e.g., Carlson et al., 2016) or (b) employ and assemble different measures at different ages to retain construct validity invariance (e.g., Petersen et al., 2016). Ignoring heterotypic continuity has been shown to result in (a) measures that are less able to detect growth and (b) incorrect developmental inferences, compared with approaches that account for heterotypic continuity (Chen & Jaffee, 2015; Petersen, LeBeau, et al., 2021; Petersen et al., 2018). For instance, in a study of internalizing problems (such as anxiety and depression) from adolescence to adulthood, no group-level change was observed when using the same measures across ages, whereas internalizing problems showed a group-level decrease when using different, construct-valid measures across ages (Petersen et al., 2018). Additionally, in a simulation study of externalizing problems from early childhood to adolescence where the true slope was specified to be negative, use of the same measures across development incorrectly yielded positive slopes at the group-level (Petersen, LeBeau, et al., 2021). Prior studies have demonstrated ways to account for heterotypic continuity in development by using changing, age-appropriate measures to ensure construct validity invariance and statistical approaches to ensure statistical equivalence (McArdle et al., 2009; Petersen, Bates, Dodge, et al., 2015; Petersen et al., 2016; Petersen & LeBeau, in press; Petersen et al., 2018). For example, developmental scaling has been used to link different measures of externalizing problems (e.g., aggression, rule-breaking) across development on the same scale to retain construct validity and account for heterotypic continuity (Petersen & LeBeau, in press).

Strengths and Limitations

One major strength of the study is its design, using three measurement points with repeated use of multiple inhibitory control measures, including performance-based measures and ratings from multiple informants. Importantly, the secondary caregiver ratings of inhibitory control were strongly related to the latent inhibitory control factor, which suggests that secondary caregivers' ratings are a key complementary assessment of children's inhibitory control. This finding is also consistent with prior work showing the incremental validity of teachers'/secondary caregivers' ratings over and above parents' ratings of children's behavior (McQuillan et al., 2018), including children's temperament as assessed on the CBQ (Rudasill et al., 2014). Secondary caregivers may evaluate child behavior in more structured, academic contexts and in comparison

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

with other age mates, providing a description of child strengths and weaknesses that may complement or, for some purposes, even surpass the informational utility of parent ratings. A second strength of the study was using carefully selected performance-based measures of inhibitory control based on prior research. Third, the longitudinal nature of the study allowed us to examine changes in convergent and discriminant validity of the measures across time, along with rank-order stability. Fourth, the multisite nature of the study and its concomitantly larger sample increase the potential generalizability of the findings. Fifth, the findings that the Bird/Alligator no-go scores changed in meaning with development were corroborated both within the Bird/Alligator task with respect to go scores and in relation to three other inhibitory control measures examined here. Sixth, our findings are consistent with, and extend, the conclusion from a meta-analysis that inhibitory control demonstrates heterotypic continuity (Petersen et al., 2016).

One key limitation of the study is that we were unable to test longitudinal factorial invariance, which is valuable when examining homotypic and heterotypic continuity and was central to our analytic plan. Although we attempted to test longitudinal measurement invariance, the longitudinal measurement model did not converge (because the correlation between the latent inhibitory control factor across ages was greater than 1.0). However, we were able to examine changes across time in the intercepts and factor loadings. Across bivariate correlations, covariance between growth parameters, and changes in measures' intercepts and factor loadings in relation to a latent construct, we found considerable evidence that the inhibitory control measures tested here changed in meaning with age and that the inhibitory control construct shows heterotypic continuity. No prior studies have provided empirical evidence of heterotypic continuity of inhibitory control, so, based on the literature we have seen, this remains a novel and important contribution.

Another possible limitation of the study is that a few lab tasks showed evidence of restricted range, as in ceiling or floor effects. We observed a ceiling effect in activation at 42 months of age on the Bird/Alligator task. However, we observed similar developmental shifts in construct validity when considering inhibition in the Shape Stroop and Grass/Snow tasks, suggesting that the developmental changes in meaning of Bird/Alligator go scores were not fully explained by range restriction. Moreover, the pattern of findings was mostly consistent when examining Spearman's rho and when excluding scores at floor or ceiling (see Supplementary Materials S2), suggesting that neither extreme values nor restricted range drove the changing pattern of associations with development. Furthermore, we followed conventions in administering and scoring the tasks, so the present study represents how they have been used by many researchers. In the future to avoid range restriction, we could consider additional rule changes to increase or decrease task difficulty (e.g., additional trials, rule switches, or less time to respond). To test this possibility, we added a rule switch in the Bird/Alligator task at older ages (36 and 42 months) as a preliminary effort to adapt the measure to older children's growing regulatory abilities. However, in a separate analysis (the findings presented in the present study only considered the prerule switch trials), the rule switch did not affect the pattern of associations observed in the present study, which further suggests that range restriction did not account for the changes in associations with age that we observed. Although Shape Stroop, Grass/Snow, and CBQ–IC converged with Bird/Alligator inhibition in interesting ways, the lack of convergence among the Shape Stroop, Grass/Snow, and parent-reported CBQ–IC scores was unexpected. This nonconvergence may owe to measurement limitations or to specific task demands that will require further research to understand.

Third, inhibitory control scores were more likely to be missing for children whose primary caregiver was Hispanic or African American and children who were from lower SES families. This may limit the generalizability of the findings. It will be important for future research to examine whether inhibitory control changes in manifestation in more ethnically and socioeconomically diverse samples. Nevertheless, our findings held when controlling for SES.

Fourth, in the present study, we used multiple measures tested repeatedly at three different measurement points, but attention to the number and timing of assessments is critical. We used three measurement points at 6-month intervals across a year of early childhood because early childhood is characterized by rapid growth in inhibitory control (Goswami, 2011). However, future research that examines whether additional measurement points at finer (< 6 months) and broader intervals (> 6 months), and at different ages, might reveal additional nuances in the measurement and manifestation of inhibitory control. For example, although we found that the measures converged into a latent factor by 36 months, our selected measurement points do not allow us to know the precise age at which the measures converge or whether inhibitory control also changes in manifestation at later ages. Another valuable future direction will be to examine changes in item-level parameters such as item difficulty and discrimination (Millsap, 2010).

Conclusions

To our knowledge, this is the first study to empirically demonstrate that measures of inhibitory control change in meaning with development. The Bird/Alligator task changed in its meaning with development, within its go and no-go components and in relation to the Shape Stroop, Grass/Snow, and CBQ–IC measures. It appears that no-go inhibition performance at 30 months does not clearly reflect self-regulatory inhibition, but by 42 months, it does clearly reflect self-regulatory inhibition. Our findings are consistent with the interpretation that inhibitory control shows heterotypic continuity in early childhood (Petersen et al., 2016). If this is the case, no-go scores from go/no-go tasks may not be conceptually comparable across development from 30 to 42 months because observed inhibition appears to mean different things in relation to the inhibitory control construct at different ages. Moreover, as a set, the inhibitory control measures do not meet a standard criterion of construct validity invariance over the 30- to 42-month age span.

The common practice in developmental psychology of using the same measure across ages is useful for some purposes—for example, growth curve modeling, but it can also introduce the risk that the measure may actually assess somewhat different constructs over time. With improved capacity to assess the development of inhibitory control, our understanding of how inhibitory control characteristics develop will also improve, as will our understanding of the developmental implications of inhibitory control. Our findings empirically demonstrate the heterotypic continuity of inhibitory control, as it appears to manifest differently at different ages, improves in the year from age two-and-a-half to age three-and-a-half years, and becomes a more coherent trait, with greater stability in individual differences and greater cross-measure convergence.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

References

Acar, I. H., Frohn, S., Prokasky, A., Molfese, V. J., & Bates, J. E. (2019). Examining the associations between performance based and ratings of focused attention in toddlers: Are we measuring the same constructs? *Infant and Child Development*, *28*(1), e2116. https://doi.org/10.1002/icd.2116

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth, and Families.

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology*, *50*(10), 2368–2379. https://doi.org/10.1037/a0037493

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. Guilford Press.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, *28*(2), 561–571. https://doi.org/10.1207/s15326942dn2802_1

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

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., Faja, S., & Beck, D. M. (2016). Incorporating early development into the measurement of executive function: The need for a continuum of measures across development. In J. A. Griffin, P. McCardle, & L. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 45–64). American Psychological Association. https://doi.org/10.1037/14797-003

Caruso, J. C., & Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurement*, *57*(4), 637–654. https://doi.org/10.1177/0013164497057004009

Caspi, A., & Shiner, R. L. (2006). Personality development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (6th ed., Vol. 3, pp. 300–365). Wiley. https://doi.org/10.1002/9780470147658.chpsy0306

Caughy, M. O. B., Mills, B., Owen, M. T., & Hurst, J. R. (2013). Emergent self-regulation skills among very young ethnic minority children: A confirmatory factor model. *Journal of Experimental Child Psychology*, *116*(4), 839–855. https://doi.org/10.1016/j.jecp.2013.07.017

Chang, H., Shaw, D. S., & Cheong, J. (2015). The development of emotional and behavioral control in early childhood: Heterotypic continuity and relations to early school adjustment. *Journal of Child and Adolescent Behavior*, *3*(3), 1–10. https://doi.org/10.4172/2375-4494.1000204

Chen, F. R., & Jaffee, S. R. (2015). The heterogeneity in the development of homotypic and heterotypic antisocial behavior. *Journal of Developmental and Life-Course Criminology*, *1*(3), 269–288. https://doi.org/10.1007/s40865-015-0012-3

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, *70*(1), 6–20. https://doi.org/10.1037/0022-006X.70.1.6

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195134971.003.0029

Durbin, C. E., Hayden, E. P., Klein, D. N., & Olino, T. M. (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion*, *7*(2), 388–399. https://doi.org/10.1037/1528-3542.7.2.388

Eigsti, I.-M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M. B., Davidson, M. C., Aber, J. L., & Casey, B. J. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science*, *17*(6), 478–484. https://doi.org/10.1111/j.1467-9280.2006.01732.x

Elliott, C. D. (1997). The Differential Ability Scales. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 183–208). Guilford Press.

Elliott, C. D. (2007). *Differential ability scales–2nd ed.: Administration and scoring manual*. Harcourt Assessment.

Espy, K. A. (2016). The changing nature of executive control in preschool *Monographs of the Society for Research in Child Development*, *81*(4), 1–6. https://doi.org/10.1111/mono.12267

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Geeraerts, S. B., Endendijk, J. J., Deković, M., Huijding, J., Deater-Deckard, K., & Mesman, J. (2021). Inhibitory control across the preschool years: Developmental changes and associations with parenting. *Child Development*, *92*(1), 335–350. https://doi.org/10.1111/cdev.13426

Goswami, U. (2011). *The Wiley-Blackwell handbook of childhood cognitive development*. Wiley-Blackwell. https://doi.org/10.1002/9781444325485ch

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion*, *4*(3), 269–288. https://doi.org/10.1080/02699939008410799

Hollingshead, A. B. (1975). *Four factor index of social status*. Yale University, Department of Sociology.

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*(2), 561–581. https://doi.org/10.1111/j.1540-5907.2010.00447.x

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1–47. https://doi.org/10.18637/jss.v045.i07

Joyce, A. W., Kraybill, J. H., Chen, N., Cuevas, K., Deater-Deckard, K., & Bell, M. A. (2016). A longitudinal investigation of conflict and delay inhibitory control in toddlers and preschoolers. *Early Education and Development*, *27*(6), 788–804. https://doi.org/10.1080/10409289.2016.1148481

Kaufman, R. L. (2013). *Heteroskedasticity in regression: Detection and correction*. Sage. https://doi.org/10.4135/9781452270128

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

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. https://doi.org/10.2307/1131828

Kraybill, J. H. (2013). *A latent factor analysis of preschool executive functions: Investigations of antecedents and outcomes* [Doctoral dissertation, VA Polytechnic Institute and State University].

Kuczynski, L., & Kochanska, G. (1990). Development of children's noncompliance strategies from toddlerhood to age 5. *Developmental Psychology*, *26*(3), 398–408. https://doi.org/10.1037/0012-1649.26.3.398

Lipszyc, J., & Schachar, R. (2010). Inhibitory control and psychopathology: A meta-analysis of studies using the stop signal task. *Journal of the*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

International Neuropsychological Society, *16*(6), 1064–1076. https://doi.org/10.1017/S1355617710000895

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

Lumley, T. (2010). mitools: Tools for multiple imputation of missing data. https://cran.r-project.org/package=mitools

Marek, S., Hwang, K., Foran, W., Hallquist, M. N., & Luna, B. (2015). The contribution of network organization and integration to the development of cognitive control. *PLoS Biology*, *13*(12), e1002328 https://doi.org/10.1371/journal.pbio.1002328

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

McQuillan, M. E., Kultur, E. C., Bates, J. E., O'Reilly, L. M., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2018). Dysregulation in children: Origins and implications from age 5 to age 28. *Development and Psychopathology*, *30*(2), 695–713. https://doi.org/10.1017/S0954579417001572

Miller, J. L., Vaillancourt, T., & Boyle, M. H. (2009). Examining the heterotypic continuity of aggression using teacher reports: Results from a national Canadian study. *Social Development*, *18*(1), 164–180. https://doi.org/10.1111/j.1467-9507.2008.00480.x

Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4*(1), 5–9. https://doi.org/10.1111/j.1750-8606.2009.00109.x

Moilanen, K. L., Shaw, D. S., Dishion, T. J., Gardner, F., & Wilson, M. (2010). Predictors of longitudinal growth in inhibitory control in early childhood. *Social Development*, *19*(2), 326–347. https://doi.org/10.1111/j.1467-9507.2009.00536.x

Moriguchi, Y., & Hiraki, K. (2013). Prefrontal cortex and executive function in young children: A review of NIRS studies. *Frontiers in Human Neuroscience*, *7*, 867. https://doi.org/10.3389/fnhum.2013.00867

Muris, P., Meesters, C., de Kanter, E., & Timmerman, P. E. (2005). Behavioural inhibition and behavioural activation system scales for children: Relationships with Eysenck's personality traits and psychopathological symptoms. *Personality and Individual Differences*, *38*(4), 831–841. https://doi.org/10.1016/j.paid.2004.06.007

Muthén, L. K., & Muthén, B. O. (2019). Mplus (Version 8.4) [Computer software]. https://www.statmodel.com/programs.shtml

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, *86*(4), 852–875. https://doi.org/10.1037/0033-2909.86.4.852

Patterson, G. R. (1993). Orderly change in a stable world: The antisocial trait as a chimera. *Journal of Consulting and Clinical Psychology*, *61*(6), 911–919. https://doi.org/10.1037/0022-006X.61.6.911

Petersen, I. T., & LeBeau, B. (in press). Creating a developmental scale to chart the development of psychopathology with different informants and measures across time. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000649

Petersen, I. T., Bates, J. E., & Staples, A. D. (2015). The role of language ability and self-regulation in the development of inattentive-hyperactive behavior problems. *Development and Psychopathology*, *27*(1), 221–237. https://doi.org/10.1017/S0954579414000698

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

Petersen, I. T., Bates, J. E., McQuillan, M. E., Hoyniak, C., Staples, A. D., Rudasill, K. M., Molfese, D. L., & Molfese, V. J. (2021). *Heterotypic continuity of inhibitory control in early childhood: Evidence from four widely used measures* [Data dictionary]. https://doi.org/10.17605/OSF.IO/A52J4

Petersen, I. T., Choe, D. E., & LeBeau, B. (2020). Studying a moving target in development: The challenge and opportunity of heterotypic continuity. *Developmental Review*, *58*, 100935. https://doi.org/10.1016/j.dr.2020.100935

Petersen, I. T., Hoyniak, C. P., McQuillan, M. E., Bates, J. E., & Staples, A. D. (2016). Measuring the development of inhibitory control: The challenge of heterotypic continuity. *Developmental Review*, *40*, 25–71. https://doi.org/10.1016/j.dr.2016.02.001

Petersen, I. T., LeBeau, B., & Choe, D. E. (2021). Creating a developmental scale to account for heterotypic continuity in development: A simulation study. *Child Development*, *92*(1), e1–e19. https://doi.org/10.1111/cdev.13433

Petersen, I. T., Lindhiem, O., LeBeau, B., Bates, J. E., Pettit, G. S., Lansford, J. E., & Dodge, K. A. (2018). Development of internalizing problems from adolescence to emerging adulthood: Accounting for heterotypic continuity with vertical scaling. *Developmental Psychology*, *54*(3), 586–599. https://doi.org/10.1037/dev0000449

Pinheiro, J. B. D., DebRoy, S., & Sarkar, D., and the R Core team. (2009). nlme: Linear and nonlinear mixed effects models. R package version 3.1-93. https://cran.r-project.org/package=nlme

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

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development*, *17*(4), 387–405. https://doi.org/10.1002/icd.582

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Rudasill, K. M., Prokasky, A., Tu, X., Frohn, S., Sirota, K., & Molfese, V. J. (2014). Parent vs. teacher ratings of children's shyness as predictors of language and attention skills. *Learning and Individual Differences*, *34*, 57–62. https://doi.org/10.1016/j.lindif.2014.05.008

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society*, *17*(1), 120–132. https://doi.org/10.1017/S1355617710001335

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (Vol. 72). Chapman & Hall/CRC. https://doi.org/10.1201/9781439821862

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–337. https://doi.org/10.3200/JOER.99.6.323-338

Sophian, C. (1997). Beyond competence: The significance of performance for conceptual development. *Cognitive Development*, *12*(3), 281–303. https://doi.org/10.1016/S0885-2014(97)90001-0

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods*, *15*(1), 96–110. https://doi.org/10.1037/a0018721

Willoughby, M. T., Kuhn, L. J., Blair, C. B., Samek, A., & List, J. A. (2017). The test-retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators. *Child Neuropsychology*, *23*(7), 822–837. https://doi.org/10.1080/09297049.2016.1205009

Wright, L., Lipszyc, J., Dupuis, A., Thayapararajah, S. W., & Schachar, R. (2014). Response inhibition and psychopathology: A meta-analysis of go/no-go task performance. *Journal of Abnormal Psychology*, *123*(2), 429–439. https://doi.org/10.1037/a0036295

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*(1), 117–130. https://doi.org/10.1037/a0014657

Zalewski, M., Lengua, L. J., Fisher, P. A., Trancik, A., Bush, N. R., & Meltzoff, A. N. (2012). Poverty and single parenting: Relations with preschoolers' cortisol and effortful control. *Infant and Child Development*, *21*(5), 537–554. https://doi.org/10.1002/icd.1759

Zambrana, I. M., Ystrom, E., & Pons, F. (2012). Impact of gender, maternal education, and birth order on the development of language comprehension: A longitudinal study from 18 to 36 months of age. *Journal of Developmental and Behavioral Pediatrics*, *33*(2), 146–155. https://doi.org/10.1097/DBP.0b013e31823d4f83

Received May 1, 2020
Revision received June 13, 2021
Accepted July 15, 2021 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Supplementary Materials S1. Tests of systematic missingness and descriptions of missing data handling.

The number of observed scores on each measure is in Table S1. Percent of those with scores (out of those who had a laboratory visit) are provided in Table S3.

Tests of Systematic Differences between the Recruited Sample and Final Sample

First, we examined whether there were systematic differences between the recruited sample and the final sample. Children who were and were not included in the final sample did not differ in terms of age ($t[54.51] = -0.01$, $p = .996$), sex ($\chi^2[1] = 2.11$, $p = .146$), or socioeconomic status (SES; $t[4.04] = 1.05$, $p = .353$), or ethnicity of the primary caregiver ($\chi^2[6] = 0.96$, $p = .987$) or parenting partner ($\chi^2[6] = 0.48$, $p = .998$). Children who were and were not included in the final sample did not differ in terms of their parents' marital status ($\chi^2[4] = 22.27$, $p < .001$). Among primary caregivers who reported marital status ($N = 514$), children whose parents were separated were less likely to be included in the final sample (75%; or 3 out of 4 children) compared to parents who were single (95%), married (99%), divorced (100%), or re-married (100%). However, among families whose primary caregivers reported the parents were separated, only one child was not included in the final sample.

Tests of Systematic Missingness in Inhibitory Control Scores

Second, we tested whether there was systematic missingness in inhibitory control scores. We examined whether the number of a child's missing scores across inhibitory control variables (Bird/Alligator no-go, Shape Stroop, Grass/Snow, and parent- and secondary caregiver-report on the Inhibitory Control subscale of the Children's Behavior Questionnaire) was related to other variables. The number of a child's missing inhibitory control scores was positively correlated with age ($r[1536] = .07$, $p = .006$), such that older children tended to have more missing scores than younger children, which likely reflects attrition. The number of missing scores was not significantly related to sex (t[495.41] = 1.35, p = .178). The number of missing scores was negatively related to a continuous index of SES (r[500] = -.13, p = .004), such that children from lower SES families tended to have more missing scores than children from higher SES families.

We conducted linear regression models to examine whether the number of missing inhibitory control scores was related to the family's marital status or the ethnicity of the parents. In terms of marital status, number of missing scores tended to be higher for children whose primary caregiver was single (B[502] = 3.01, p < .001) or separated (B[502] = 3.67, p < .001), compared to children whose primary caregiver was married. The number of missing scores did not differ for children whose primary caregiver was divorced (B[502] = 1.40, p = .178) or re-married (B[502] = -0.06, p = .972), compared to children whose primary caregiver was married.

Primary caregivers' ethnicity showed some association with children's number of missing scores. Children whose primary caregiver was Hispanic (B[497] = 2.12, p = .021) or African American (B[497] = 3.31, p = .002) tended to have more missing scores compared to children whose primary caregiver was non-Hispanic White. Children whose primary caregiver was mixed race (B[497] = 2.74, p = .081), American Indian (B[497] = -1.69, p = .681), Asian American (B[497] = 0.56, p = .642), or "other" ethnicity" (B[497] = 3.81, p = .066) did not have statistically significantly more missing scores compared to children whose primary caregiver was non-Hispanic White. Parenting partners' ethnicity was unrelated to children's number of missing scores: Hispanic (B[459] = 0.29, p = .790), African American (B[459] = 1.14, p = .250), Asian American (B[459] = 0.32, p = .750), mixed race (B[459] = 1.88, p = .164), American Indian (B[459] = -1.82, p = .366), "other" ethnicity (B[459] = -0.23, p = .920).

Missing Data Handling

As a final sensitivity analysis, we examined models with multiple imputation using the Amelia package (Honaker et al., 2011) in R. Amelia uses an expectation–maximization with bootstrapping algorithm, and is well suited for longitudinal data (Honaker & King, 2010). All model variables were used to create imputed values for 100 data sets. The HLM growth curve models were run on each imputed data set separately, and then the results were combined using the mitools (Lumley, 2010) and mix (Schafer, 1997) packages in R, which use Rubin's (1987) rules for combining results of analyses on multiply imputed data sets. The substantive findings were unchanged when using multiple imputation, providing greater confidence in the findings. Thus, results from the raw data are presented.

Supplementary Materials S2. Sensitivity analyses: Spearman's rho, covariates, and exclusion of scores at ceiling or floor.

Spearman's Rho

Results

As a sensitivity test, we examined intra- and inter-measure associations with Spearman's rho to determine the extent to which our findings may have been driven by extreme values. Spearman's rho is less influenced by extreme values compared to Pearson correlations (Caruso & Cliff, 1997). Spearman correlations are in Table S5. Findings were consistent when examining Spearman's rho, with one exception: Bird/Alligator go and no-go scores were positively associated at 42 months.

Discussion

Overall, the changing patterns of associations with development were consistent when examining Spearman's rho, suggesting that the changing patterns of associations with development did not reflect extreme values.

Exclusion of Scores at Ceiling or Floor

Results

As an additional sensitivity test, we examined whether the intra-measure and inter-measure associations differed when excluding scores on the lab tasks that reflected potential ceiling effects (maximum possible score) or floor effects (minimum possible score). After excluding scores at ceiling or floor, some estimates of cross-time stability were somewhat stronger, including Bird/Alligator go scores from 36 to 42 months ($r[50] = .34$, $p = .013$), and Bird/Alligator no-go scores from 30 to 42 months ($r[56] = .24$, $p = .066$). Other estimates of cross-time stability were somewhat weaker, including Bird/Alligator no-go scores from 30 to 36 months (r[83] = .10, p = .345), Shape Stroop scores from 30 to 36 months (r[81] = .15, p = .172), and Shape Stroop scores from 30 to 42 months (r[45] = .15, p = .321). Nevertheless, the pattern of findings was generally consistent. Bird/Alligator no-go scores continued to show a more negative association with Bird/Alligator go scores at 30 months than 36 and 42 months. In general, inter-measure associations continued to show stronger associations at 36 months than 30 months. Inter-measure associations were somewhat weakened at 42 months, which may reflect the further restricted range when excluding minimum and maximum possible scores—the percent of scores that were excluded at 42 months was 65% for Bird/Alligator, 89% for Shape Stroop, and 72% for Grass/Snow.

Discussion

Although some associations became somewhat stronger or weaker when excluding scores at ceiling or floor, the changing patterns of associations with development were generally consistent. This suggests that the changing patterns of associations with development did not reflect ceiling or floor effects.

Covariates

Results

As a final sensitivity test, we also controlled for child language ability, SES, and sex in a set of partial correlations. We found that the pattern of associations within and between measures was not changed when we controlled for language ability, SES, or sex. Neither language ability, nor SES, nor sex accounted for the different associations between Bird/Alligator no-go scores, Bird/Alligator go scores, Shape Stroop scores, Grass/Snow, and CBQ–IC scores over time, with one exception: Bird/Alligator no-go scores were not associated with Shape Stroop scores at 36 months when controlling for language ability (r[356] = .03, p = .572). When controlling for language ability, the correlation between Bird/Alligator no-go scores and Shape Stroop scores at 36 months changed from $r = .14$ to .03. Thus, overall, the changing patterns of associations with development did not reflect extreme values or differences in language ability, SES, or sex.

Discussion

The change in patterns of association could not be explained by differences in SES or sex. Although the present sample had a range of SES, the preponderance of the sample was middle class, so associations with SES may have been somewhat attenuated. The change in patterns of association also could not be explained by children's differences in language ability, which suggests that the different associations over time could not be attributed to differences in the comprehension of task rules.

Language ability did appear to play a role in the association between Shape Stroop and Bird/Alligator no-go scores at 36 months. This provides another illustration of how the meanings of the inhibitory control measures appear to change across development. The age of 36 months may be an interesting epoch during which the child's inhibitory control has more to do with the child's language development at this age than six months earlier or six months later. Prior work has shown that language ability appears to be important for the development of inhibitory control (Petersen et al., 2015). This finding is also consistent with findings reported in a monograph edited by Espy (2016): Scores on inhibitory control measures at 3 years of age were more strongly associated with children's general cognitive abilities including language skills, and less strongly associated with a core, separate construct of inhibitory control than it was at later ages. One possibility is that inhibitory control becomes more efficient with practice across development. Early in development, inhibitory control is effortful and likely depends heavily on other systems (e.g., language). However, with development, these effortful processes may become embedded in the cognitive, neural, and behavioral systems in which inhibition is needed (within attention, behavioral control, etc.). Thus, as children more generally attain language skills with development, inhibitory control and its individual differences may emerge as separable from other systems.

Supplementary Materials S3. Analyses examining sex-related differences.

Results

There were modest sex-related differences in inhibitory control (see Table 1). Compared to boys, girls showed higher Shape Stroop scores at 36 months, higher Bird/Alligator no-go scores at 42 months, and higher CBQ–IC scores at all ages. By contrast, boys showed higher Bird/Alligator no-go scores than did girls at 30 months. Consistent with considerable prior research (e.g., Zambrana et al., 2012), girls showed greater language ability on average compared to boys ($t[1,294.80] = -2.75$, $p = .006$). Therefore, in a set of partial correlations, we examined whether the sex-related differences in inhibitory control remained when language ability was statistically controlled. Girls no longer showed significantly higher Shape Stroop scores than boys at 36 months ($r[403] = .07$, $p = .12$) or higher Bird/Alligator scores at 42 months ($r[387] = .08$, $p = .09$) at the $p < .05$ level, when controlling for language ability. Girls continued to show higher parent-reported CBQ–IC scores than boys at 30 ($r[495] = .23$, $p < .001$), 36 ($r[423] = .26$, $p < .001$), and 42 ($r[406] = .26$, $p < .001$) months and secondary caregiver-reported CBQ–IC scores than boys at 30 ($r[495] = .23$, $p < .001$), 36 ($r[423] = .26$, $p < .001$), and 42 ($r[406] = .26$, $p < .001$) months, even controlling for language ability. Boys continued to show higher Bird/Alligator no-go scores than girls at 30 months ($r[356] = -.14$, $p = .01$), even controlling for language ability.

In sum, girls showed better inhibitory control than boys for some measures at some ages. Girls' tendency to have better language ability than boys accounted for some but not all of the observed sex-related differences in inhibitory control. By contrast, boys showed higher Bird/Alligator no-go scores than girls at 30 months.

Discussion

We observed sex-related differences in inhibitory control, with girls showing modestly higher scores than boys, on average, on Shape Stroop at 36 months, Bird/Alligator no-go scores at 42 months, and on parents' and secondary caregivers' ratings of inhibitory control at each age. This set of results is consistent with prior work (Kochanska et al., 1997). We found that the apparent female advantages in inhibitory control could, in part, reflect their tendency to have better language ability compared to males, which could point to language skills as an important intervention target. However, in the present study, girls continued to show higher parent- and secondary caregiver-rated inhibitory control even when we controlled for language ability.
We also observed that boys tended to have modestly higher scores on Bird/Alligator no-go scores, but this sex-related difference was not accounted for by language ability. The male advantage in no-go scores at 30 months could reflect that girls were more likely than boys to activate on the go (and no-go) trials at 30 months. That is, girls had better scores on go trials at 30 months than boys. Interestingly, go scores at 30 months were more strongly associated (compared to no-go scores at that age) with no-go scores at 42 months (see Table 1), even controlling for language ability. Thus, activation performance, compared to inhibition performance, may at 30 months be more prognostic of later inhibitory control. Perhaps this is because activation is a more developmentally relevant yardstick at that age. That is, even though boys showed better inhibition on no-go trials at 30 months than girls, girls may have been more *competent* on the task overall in their ability to appropriately activate and behaviorally comply with prompts from the puppets, consistent with prior research that girls show higher compliance than boys (Kuczynski & Kochanska, 1990). Indeed, girls showed better inhibition on no-go trials than boys at a later age (42 months). This potential distinction between task performance and task competence (Sophian, 1997) is an intriguing finding that will be important for future

research to replicate and further examine to determine how performance should be interpreted in light of competence.

Supplementary Materials S4. Growth curve analyses.

Method

To examine mean-level growth in inhibitory control, we examined growth curve models in hierarchical linear modeling (HLM) using the nlme package (Pinheiro et al., 2009) in R (R Core Team, 2020). We examined quantile–quantile plots of the residuals from the growth curve models to examine whether the inhibitory control measures showed range restriction (Kaufman, 2013).

Results

Mean-Level Change

Results from the hierarchical linear modeling (HLM) growth curve models are provided in Table S6. On all four inhibitory control measures, children showed improvement with age, as would be expected by theory for measures of inhibitory control at these ages. Children showed particularly rapid growth in inhibitory control scores on the lab tasks ($.22 \leq \beta s \leq .49$).

Intra-Measure Correlations

We also examined the intra-measure correlations between the intercept and slope parameters of the HLM growth curve models (see Table 3). Among the intra-measure correlations, the intercepts and slopes of the lab tasks were strongly correlated ($|r\text{s}| > .70$), and the correlation between the intercepts and slopes for parent-reported CBQ–IC scores was at a trend level, and was small in magnitude. The direction of the correlations between intercepts and slopes differed by measure. For the Bird/Alligator go, Shape Stroop, and secondary caregiver-reported CBQ–IC scores, the correlations between intercepts and slopes were negative, indicating that children who scored higher initially showed lower rates of growth in scores. For the Bird/Alligator no-go, Grass/Snow, and parent-reported CBQ–IC scores, by contrast, the correlations between intercepts and slopes were positive, indicating that children who started at higher levels showed steeper growth. A negative correlation between intercepts and slopes might indicate a ceiling effect at later ages (i.e., the task may have been too easy), whereas a positive correlation between intercepts and slopes might indicate a floor effect at earlier ages (i.e., the task may have been too difficult).

Quantile–Quantile Plots of the Residuals

Quantile–quantile plots of the residuals from the growth curve models are depicted in Figure S1. The residuals were non-normally distributed in ways that were consistent with range restriction. The residuals for Bird/Alligator Go and Shape Stroop scores showed a convex bend, consistent with a negatively skewed distribution and a ceiling effect. The residuals for Bird/Alligator no-go scores were relatively "S"-shaped, consistent with heavy tails (i.e., more scores on the extremes than would be expected in a normal distribution) and range restriction. The flattened slope of the residuals (relative to the diagonal line) indicates a smaller standard deviation of residuals than would be expected in a normal distribution, consistent with range restriction. Furthermore, examination of the residuals as a function of age and performance on the inhibitory control task demonstrated that the residuals were heteroscedastic as a function of both age (i.e., the predictor in the growth curve models) and inhibitory control performance (i.e., the outcome). Variability of residuals decreased as a function of age and performance on the task for the Bird/Alligator go and Shape Stroop scores, consistent with a ceiling effect. Variability of residuals increased as a function of age and performance on the task for the Bird/Alligator no-go and Grass/Snow scores, consistent with a floor effect.

Inter-Measure Correlations

We next examined inter-measure associations of the intercept and slope parameters (see

Table 3). The intercepts of Bird/Alligator no-go scores were positively associated with the intercepts of Shape Stroop, Grass/Snow, and parent- and secondary caregiver-reported CBQ–IC scores, but were not associated with intercepts of Bird/Alligator go scores. The slopes of Bird/Alligator no-go scores were positively associated with the slopes of Grass/Snow and parent-reported CBQ–IC scores, demonstrating convergent validity of these measures; however, slopes of Bird/Alligator no-go scores were negatively associated with slopes of Shape Stroop and secondary caregiver-reported CBQ–IC scores. Intercepts and slopes of Bird/Alligator go scores, on the other hand, were less strongly associated with the intercepts and slopes of the other measures, demonstrating discriminant validity of the inhibitory control measures in relation to the Bird/Alligator go scores. The intercepts of the Shape Stroop and Grass/Snow scores did not significantly relate to one another, but their slopes were negatively associated, perhaps due to range restriction. The intercepts and slopes of the Shape Stroop and Grass/Snow scores were not significantly associated with their counterpart intercepts and slopes or with intercepts or slopes of parent-reported CBQ–IC scores. The intercepts of Shape Stroop, Grass/Snow, and parent-reported CBQ–IC scores were positively associated with intercepts of secondary caregiver-reported CBQ–IC scores. The slopes of Shape Stroop scores were positively associated with slopes of secondary caregiver-reported CBQ–IC scores. In sum, the pattern of inter-measure associations of the intercept and slope parameters across measures provided some evidence of convergent and discriminant validity of the measures, but some of the intercept and slope parameters were not associated in expected ways.

Discussion

Children showed particularly rapid growth in inhibitory control on the lab tasks, consistent with theory of inhibitory control development during early childhood (Goswami, 2011). The theoretically consistent forms of skill growth across 30 to 42 months provide further evidence that the measures show construct validity, at least at some ages. Regarding associations among the intercepts, we found that young children who started from a higher level of inhibitory control on the Bird/Alligator task generally did better also on the Shape Stroop and Grass/Snow inhibition tasks, and had higher ratings of inhibitory control by parents and secondary caregivers on the CBQ–IC, showing some convergent validity. In addition, children who started from a higher level of inhibitory as reported by secondary caregivers generally did better on the Shape Stroop and Grass/Snow tasks. However, the intercepts of the Shape Stroop, Grass/Snow, and parent-reported CBQ–IC scores did not significantly relate to one another.

Regarding associations among the slopes, we found that the more a child grew over a year in inhibitory control in the Bird/Alligator task, the more the child also grew in Grass/Snow inhibitory control and parents' and secondary caregivers' ratings of their inhibitory control, but the less the child grew in Shape Stroop inhibitory control. In addition, the more a child grew over a year in inhibitory control in the Shape Stroop task, the more the child also grew in secondary caregivers' ratings of their inhibitory control. The negative association between slopes of Shape Stroop scores and slopes of the Bird/Alligator and Grass/Snow tasks could reflect ceiling effects on the Shape Stroop task at later ages. Overall, the intercepts and slopes of Bird/Alligator no-go scores were more strongly associated with the intercepts and slopes of Grass/Snow, Shape Stroop, and CBQ–IC scores than they were with intercepts and slopes of Bird/Alligator go scores, consistent with convergent and discriminant validity. Thus, although the present results suggest that the meanings of these inhibitory control measures change over time, they do appear to be part of an enduring inhibitory control construct, at least by the older ages.

In the growth curve models, we observed some evidence of possible ceiling and floor effects, which is consistent with suggestions that perceptual inhibition (as assessed by Shape Stroop) may develop earlier than performance inhibition (as assessed by Bird/Alligator) and association inhibition (as assessed by Grass/Snow; Petersen et al., 2016). Thus, studies that have used the Grass/Snow task in children younger than 36 months of age may be difficult to interpret based on the present findings that suggest that inhibitory control may not be accurately assessed with this task at this age (Caughy et al., 2013; Petersen et al., 2015; Zalewski et al., 2012).

Supplementary Materials S5. Latent construct analyses.

Method

To determine whether the measures demonstrated longitudinal factorial invariance, we examined a latent inhibitory control construct using structural equation modeling (SEM) in Mplus 8.4 (Muthén & Muthén, 2019). Mplus implements full information maximum likelihood (FIML) estimation, which is a robust estimation method when data are missing at random or completely at random. We set the mean and variance of the latent factor to zero and one, respectively, which allowed the measures' factor loadings to vary freely. We also examined the association between the Bird/Alligator go scores and the latent inhibitory control construct as a further test of the cross-time associations between an index of activation and a latent construct of inhibition. Model fit was evaluated using Root Mean Square Error of Approximation (RMSEA < .08), Comparative Fit Index (CFI ≥ .95), and Standardized Root Mean Residual (SRMR ≤ .08) values, according to established cutoffs (Schreiber et al., 2006). We used maximum likelihood estimation with robust standard errors to account for non-normally distributed data.

Results

We attempted to fit a confirmatory factor analysis (CFA) model using structural equation modeling (SEM) to examine inhibitory control as a latent construct, composed of the common variance among the Bird/Alligator no-go, Shape Stroop, Grass/Snow, and CBQ–IC scores. We were unable to fit a converging model using a longitudinal CFA model, perhaps due to the age-dependent pattern of association between performance on Bird/Alligator, Shape Stroop, Grass/Snow, and CBQ–IC, and weak inter-measure associations at 30 months. Thus, we were unable to test longitudinal factorial invariance of the latent inhibitory control construct. Because we were unable to fit a longitudinal CFA model, we attempted to fit separate CFA models at each age. Intercepts and factor loadings are in Table 4.

At 30 months, the CFA model failed to converge because the Bird/Alligator no-go scores had a negative residual variance. This likely reflects that the inter-measure associations were weak to modest, resulting in the inability to estimate a latent variable from the limited common variance (i.e., secondary caregiver-reported CBQ–IC scores had a much stronger loading from the latent variable compared to Bird/Alligator no-go, Shape Stroop, Grass/Snow, and parent-reported CBQ–IC scores, which—with the exception of Shape Stroop—had loadings near zero). Moreover, the model did not converge even after removing any of the lab tasks.

The CFA model at 36 months is depicted in Figure 3. The CFA model converged and fit well according to SRMR (.054), but fit less well according to RMSEA (.085; 90% confidence interval: .058–.115) and CFI (.526). The factor loadings from the latent inhibitory control construct were positive for Bird/Alligator no-go scores ($B = 0.41$, $\beta = .34$, $p = .043$), Shape Stroop scores ($B = 0.14$, $\beta = .25$, $p = .020$), Grass/Snow scores ($B = 0.07$, $\beta = .22$, $p = .094$), parent-reported CBQ–IC scores ($B = 0.31$, $\beta = .39$, $p < .001$), and secondary caregiver-reported CBQ–IC scores ($B = 0.72$, $\beta = .67$, $p = .002$). Bird/Alligator go scores were not significantly associated with the latent inhibitory control construct ($B = 0.01$, $\beta = .03$, $SE = 0.23$, $p = .816$).

The CFA model at 42 months is depicted in Figure 3. The CFA model converged and fit well according to RMSEA (.064; 90% confidence interval: .035–.095) and SRMR (.046), but fit less well according to CFI (.801). The factor loadings from the latent inhibitory control construct were positive for Bird/Alligator no-go scores ($B = 0.74$, $\beta = .68$, $p < .001$), Shape Stroop scores ($B = 0.12$, $\beta = .29$, $p = .001$), Grass/Snow scores ($B = 0.08$, $\beta = .24$, $p = .002$), parent-reported CBQ–IC scores ($B = 0.26$, $\beta = .33$, $p < .001$), and secondary caregiver-reported CBQ–IC scores ($B = 0.47$, $\beta = .46$, $p < .001$). Bird/Alligator go scores showed a trend of a positive association with the latent inhibitory control construct (*B* = 0.06, β = .17, *SE* = 0.04, *p* = .068).

In sum, CFA models provided evidence that the measures coherently reflected a common construct at 36 and 42 months, but not 30 months. Because the CFA model converged at 36 and 42 months (but not 30 months), we attempted to fit a longitudinal CFA model across 36 to 42 months. However, we were unable to fit a converging model using a longitudinal CFA model across 36 to 42 months. Specifically, the estimated correlation between the latent inhibitory control factor at 36 and 42 months (i.e., rank-order stability) was above one ($\phi = 1.13$). Moreover, the estimated correlation of the latent factor across time continued to be above one even after allowing residuals of the same measure to be correlated across time (Little, 2013). Interestingly, prior work using longitudinal CFA has observed implausibly high cross-time correlations of executive functioning ($\phi = 1.0$), of which inhibitory control is a key component (Willoughby et al., 2017). At 36 and 42 months, with one exception (Grass/Snow at 36 months; β = .22, *p* = .094), all of the measures had significantly positive loadings from the construct.

Although we were unable to test longitudinal factorial invariance in the same model, examination of intercepts and factor loadings from CFA models at 36 and 42 months demonstrated differences across ages (see Table 4). For instance, the factor loadings of the performance-based tasks increased from 36 to 42 months. The Bird/Alligator task showed the strongest changes in factor loadings. By contrast, factor loadings of the parent- and secondary caregiver-reported CBQ–IC scores decreased from 36 to 42 months. These findings suggest that the measures change in their strength of association with the latent construct from ages 36 to 42 months. In addition, intercepts of the performance-based tasks increased from 36 to 42 months. The Bird/Alligator and Shape Stroop tasks showed the strongest changes in intercepts. These findings suggest that a higher score on the performance-based measures was needed at 42 months than 36 months to have the same level on the latent inhibitory control factor.

Discussion

The pattern of different associations by age was confirmed when we considered latent inhibitory control constructs. The latent inhibitory control construct did not converge at 30 months, likely due to the weak-to-modest inter-measure associations at this age. The challenges of using CFA-derived latent variables from weakly correlated tasks has also been documented in studies of executive function in children (Willoughby et al., 2017). However, latent constructs were estimated at 36 and 42 months, with each measure showing a positive loading from the latent construct. The measures appear to more consistently and coherently assess the construct of inhibitory control by 36 months of age. Further, the intercepts and factor loadings of inhibitory control measures appeared to change across time. In sum, the meanings of individual inhibitory control measures appear to change across time, consistent with heterotypic continuity, as depicted in Figure 1. This finding is also consistent with work on focused attention that has shown a more coherent factor structure by 3 years of age than at earlier ages (Acar et al., 2019). Based on evidence from the latent factor models, we see all four measures of inhibitory control as distinct but complementary ways of assessing the overarching inhibitory control construct.

References

Acar, I. H., Frohn, S., Prokasky, A., Molfese, V. J., & Bates, J. E. (2019). Examining the associations between performance based and ratings of focused attention in toddlers: Are we measuring the same constructs? *Infant and Child Development*, *28*(1), e2116. https://doi.org/10.1002/icd.2116

Caruso, J. C., & Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurement*, *57*(4), 637–654. https://doi.org/10.1177/0013164497057004009

Caughy, M. O. B., Mills, B., Owen, M. T., & Hurst, J. R. (2013). Emergent self-regulation skills among very young ethnic minority children: A confirmatory factor model. *Journal of Experimental Child Psychology*, *116*(4), 839–855. https://doi.org/10.1016/j.jecp.2013.07.017

Espy, K. A. (2016). The changing nature of executive control in preschool. *Monographs of the Society for Research in Child Development*, *81*(4). https://doi.org/10.1111/mono.12267

Goswami, U. (2011). *The Wiley-Blackwell handbook of childhood cognitive development*. Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*(2), 561–581. https://doi.org/10.1111/j.1540-5907.2010.00447.x

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1–47. https://doi.org/10.18637/jss.v045.i07

Kaufman, R. L. (2013). *Heteroskedasticity in regression: Detection and correction*. Sage Publications, Inc. https://doi.org/10.4135/9781452270128

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/10.1111/j.1467-8624.1997.tb01939.x

Kuczynski, L., & Kochanska, G. (1990). Development of children's noncompliance strategies from toddlerhood to age 5. *Developmental Psychology*, *26*(3), 398–408. https://doi.org/10.1037/0012-1649.26.3.398

Little, T. D. (2013). *Longitudinal structural equation modeling*. The Guilford Press. https://www.guilford.com/books/Longitudinal-Structural-Equation-Modeling/Todd-Little/9781462510160

Lumley, T. (2010). *mitools: Tools for multiple imputation of missing data*. https://cran.r-project.org/package=mitools

Muthén, L. K., & Muthén, B. O. (2019). *Mplus version 8.4*. Muthén & Muthén. https://www.statmodel.com/programs.shtml

Petersen, I. T., Bates, J. E., & Staples, A. D. (2015). The role of language ability and self-regulation in the development of inattentive-hyperactive behavior problems. *Development and Psychopathology*, *27*(1), 221–237. https://doi.org/10.1017/S0954579414000698

Petersen, I. T., Hoyniak, C. P., McQuillan, M. E., Bates, J. E., & Staples, A. D. (2016). Measuring the development of inhibitory control: The challenge of heterotypic continuity. *Developmental Review*, *40*, 25–71. https://doi.org/10.1016/j.dr.2016.02.001

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., and the R Core team. (2009). *nlme: Linear and nonlinear mixed effects models. R package version 3.1-93.* https://cran.r-project.org/package=nlme

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316696.fmatter

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (Vol. 72). Chapman & Hall/CRC. https://doi.org/10.1201/9781439821862

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–337. https://doi.org/10.3200/JOER.99.6.323-338

Sophian, C. (1997). Beyond competence: The significance of performance for conceptual development. *Cognitive Development*, *12*(3), 281–303. https://doi.org/10.1016/S0885-2014(97)90001-0

Willoughby, M. T., Kuhn, L. J., Blair, C. B., Samek, A., & List, J. A. (2017). The test–retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators. *Child Neuropsychology*, *23*(7), 822–837. https://doi.org/10.1080/09297049.2016.1205009

Zalewski, M., Lengua, L. J., Fisher, P. A., Trancik, A., Bush, N. R., & Meltzoff, A. N. (2012). Poverty and single parenting: Relations with preschoolers' cortisol and effortful control. *Infant and Child Development*, *21*(5), 537–554. https://doi.org/10.1002/icd.1759

Zambrana, I. M., Ystrom, E., & Pons, F. (2012). Impact of gender, maternal education, and birth order on the development of language comprehension: A longitudinal study from 18 to 36 months of age. *Journal of Developmental & Behavioral Pediatrics*, *33*(2), 146–155. https://doi.org/10.1097/DBP.0b013e31823d4f83

Table S1. Descriptive statistics.

Variable	Age (months)	*N*	*M*	*SD*	Lower	Upper
Bird/Alligator Go	30	364	2.43	0.65	2.39	2.46
Bird/Alligator No-Go	30	361	0.77	0.91	0.73	0.82
Shape Stroop	30	477	1.11	0.73	1.07	1.14
Grass/Snow	30	376	0.40	0.29	0.39	0.42
CBQ–IC: P	30	494	4.33	0.77	4.29	4.36
CBQ–IC: SC	30	232	4.45	1.17	4.38	4.53
Language Ability	30	487	48.92	8.14	48.55	49.29
SES	30	488	48.14	13.26	47.54	48.74
Bird/Alligator Go	36	370	2.69	0.48	2.67	2.72
Bird/Alligator No-Go	36	367	1.38	1.21	1.32	1.45
Shape Stroop	36	406	1.58	0.58	1.55	1.61
Grass/Snow	36	381	0.44	0.32	0.42	0.45
CBQ–IC: P	36	417	4.42	0.78	4.38	4.45
CBQ–IC: SC	36	205	4.68	1.08	4.60	4.75
Language Ability	36	417	52.55	8.16	52.15	52.95
SES	36	135	48.40	12.65	47.31	49.49
Bird/Alligator Go	42	390	2.82	0.37	2.80	2.84
Bird/Alligator No-Go	42	390	2.26	1.09	2.20	2.31
Shape Stroop	42	402	1.80	0.42	1.78	1.82
Grass/Snow	42	395	0.58	0.33	0.56	0.59
CBQ–IC: P	42	399	4.59	0.80	4.55	4.63
CBQ–IC: SC	42	213	4.79	1.02	4.72	4.86
Language Ability	42	402	54.31	7.67	53.93	54.70
SES	42	91	50.63	10.38	49.54	51.71

Note: “Lower” is the lower bound of the 95% confidence interval; “Upper” is the upper bound of the 95% confidence interval; “CBQ–IC: P” and “CBQ–IC: SC” are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children’s Behavior Questionnaire–Short Form; “SES” is the index of socioeconomic status. SES scores are presented disaggregated by age even though SES scores were averaged across ages for analysis. Some children ($n = 6$) had go scores but not no-go scores because they activated on one or a few go trials but did not participate on any of the trials thereafter.

Table S2. Descriptive statistics of language ability and SES among those who have inhibitory control scores (separated by measure).

Variable	Bird/Alligator		
	Age (months)	*M*	*SD*
Language Ability	30	50.35	7.42
SES	30	48.34	13.05
Language Ability	36	53.24	7.81
SES	36	48.45	12.47
Language Ability	42	54.48	7.34
SES	42	50.10	10.23

Variable	Shape Stroop		
Language Ability	30	49.11	7.96
SES	30	48.13	13.10
Language Ability	36	52.64	8.20
SES	36	48.31	12.77
Language Ability	42	54.36	7.69
SES	42	50.33	10.40

Variable	Grass/Snow		
Language Ability	30	50.58	7.12
SES	30	48.54	12.61
Language Ability	36	53.30	7.75
SES	36	48.86	11.93
Language Ability	42	54.52	7.47
SES	42	50.22	10.32

Variable	CBQ–IC: P		
Language Ability	30	48.94	8.04
SES	30	48.10	13.29
Language Ability	36	52.67	8.20
SES	36	48.54	12.68
Language Ability	42	54.48	7.60
SES	42	50.62	10.50

Variable	CBQ–IC: SC		
Language Ability	30	50.35	7.43
SES	30	49.84	11.04

Language Ability	36	53.42	7.76
SES	36	51.00	7.99
Language Ability	42	54.38	7.88
SES	42	50.78	9.29

Note: "CBQ IC: P" and "CBQ IC: SC" are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form; "SES" is the index of socioeconomic status. SES scores are presented disaggregated by age even though SES scores were averaged across ages for analysis.

Table S3. Percent of those with scores (out of those who had a laboratory visit), and the sex distribution by measure.

Variable	Age	% with scores (out of those who had a laboratory visit*)	% girls (of those with scores)
Bird/Alligator Go	30	72%	49%
Bird/Alligator No-Go	30	71%	49%
Shape Stroop	30	94%	46%
Grass/Snow	30	74%	47%
CBQ–IC: P	30	**	46%
CBQ–IC: SC	30	**	43%
Language Ability	30	96%	46%
SES	30	**	46%
Bird/Alligator Go	36	86%	48%
Bird/Alligator No-Go	36	85%	48%
Shape Stroop	36	94%	46%
Grass/Snow	36	89%	46%
CBQ–IC: P	36	**	46%
CBQ–IC: SC	36	**	46%
Language Ability	36	97%	46%
SES	36	**	45%
Bird/Alligator Go	42	94%	47%
Bird/Alligator No-Go	42	94%	47%
Shape Stroop	42	97%	46%
Grass/Snow	42	96%	46%
CBQ–IC: P	42	**	45%
CBQ–IC: SC	42	**	46%
Language Ability	42	97%	45%
SES	42	**	43%

Note: “Age” is the child’s age (in months); “CBQ IC: P” and “CBQ IC: SC” are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children’s Behavior Questionnaire–Short Form; “SES” is the index of socioeconomic status. SES scores are presented disaggregated by age even though SES scores were averaged across ages for analysis.

[*] Children who had a laboratory visit numbered 507, 430, and 413 at 30, 36, and 42 months, respectively.

** Not a performance-based measure; for *N*s of children with scores, see Table S1.

Table S4. Inter-rater reliability of lab tasks.

Variable	Videos with Two Coders	ICC
Bird/Alligator go scores	70%	.89
Bird/Alligator no-go scores	70%	.98
Shape Stroop scores	81%	.95
Grass/Snow scores	74%	.98

Note: “ICC” is the intra-class correlation coefficient, representing inter-rater reliability.

Table S5. Spearman correlation matrix of study variables.

			30 months							36 months							42 months						
	Sex	SES	Lang	BA Go	BA No-Go	SS	GS	PR	SCR	Lang	BA Go	BA No-Go	SS	GS	PR	SCR	Lang	BA Go	BA No-Go	SS	GS	PR	SCR
Sex	–																						
SES	-.01	–																					
Lang 30	.08†	.25***	–																				
BA Go 30	.18***	.01	.25***	–																			
BA No-Go 30	-.16***	.01	-.17***	-.46***	–																		
SS 30	.09†	.01	.39***	.14*	-.07	–																	
GS 30	.03	-.03	-.07	-.03	.12*	-.05	–																
PR 30	.22***	.03	.11*	.07	-.06	-.01	.01	–															
SCR 30	.20***	.11†	.14*	.00	.04	.02	-.05	.15*	–														
Lang 36	.08†	.14***	.65***	.25***	-.07	.30***	-.14*	.06	.21***	–													
BA Go 36	.14*	.02	.16***	.13*	-.07	.04	-.05	.05	-.04	.15*	–												
BA No-Go 36	.10†	.05	.19***	.07	.11†	.14*	.06	.14*	.22***	.26***	-.08	–											
SS 36	.12*	.04	.31***	.06	-.07	.23***	-.15*	.05	.22***	.35***	.12*	.11*	–										
GS 36	.05	-.02	-.01	.02	.04	.00	.08	.02	.08	.02	.01	.17***	-.02	–									
PR 36	.25***	.08	.10*	.11†	-.09	-.01	.06	.60***	.18*	.08	.02	.14*	.05	.05	–								
SCR 36	.23***	.03	.12	.18*	.01	.08	.12	.25***	.49***	.16*	-.02	.15†	.14†	.15†	.30***	–							
Lang 42	.10†	.14***	.55***	.24***	-.19***	.29***	-.12*	.05	.19*	.67***	.18***	.18***	.31***	-.06	.04	.08	–						
BA Go 42	.09†	.01	.12*	.08	-.03	.05	-.08	.01	.07	.17***	.14*	.12*	.14*	.00	.00	.01	.13*	–					
BA No-Go 42	.12*	.07	.29***	.26***	-.12†	.13*	-.04	.09†	.14†	.37***	.14*	.36***	.25***	.06	.11*	.16*	.37***	.15***	–				
SS 42	.09†	.00	.22***	.14*	.01	.25***	-.05	-.03	.13†	.24***	.10†	.02	.29***	-.01	-.07	.05	.26***	.11*	.19***	–			
GS 42	.06	.11*	.16***	.13*	.02	.17***	.01	-.05	.10	.22***	.00	.20***	.15***	.21***	.00	.08	.22***	.11*	.20***	.09†	–		
PR 42	.21***	.10†	.07	.01	-.07	-.05	.06	.55***	.11	.09	-.01	.11†	.11*	.10†	.67***	.28***	.07	.00	.20***	.00	.01	–	
SCR 42	.35***	.07	.10	.19*	.00	.00	.13	.26***	.45***	.15*	-.04	.24***	.11	.08	.18*	.59***	.20***	.16*	.16*	.12	.16*	.17*	–

Note. “SES” = socioeconomic status; “Lang” = language ability; “BA” = Bird/Alligator; “NG” = no-go; “SS” = Shape Stroop; “GS =

Grass/Snow"; "PR" and "SCR" are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children's Behavior Questionnaire–Short Form. Sex is coded with male = 0 and female = 1. *** $p < .001$; * $p < .05$; † $p < .10$; all *p*s two-tailed.

Table S6. Linear growth curve models.

Bird/Alligator go scores	B	β	SE	df	p
intercept	2.43	-.01	0.031	636	< .001
time	0.03	.31	0.003	636	< .001
Model Pseudo-R^2	.60				
Bird/Alligator no-go scores					
intercept	0.71	-.01	0.047	630	< .001
time	0.13	.49	0.006	630	< .001
Model Pseudo-R^2	.58				
Shape Stroop scores					
intercept	1.13	-.01	0.031	774	< .001
time	0.06	.44	0.003	774	< .001
Model Pseudo-R^2	.65				
Grass/Snow scores					
intercept	0.38	-.01	0.014	660	< .001
time	0.02	.22	0.002	660	< .001
Model Pseudo-R^2	.44				
CBQ–IC: P scores					
intercept	4.32	-.01	0.034	804	< .001
time	0.02	.13	0.003	804	< .001
Model Pseudo-R^2	.84				
CBQ–IC: SC scores					
intercept	4.48	.02	0.069	342	< .001
time	0.03	.13	0.007	306	< .001
Model Pseudo-R^2	.79				

Note: “CBQ–IC: P” and “CBQ–IC: SC” are parent and secondary caregiver reports, respectively, on the Inhibitory Control scale from the Children’s Behavior Questionnaire–Short Form. Linear growth curve models were fit in HLM. Time is centered so that intercepts are set at 30 months of age. Standard errors (*SE*s) are presented to three decimal places to distinguish between those with small values (SEs < 0.01).

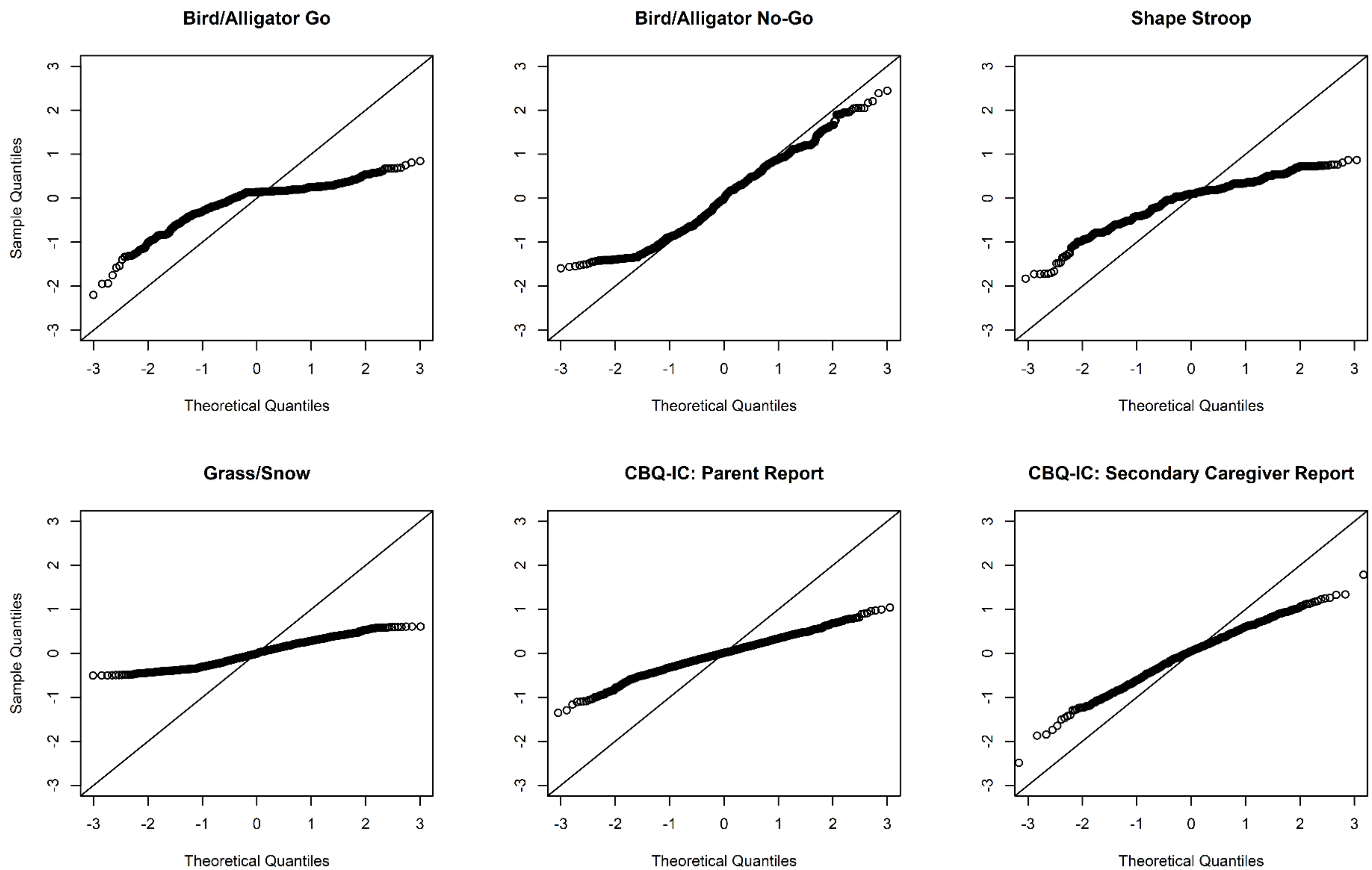


Figure S1. Quantile–quantile (Q–Q) plots of the residuals from the growth curve models (i.e., controlling for age). Q–Q plots depict the expected residuals for a normal distribution (x-axis) in relation to the actual residuals (y-axis). The diagonal line represents the comparison to a normal distribution. Examining a Q–Q plot can be helpful for determining whether residuals from the growth curve models are normally distributed, which is an assumption of the modeling approach. The plots above indicate that the residuals show some deviations from normality in terms of skewness (i.e., asymmetric distribution with more residuals to the left or right than would be expected in a normal distribution) and/or kurtosis (i.e., light or heavy tails). For more detailed discussion, see Supplementary Materials S4. “CBQ–IC” is the Inhibitory Control scale from the Children’s Behavior Questionnaire–Short Form.