Developmental Review 40 (2016) 25–71

Contents lists available at ScienceDirect

Developmental Review

journal homepage: www.elsevier.com/locate/dr

Measuring the development of inhibitory control: The challenge of heterotypic continuity

Isaac T. Petersen [a,*], Caroline P. Hoyniak [a], Maureen E. McQuillan [a], John E. Bates [a], Angela D. Staples [b]

[a] Indiana University, USA
[b] Eastern Michigan University, USA

ARTICLE INFO

Article history:
Received 5 March 2015
Revised 17 February 2016
Available online 10 March 2016

Keywords:
Inhibitory control
Self-regulation
Executive function
Meta-analysis
Changing measures
Longitudinal factorial invariance

ABSTRACT

Inhibitory control is thought to demonstrate heterotypic continuity, in other words, continuity in its purpose or function but changes in its behavioral manifestation over time. This creates major methodological challenges for studying the development of inhibitory control in childhood including construct validity, developmental appropriateness and sensitivity of measures, and longitudinal factorial invariance. We meta-analyzed 198 studies using measures of inhibitory control, a key aspect of self-regulation, to estimate age ranges of usefulness for each measure. The inhibitory control measures showed limited age ranges of usefulness owing to ceiling/floor effects. Tasks were useful, on average, for a developmental span of less than 3 years. This suggests that measuring inhibitory control over longer spans of development may require use of different measures at different time points, seeking to measure heterotypic continuity. We suggest ways to study the development of inhibitory control, with overlapping measurement in a structural equation modeling framework and tests of longitudinal factorial or measurement invariance. However, as valuable as this would be for the area, we also point

The Toddler Development Study has been funded by Indiana University and grants MH099437 from the National Institute of Mental Health and HD073202 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. Isaac T. Petersen was supported by Grant Number TL1 TR000162 (A. Shekhar, PI) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award, and by a Ruth L. Kirschstein National Research Service Award grant from the National Institute of Mental Health (1 F31 MH100814-01A1). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health or Indiana University.

* Corresponding author. Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, IN 47405, USA. Fax: +(812) 855-4691.
E-mail address: itpeters@indiana.edu (I.T. Petersen).

http://dx.doi.org/10.1016/j.dr.2016.02.001

out that establishing longitudinal factorial invariance is neither sufficient nor necessary for examining developmental change. Any study of developmental change should be guided by theory and construct validity, aiming toward a better empirical and theoretical approach to the selection and combination of measures.

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 in the domain of self-regulation (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). Self-regulation is a broad construct encompassing physiological, attentional, cognitive, emotional, and behavioral regulatory processes that promote adaptive or goal-directed behavior (Berger, 2011; Calkins & Fox, 2002). Self-regulation and inhibitory control measures have been associated with many important adjustment outcomes, including school readiness (Blair, 2002), health (Moffitt et al., 2011), and psychopathology (Dale & Baumeister, 1999), including in longitudinal studies (e.g., Brocki, Nyberg, Thorell, & Bohlin, 2007). Studies of self-regulation in children often include batteries of behavioral tasks thought to assess executive functioning, executive control, inhibitory control, response inhibition, behavioral regulation, and effortful control, among others (Carlson, 2005; Carlson & Moses, 2001; Kochanska, Murray, & Harlan, 2000). Despite the different terms used for similar measures, the different self-regulation constructs primarily reflect differences in research tradition rather than construct differences (Zhou, Chen, & Main, 2012). It has been difficult to achieve conceptual clarity among the different self-regulation constructs (McClelland & Cameron, 2012). The present review focuses on issues in measurement of inhibitory control, a key component of self-regulation across many research traditions, in longitudinal research. It is important to consider how to measure the development of inhibitory control because of the widespread use of inhibitory control measures in early childhood and its relevance to the development of psychopathology.

Inhibitory control, also referred to as response inhibition, is one of the separable cognitive processes thought to comprise the construct of executive function. In the past, inhibitory control was thought to emerge only in middle to late childhood, corresponding with improvements in the child's ability to execute complex, higher-order integrative tasks (Welsh, Friedman, & Spieker, 2006). However, response inhibition, in its most basic form, is present in the first year of life including the inhibition of neonatal reflexes and the inhibition of predominant behavioral reaching responses (Diamond, 1990), and more robustly in the preschool years. The vast majority of research on the development of inhibitory control often focuses on development from early childhood to adolescence (Durston et al., 2002; Williams, Ponesse, Schachar, Logan, & Tannock, 1999), with behavioral measures of inhibitory control demonstrating dramatic improvement over this time span (Welsh et al., 2006; Williams et al., 1999). The age at which performance on inhibitory control tasks reaches adult levels depends largely upon task complexity and difficulty, with performance on some basic tasks reaching adult levels in early childhood, while performance on other tasks, which require the integration of multiple executive functions, continues to improve until adolescence (for a review, see Garon, Bryson, & Smith, 2008).

Executive function has been defined as "goal-directed cognitive control of thought, action, and emotion" (Zelazo et al., 2013, p. 16). In adults, executive function is thought to be a multi-dimensional construct comprised of separable cognitive processes including response inhibition, working memory, and cognitive flexibility (Miyake et al., 2000). However, it is less clear whether such a factor structure characterizes executive function in early childhood, with some studies suggesting that executive function is best represented as a single, unitary construct in early childhood (Fuhs & Day, 2011; Shing, Lindenberger, Diamond, Li, & Davidson, 2010; Wiebe, Espy, & Charak, 2008; Wiebe et al., 2011; Willoughby, Blair, Wirth, & Greenberg, 2010; Willoughby, Wirth, & Blair, 2012) that likely fractures into a multi-dimensional construct in later childhood, with other studies suggesting that a multi-dimensional factor structure similar to adults is present in early childhood (Lee, Bull, & Ho, 2013; Lee et al., 2012; Miller, Giesbrecht, Müller, McInerney, & Kerns, 2012).

The behavioral manifestation of inhibitory control changes across development, consistent with the notion of heterotypic continuity. Heterotypic continuity is defined as the "continuity of an inferred genotypic attribute presumed to underlie diverse phenotypic behaviors" (Caspi, 2006, pp. 349–50) or as "the manifestation of the same underlying process through different behavioral presentations at different developmental periods" (Cicchetti & Rogosch, 2002, p. 13). In other words, the capacity for inhibitory control may appear different in toddlers compared to older children. For example, in toddlers, observed inhibitory control may include using words rather than physical aggression to get a toy. For older children, it may include the ability to inhibit an action based on an abstraction despite a concrete command (in the case of the game, "Simon Says") or despite social pressure (as for example when a peer invites a child to throw rocks at a window).

Measures of inhibitory control assess the capacity to withhold or delay a prepotent response or the capacity to perform a subdominant, yet correct response in the face of a conflicting and often prepotent (i.e., probabilistically dominant or habitual), yet incorrect response (Carlson & Moses, 2001). How does this important dimension of self-regulation develop? We think useful answers to this question will come from longitudinal studies. Longitudinal assessment of inhibitory control presents many challenges. The behavioral manifestation of inhibitory control changes across development, and measures must be valid for the developmental time frame of interest. Behavioral tasks, however, are likely to be appropriate for a restricted age range. The present meta-analytic review (1) determines the age ranges of usefulness for inhibitory control measures in childhood, (2) describes conceptual and methodological challenges in the longitudinal assessment of inhibitory control, and (3) offers recommendations for researchers studying the development of inhibitory control, with a hypothetical example of the longitudinal assessment of inhibitory control. Although this review focuses on the assessment of inhibitory control, the issues and recommendations may be common to the longitudinal assessment of many other domains and methodologies involving developmental change.

Measuring the development of inhibitory control: empirical considerations of age ranges of usefulness for inhibitory control measures

We first consider the age ranges for which common inhibitory control measures are useful for detecting individual differences in typically-developing children, focusing on measures of inhibitory control or response inhibition. Inhibitory control has been measured by response inhibition tasks that involve the withholding or delay of a prepotent response (for a review, see Garon et al., 2008). Consistent with Garon et al. (2008), we differentiate between simple response inhibition and complex inhibitory control paradigms. Simple response inhibition paradigms have minimal working memory demands, in which a child is given the opportunity to delay or the rule to inhibit a prepotent response. Examples of simple response inhibition tasks include delay of gratification (i.e., "Delay" tasks; Carlson & Moses, 2001) and "Don't" paradigms. By contrast, complex inhibitory control (i.e., "Conflict" tasks; Carlson & Moses, 2001) paradigms have greater working memory demands because children are instructed not only to inhibit a prepotent response, but also to respond in a certain way to a salient, conflicting response option (Carlson & Moses, 2001; Garon et al., 2008). Here we focus on complex inhibitory control tasks involving conflict rather than delay (Carlson & Moses, 2001) – specifically those tasks that require remembering a rule and responding according to the rule rather than to the competing, prepotent response (Garon et al., 2008). We chose to focus on complex inhibitory control rather than simple response inhibition tasks because (a) complex inhibitory control tasks tend to assess behavioral accuracy whereas simple response inhibition tasks often assess duration of delay, (b) we expected greater developmental growth up to 8 years of age in complex than simple tasks (thus avoiding ceiling effects), and (c) simple response inhibition tasks have a different motivational component than complex inhibitory control tasks and likely have different antecedents (Rochette & Bernier, 2014).

Examples of complex inhibitory control tasks involving conflict include: Baby Stroop, Bear/Dragon, Grass/Snow, Green-Red Signs, Day/Night, Hand Game, Knock-Tap, Less is More, Reverse Categorization, Shape Stroop, Simon Says, Spatial Conflict, and Tower (Carlson & Moses, 2001; Garon et al., 2008; Murray & Kochanska, 2002). Although the emphasis of the present review is on behavioral tasks, we also considered a questionnaire measure, the Inhibitory Control subscale of the Children's

Behavior Questionnaire (CBQ; Rothbart, Ahadi, & Hershey, 1994),[1] because of its relevance to the construct and because of the importance of incorporating multiple methods of assessment to estimate the amount of method variance (Strauss & Smith, 2009). Researchers have argued that, compared to structured lab tasks, naturalistic measures may more accurately reflect a child's adaptive executive functioning in real life (Barkley, 2012). There are a number of temperament questionnaires of analogous constructs (e.g., Behavior Rating Inventory of Executive Function; Gioia, Isquith, Guy, & Kenworthy, 2000). For a point of comparison to behavioral measures of inhibitory control, we included the CBQ because it is very widely used. The CBQ measure of temperamental inhibitory control reflects caregivers' impressions of child behavior across many relevant situations over long periods of time, unlike the structured lab tasks. Thus, the CBQ may be more consistent with a naturalistic approach to measurement of inhibitory control, and may be an important complement to behavioral tasks. See Table 1 for a description of these inhibitory control measures along with variants of each task that did not fundamentally alter the nature of the task. The table is not an exhaustive list of inhibitory control measures, but rather a list of some of the most prominent inhibitory control measures identified by Garon et al. (2008).[2]

In September 2014, we reviewed studies using the inhibitory control measures we selected (or similar variants; Table 1). Table 2 lists the studies we identified, the inhibitory control measures that each examined, and, if the study was excluded from the meta-analysis, the reason for exclusion. Table 3 lists which studies examined each measure. We conducted a meta-analysis of these previous studies to estimate the age range for which each measure is useful, based on the measures' average percentage-of-maximum (POM) score at each age (Little, 2013). We calculated POM for each study and inhibitory control measure by first scaling the measure's metric to have a meaningful zero point (i.e., subtracting the minimum possible score from the metric). Then, to create a proportion score, we divided the zero-scaled observed mean score by the zero-scaled maximum possible score (and multiplied by 100 to convert it to a percentage score). For studies that dichotomized task scores into pass/fail, we calculated POM as the percentage of cases that were given a score of "pass."

Inclusion criteria for studies in the meta-analysis included: (1) the study examined one or more of the inhibitory control measures from the list in Table 1 and (2) the sample included children younger than 8 years of age because inhibitory control develops most rapidly at younger ages (Diamond, 2002). Exclusion criteria included: (1) having identical data to (or fully subsumed within) that of another included paper, (2) not reporting POM or the information necessary to compute POM (mean performance and possible minimum and maximum) for unstandardized behavioral accuracy (not latency) on the inhibition trials of the individual task (as opposed to a composite of multiple measures or trial types), (3) including only a clinical sample, or (4) providing a single mean score for a sample of children whose ages spanned more than 3 years. We focused on scores on inhibition trials only (not go/activation trials) in order to ensure the scores were on a conceptually similar metric relating to inhibitory control. We excluded clinical subsamples from the full study sample if all children in the subsample had a clinical disorder, had low birth weight, or were born prematurely; however, we included control subsamples from these studies. We excluded studies providing a single mean score for ages spanning more than 3 years in order to have fine-grained estimates of developmental progression and because inhibitory control tasks have limited age ranges of utility (McClelland & Cameron, 2012). To avoid a

[1] We only included studies using the 6-item (Putnam & Rothbart, 2006) and 13-item (Rothbart et al., 1994) versions of the CBQ designed for 3- to 7-year-old children. We did not include studies using the Infant Behavior Questionnaire (3–12 months; Rothbart, 1981), Early Childhood Behavior Questionnaire (18–36 months; Putnam, Gartstein, & Rothbart, 2006), or the Temperament in Middle Childhood Questionnaire (7–10 years; Simonds, 2006), or their short, very short, or revised versions.

[2] Although Garon et al. identified the Detour-Reaching box task (Hughes & Russell, 1993) as a measure of complex inhibitory control, we excluded it because it has been less widely used with children than the other tasks we reviewed. We also excluded two tasks even though some researchers refer to them by the same name as two of the tasks in Table 1, because they are fundamentally different tasks. We excluded the Peg-Tapping task (Diamond & Taylor, 1996; Luria, 1966; sometimes referred to as Knock-Tap) in which the child uses a dowel or pencil to tap the table once when the experimenter taps twice and twice when the experimenter taps once. This task is conceptually different from knocking when the experimenter taps and vice versa (as in Knock-Tap). We also excluded the Fruit Distraction (Cammock & Cairns, 1979; Santostefano, 1988), Fruit Stroop (Archibald & Kerns, 1999), and Animal Stroop (Wright, Waterman, Prescott, & Murdoch-Eaton, 2003) tasks that are sometimes referred to as Shape Stroop because a small object was *not* embedded within a larger, different object.

Table 1
Inhibitory control measures (behavioral tasks and one questionnaire measure) with descriptions and developmentally-appropriate and -sensitive age ranges.

Measure	Source	Description	Dependent variable	Variants	Useful age range
Baby Stroop	Hughes and Ensor (2005)	Child says (spoon/cup belongs to) "mommy" when E shows small spoon or cup, but says "baby" when E shows large spoon or cup.	Number of correct responses		24–36 (26–36)
Bear/Dragon	Kochanska et al. (1996), adapted from Reed, Pien, and Rothbart (1984)	Child follows directions from the bear puppet, but ignores commands from the dragon puppet.	Number of trials child does not move in response to inhibition (dragon) trials	Bear/Alligator, Bear/Dog, Bear/Elephant, Bear/Lizard, Bird/Alligator, Bird/Dragon, Bunny/Dragon, Dog/Dragon, Hedgehog/Snake, Horse/Cow, Monkey/Dragon, Panda/Lion, Pig/Bull, Pig/Wolf, Policeman/Princess, Simon/Oscar, Squirrel/Badger	25–54 (33–54)
Day/Night	Gerstadt, Hong, and Diamond (1994)	Child says "day" when E shows picture of moon, "night" to picture of sun.	Number of correct responses	Similar to Grass/Snow, but involves verbal response: Sun/Moon, Car/Boat, Table/Chair, Fork/Knife, Cat/Dog, Happy/Sad, Fat/Thin, Bird/Dragon, Yes/No, Boy/Girl, Big/Little, Up/Down, Red/Blue, Car/Book, Black/White, Mommy/Me	32–71 (33–71)
Grass/Snow	Carlson and Moses (2001)	Child points to white square when E says "grass," green square when E says "snow."	Number of correct responses	Similar to Day/Night, but involves pointing: Sun/Moon, Car/Boat, Table/Chair, Fork/Knife, Red/Blue, Rain/Snow, Mommy/Me	28–69 (30–69)
Green-Red Signs	Kochanska et al. (1997)	Child raises same hand as E (left or right) when E holds up green sign, child raises opposite hand as E when E holds up red sign.	Number of correct responses on inhibition (red) trials		52–71 (none)
Hand Game	Hughes (1996, 1998b), based on Luria, Pribram, and Homskaya (1964)	Child makes opposite gesture as E (fist or pointed finger).	Number of correct responses		29–63 (31–63)
Inhibitory Control (CBQ)	Rothbart et al. (13 items; 1994), Putnam and Rothbart (6 items; 2006)	"Capacity to plan and to suppress inappropriate approach responses under instructions or in novel or uncertain situations." (Rothbart et al., 2001, p. 1406)	e.g., "Can lower his/her voice when asked to do so." (Rothbart et al., 2001, p. 1406)	6 items, 13 items	24–96 (26–96); designed for 36–84
Knock-Tap	NEPSY: Korkman, Kirk, and Kemp (1998), and Klenberg, Korkman, and Lahti-Nuuttila (2001)	Child taps table with open palm when E knocks table with fist, knocks table with fist when E taps table with open palm.	Number of correct responses	Child responds with an opposite action to one action of the E and does not respond at all to the second action of the E	36–61 (40–61)

(*continued on next page*)

Table 1 (*continued*)

Measure	Source	Description	Dependent variable	Variants	Useful age range
Less is More	Carlson et al. (2005)	Child chooses between small and large trays of treats, child receives the treat tray not pointed to.	Number of small tray selections	Jelly Beans, chocolate chips, rocks, dots, mouse/elephant, stickers, marbles, goldfish, cereal. The game has been played with the child by themselves, in competition with an experimenter, on a team with an experimenter, or with a stuffed animal (monkey, bear). It has also been played with mouse/elephant stimuli instead of trays.	23–75 (35–71)
Reverse Categorization	Carlson et al. (2004)	Child sorts objects into two categories using a sorting rule (e.g. sorting red blocks into a blue bucket and blue blocks into a red bucket) that is opposite a prepotent sorting rule (e.g. sorting the red blocks into the red bucket and sorting the blue blocks into the blue bucket).	Number of correct responses	Goldfish/Teddy Grahams, mommy/baby animals, large/small objects, green apples/ green bananas, green pigs/green fish	18–60 (20–60)
Shape Stroop	Kochanska et al. (1997)	Child points to pictures of small objects embedded within pictures of larger, different objects.	Number of correct responses	Fruit, shapes, animals, letters, numbers	20–43 (20–43)
Simon Says	Strommen (1973)	One E gives commands. Child performs action only when command is preceded by "Simon says."	Number of trials child does not move in response to inhibition (non-Simon) trials	E performs action, child faces away from E	58–86 (58–86)
Spatial Conflict	Gerardi-Caulton (2000)	Target stimuli (associated with different response keys) are presented on the left or right side of a monitor, and the child responds by selecting the correct key. In some trials the target stimulus is on the same side (left or right) as the correct key (congruent); on conflict trials the target is on the opposite side from the correct key (incongruent).	Number of correct incongruent trials	Cat/Dog	24–36 (24–36)
Tower	Kochanska et al. (1996)	Child takes turns placing blocks to build a tower with the E. E waits to take turn until the child signals to give E a turn.	Number of turns child yields to E		23–56 (24–56)

Note: E = experimenter. Table adapted from Garon et al. (2008). Ages in months. Age ranges in parentheses reflect age ranges with stronger empirical support (two or more studies).

Table 2
Studies reviewed using the inhibitory control measures from Table 1 (or similar variants).

Numeric index	Study year	Study authors	Inhibitory control measures	Exclusion criterion
1	1973	Strommen	SI	0
2	1977	Camp	SI	0
3	1981	LaVoie, Anderson, Fraze, & Johnson	SI	0
4	1984	Reed, Pien, & Rothbart	BD	0
5	1993	Ahadi, Rothbart, & Ye	IC	0
6	1994	Gerstadt, Hong, & Diamond	DN	0
7	1994	Rothbart, Ahadi, & Hershey	IC	2
8	1996	Diamond & Taylor	DN	0
9	1996	Hughes	HG	0
10	1996	Kochanska, Murray, Jacques, Koenig, & Vandegeest	BD, IC, TO	0, 0, 2
11	1996	Korkman, Liikanen, & Fellman	KT	4
12	1997	Kochanska, Murray, & Coy	GR, IC, SH, SI	0, 0, 0, 0
13	1998a	Hughes	HG	0
14	1998b	Hughes	HG	0
15	1998	Hughes, Dunn, & White	HG	0
16	2000	Berger, Jones, Rothbart, & Posner	SC	0
17	2000	Cole & Mitchell	BD	0
18	2000	Gerardi-Caulton	IC, SC, TO	2, 0, 2
19	2000	Kochanska, Murray, & Harlan	IC, SH, TO	0, 0, 2
20	2001	Carlson & Moses	BD, DN, GS, IC, SC, TO	0, 0, 0, 2, 0, 0
21	2001	Diamond	DN	1
22	2001	Floyd & Kirby	BD, TO	0, 0
23	2001	González, Fuentes, Carranza, & Estévez	IC	0
24	2001	Klenberg, Korkman, & Lahti-Nuuttila	KT	0
25	2001	Kochanska, Coy, & Murray	BD, GR, IC, SI, TO	2, 2, 0, 2, 2
26	2001	Rothbart, Ahadi, Hershey, & Fisher	IC	0
27	2002	Alexander, Goodman, Schaaf, Edelstein, Quas, & Shaver	DN, IC	2, 2
28	2002	Carlson, Moses, & Breton	BD	0
29	2002	Diamond, Kirkham, & Amso	DN	0
30	2002	Lang & Perner	HG	2
31	2002	Lengua	IC, SI	4, 4
32	2002	Murray & Kochanska	BD, GR, SH, SI, TO	1, 1, 1, 1, 0
33	2002	Perner, Kain, & Barchfeld	KT	0
34	2002	Sonuga-Barke, Dalen, Daley, & Remington	BD	0
35	2003	Blair & Peters	DN	2
36	2003	Deák & Narasimham	DN	0
37	2003	Fahie & Symons	HG	4
38	2003	Hala, Hug, & Henderson	DN	0
39	2003	Jones, Rothbart, & Posner	BD, IC	0, 2
40	2003	Kochanska & Knaack	BD, GR, SH, TO	2, 2, 2, 2
41	2003	Lengua	IC, SI	4, 4
42	2003	Peskin & Ardino	DN	0
43	2003	Rothbart, Ellis, Rueda, & Posner	SC	0
44	2003	Walker	BD, DN, GS, TO	0, 0, 2, 0
45	2004	Aksan & Kochanska	BD	2
46	2004	Blair, Peters, & Granger	IC	2
47	2004	Bohm, Smedler, & Forssberg	HG, KT	2, 2
48	2004	Carlson, Mandell, & Williams	BD, GS, HG, IC, RC, SH, TO	0, 0, 0, 0, 0, 0, 2
49	2004	Carlson, Moses, & Claxton	BD	0
50	2004	Dalen, Sonuga-Barke, Hall, & Remington	BD	2
51	2004	Flynn, O'Malley, & Wood	HG	0
52	2004	Jahromi	DN, IC	0, 0
53	2004	Joseph & Tager-Flusberg	DN, KT	3, 3
54	2004	Simpson, Riggs, & Simon	DN, GS	0, 0
55	2004	Wolfe & Bell	DN, IC	0, 0
56	2005	Berwid, Curko Kera, Marks, Santra, Bender, & Halperin	DN	0

(*continued on next page*)

Table 2 (*continued*)

Numeric index	Study year	Study authors	Inhibitory control measures	Exclusion criterion
57	2005	Carlson	BD, DN, GS, HG, LM, RC, SC, SH, SI, TO	0, 0, 0, 0, 0, 0, 0, 0, 0, 0
58	2005	Carlson, Davis, & Leach	BD, GS, LM, TO	2, 2, 0, 2
59	2005	Hughes & Ensor	BS	2
60	2005	Joseph, McGrath, & Tager-Flusberg	DN, KT	4, 4
61	2005	Lengua & Kovacs	IC	4
62	2005	Monks, Smith, & Swettenham	DN	2
63	2005	Olson, Sameroff, Kerr, Lopez, & Wellman	IC, TO	0, 0
64	2005	Pears & Fisher	DN	2
65	2005	Roberts & Powell	DN	0
66	2005	Roebers & Schneider	BD, DN, GS, HG	2, 2, 2, 2
67	2005a	Simpson & Riggs	DN	0
68	2005b	Simpson & Riggs	DN	0
69	2005	Tager-Flusberg & Joseph	DN, KT	3, 3
70	2006	Chasiotis, Kiessling, Hofer, & Campos	BD, DN, HG	0, 0, 0
71	2006	Chasiotis, Kiessling, Winter, & Hofer	BD, DN, HG	0, 0, 0
72	2006	Cromer, Stevens, DePrince, & Pears	KT	2
73	2006	Dennis	IC	0
74	2006	Hoff Esbjørn, Hansen, Greisen, & Mortensen	KT	0
75	2006	Hughes & Ensor	BS	2
76	2006	Livesey, Keen, Rouse, & White	DN	0
77	2006	Melinder, Endestad, & Magnussen	DN	4
78	2006	Pellicano, Maybery, Durkin, & Maley	HG	4
79	2006	Putnam & Rothbart	IC	0
80	2006	Sabbagh, Moses, & Shiverick	BD	0
81	2006	Sabbagh, Xu, Carlson, Moses, & Lee	BD, DN, GS, TO	2, 2, 2, 2
82	2006	Simpson & Riggs	DN	0
83	2006	Thorell & Wåhlstedt	DN	0
84	2007a	Bell & Wolfe	DN	1
85	2007b	Bell & Wolfe	DN	0
86	2007	Bernstein, Atance, Meltzoff, & Loftus	BD, DN	0, 0
87	2007	Bibok	SH, RC	0, 0
88	2007	Blair & Razza	IC	2
89	2007	Brocki, Nyberg, Thorell, & Bohlin	DN, KT	2, 2
90	2007	Carlson & Wang	SI	0
91	2007	Domitrovich, Cortes, & Greenberg	DN	0
92	2007	Flynn	BD, HG	0, 0
93	2007	Hinnant & O'Brien	DN	0
94	2007	Hrabok, Kerns, & Müller	SC	0
95	2007	Hughes & Ensor	BS, DN, HG	0, 0, 0
96	2007	Kochanska, Aksan, Penney, & Doobay	BD, GR, GS, SH, SI, TO	2, 2, 2, 2, 2, 2
97	2007	Lengua, Honorado, & Bush	BD, DN, GS	0, 0, 0
98	2007	Lewis, Dozier, Ackerman, & Sepulveda-Kozakowski	DN	0
99	2007	Marlow, Hennessy, Bracewell, Wolke, & The EPICure Study Group	KT	0
100	2007	Pellicano	HG	0
101	2007	Simpson & Riggs	DN	0
102	2007	Smith-Donald, Raver, Hayes, & Richardson	TO	0
103	2007	Tardif, So, & Kaciroti	DN, HG	0, 0
104	2007	Visu-Petra, Benga, & Miclea	KT	0
105	2007	von Stauffenberg & Campbell	DN	0
106	2007a	Wolfe & Bell	DN, IC	1, 2
107	2007b	Wolfe & Bell	BD, DN, IC	0, 0, 0
108	2008	Albertson & Shore	BD, GS	2, 0
109	2008	Bolnick	BD, IC, KT, SI	0, 0, 0, 0
110	2008	Brooker	TO	0
111	2008	Carlson & Meltzoff	IC, SI	2, 0
112	2008	Holmboe, Pasco Fearon, Csibra, Tucker, & Johnson	SC	0
113	2008	Hughes & Ensor	BS, DN, HG	2, 2, 2

(*continued on next page*)

Table 2 (*continued*)

Numeric index	Study year	Study authors	Inhibitory control measures	Exclusion criterion
114	2008	Jennings, Sandberg, Kelley, Valdes, Yaggi, Abrew, & Macey-Kalcevic	TO	2
115	2008	Karpinski, Scullin, & Montgomery-Downs	DN, HG	0, 0
116	2008	Kloo, Perner, Kerschhuber, Dabernig, & Aichhorn	RC	0
117	2008	Kochanska, Aksan, Prisco, & Adams	BD, GR, GS, TO	2, 2, 2, 2
118	2008	Leerkes, Paradise, O'Brien, Calkins, & Lange	DN	0
119	2008	Lunkenheimer, Dishion, Shaw, Connell, Gardner, Wilson, & Skuban	IC	0
120	2008	Montgomery, Anderson, & Uhl	DN	0
121	2008	Oh & Lewis	DN, GS, HG, TO	0, 0, 0, 2
122	2008	Putnam, Rothbart, & Gartstein	IC	0
123	2008	Ryan & Nelson	DN	2
124	2008	Sinn, Bryan, & Wilson	KT	4
125	2008	Talwar & Lee	DN	0
126	2008	Wellman, Lopez-Duran, LaBounty, & Hamilton	BD	0
127	2009	Beck, Riggs, & Gorniak	BD, DN	0, 0
128	2009	Bibok, Carpendale, & Müller	RC, SH	2, 2
129	2009a	Hughes & Ensor	BS	2
130	2009b	Hughes & Ensor	DN	0
131	2009	Hughes, Ensor, Wilson, & Graham	DN	0
132	2009	Ito	BD, GS, LM	0, 0, 0
133	2009	Kegel, van der Kooy-Hofland, & Bus	DN	0
134	2009	Kochanska, Barry, Jimenez, Hollatz, & Woodard	GR, GS, SS	2, 2, 2
135	2009	Lackner	GS, HG, LM	0, 0, 0
136	2009	Rhoades, Greenberg, & Domitrovich	DN	0
137	2009	Sabbagh, Bowman, Evraire, & Ito	BD, GS, LM	0, 0, 0
138	2009	Siegal, Iozzi, & Surian	DN	0
139	2009	Simpson & Riggs	DN, GS	0, 0
140	2009	Tregay, Gilmour, & Charman	HG	0
141	2010	Allan	GS, LM	1, 0
142	2010	Bernier, Carlson, Bordeleau, & Carrier	BS, SH	0, 0
143	2010	Bernier, Carlson, & Whipple	BS, SH	0, 0
144	2010	Bigham	HG	0
145	2010	Cutuli, Herbers, Rinaldi, Masten, & Oberg	SI	0
146	2010	Cutuli, Wiik, Herbers, Gunnar, & Masten	SI	4
147	2010	Dennis, Hong, & Solomon	IC	0
148	2010	Eisenberg, Vidmar, Spinrad, Eggum, Edwards, Gaertner, & Kupfer	IC	0
149	2010	Hayden, Klein, Sheikh, Olino, Dougherty, Dyson, Durbin, & Singh	TO	2
150	2010	Huyder	SI	1
151	2010	Kesek	IC	2
152	2010	Kloo, Perner, & Giritzer	DN	0
153	2010	Kolnik	DN, HG, SI	0, 2, 2
154	2010	Lackner, Bowman, & Sabbagh	GS, LM	0, 0
155	2010	Lane, Wellman, Olson, LaBounty, & Kerr	GR, SH, SI	2, 0, 0
156	2010	Molfese, Molfese, Molfese, Rudasill, Armstrong, & Starkey	KT	0
157	2010	Nampijja, Apule, Lule, Akurut, Muhangi, Elliott, & Alcock	KT	0
158	2010	Obradović	SI	0
159	2010	Pellicano	HG	3
160	2010	Prado, Hartini, Rahmawati, Ismayani, Hidayati, Hikmah, Muadz, Apriatni, Ullman, Shankar, & Alcock	KT, TO	2, 2
161	2010	Qu, Audrey, Jun, & Qun	LM	0
162	2010	Rakoczy	BD	0
163	2010	Sabbagh, Hopkins, Benson, & Randall Flanagan	GS	0

(*continued on next page*)

Table 2 (*continued*)

Numeric index	Study year	Study authors	Inhibitory control measures	Exclusion criterion
164	2010	Sulik, Huerta, Zerr, Eisenberg, Spinrad, Valiente, Di Gunta, Pina, Eggum, Sallquist, Edwards, Kupfer, Lonigan, Phillips, Wilson, Clancy-Menchetti, Landry, Swank, Assel, & Taylor	BD, IC, KT	0, 0, 0
165	2010	Willoughby, Blair, Wirth, & Greenberg	DN, SC	1, 1
166	2011	Allan & Lonigan	GS, LM	0, 0
167	2011	Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood	BD, RC	0, 0
168	2011	Beck, Schaefer, Pang, & Carlson	RC	2
169	2011	Blair, Granger, Willoughby, Mills-Koonce, Cox, Greenberg, Kivlighan, & Fortunato	SC	0
170	2011	Chang, Olson, Sameroff, & Sexton	IC, TO	1, 1
171	2011	Cutuli	SI	1
172	2011	Drayton, Turley-Ames, & Guajardo	BD, GS	0, 0
173	2011	Evans, Xu, & Lee	BD, DN, TO	0, 0, 0
174	2011	Fuhs & Day	DN	0
175	2011	Geng, Hu, Wang, & Chen	IC	0
176	2011	Gleason, Fox, Drury, Smyke, Egger, Nelson, Gregas, & Zeanah	BD	2
177	2011	Gusdorf, Karreman, van Aken, Deković, & van Tuijl	IC, SH, TO	0, 0, 0
178	2011	Herbers, Cutuli, Lafavor, Vrieze, Leibel, Obradović, & Masten	SI	1
179	2011	Hughes, Ensor, & Marks	BS, DN	0, 0
180	2011	Kegel	DN	0
181	2011	King	BD	0
182	2011	Lagattuta, Sayfan, & Monsour	DN	0
183	2011	Matte-Gagné & Bernier	BS, BD, DN, SH	2, 2, 0, 2
184	2011	Monette, Bigras, & Guay	DN, KT	0, 0
185	2011	Olson, Lopez-Duran, Lunkenheimer, Chang, & Sameroff	IC, TO	1, 1
186	2011	Olson, Tardif, Miller, Felt, Grabell, Kessler, Wang, Karasawa, & Hirabayashi	DN, GS, HG	0, 0, 0
187	2011	Phillips, Ruhl, Montague, Gasparovic, Caprihan, Ohls, Schrader, & Lowe	BD	2
188	2011	Poulin-Dubois, Blaye, Coutya, & Bialystok	RC, SH	0, 0
189	2011	Qu	LM	0
190	2011	Raver, Jones, Li-Grining, Zhai, Bub, & Pressler	SI	2
191	2011	Reck & Hund	BD, DN, IC	4, 4, 4
192	2011	Rhoades, Greenberg, Lanza, & Blair	SC	1
193	2011	Sulik	BD, KT	0, 0
194	2011	Thompson	BD, DN	0, 0
195	2011	Wellman, Lane, LaBounty, & Olson	GR, SH, SI	1, 1, 1
196	2011	White, McDermott, Degnan, Henderson, & Fox	DN, GS, IC	0, 0, 0
197	2011	Wiebe, Sheffield, Nelson, Clark, Chevalier, & Espy	SH	0
198	2012	Bellagamba, Laghi, Lonigro, & Pace	RC	1
199	2012	Bernier, Carlson, Deschênes, & Matte-Gagné	BS, BD, DN, SH	1, 2, 0, 0
200	2012	Calderon, Angeard, Moutier, Plumet, Jambaqué, & Bonnet	KT	0
201	2012	Carroll, Riggs, Apperly, Graham & Geoghegan	GS	0
202	2012	Conway & Stifter	DN	0
203	2012	Cuevas, Hubble, & Bell	GS, IC	0, 0
204	2012	Duncan	DN	0
205	2012	Duvall	BD, RC	0, 0
206	2012	Dyson, Olino, Durbin, Goldsmith, & Klein	TO	2
207	2012	Ford, Driscoll, Shum, & Macaulay	BD, DN, HG	0, 0, 0
208	2012	Hammond, Müller, Carpendale, Bibok, & Liebermann-Finestone	BD, HG, RC, SH	2, 1, 1, 0
209	2012	Hardaway, Wilson, Shaw, & Dishion	IC	0
210	2012	Huyder & Nilsen	SI	4
211	2012	Kim & Kochanska	TO	2

(*continued on next page*)

Table 2 (*continued*)

Numeric index	Study year	Study authors	Inhibitory control measures	Exclusion criterion
212	2012	Lackner, Sabbagh, Hallinan, Liu, & Holden	GS, HG, LM	1, 1, 1
213	2012	Lahat, Degnan, White, McDermott, Henderson, Lejuez, & Fox	DN, GS	0, 0
214	2012	Masten, Herbers, Desjardins, Cutuli, McCormick, Sapienza, Long, & Zelazo	SI	0
215	2012	Monette	DN, KT	1, 1
216	2012	Müller, Liebermann-Finestone, Carpendale, Hammond, & Bibok	BD, HG, RC, SH	2, 0, 0, 0
217	2012	Nilsen & Graham	DN, GS	0, 0
218	2012	Röthlisberger, Neuenschwander, Cimeli, Michel, & Roebers	GS, HG, SI	2, 2, 2
219	2012	Sage	DN, GS	0, 0
220	2012	Smith, Sheikh, Dyson, Olino, Laptook, Durbin, Hayden, Singh, & Klein	TO	2
221	2012	Steele, Karmiloff-Smith, Cornish, & Scerif	SC	0
222	2012	van der Kooy-Hofland, van der Kooy, Bus, van Ijzendoorn, & Bonsel	DN	0
223	2012	Viterbori, Gandolfi, & Usai	TO	0
224	2012	Willoughby, Wirth, & Blair	DN, SD	2, 2
225	2012	Zalewski, Lengua, Fisher, Trancik, Bush, & Meltzoff	BD, DN, GS	0, 0, 0
226	2013	Allan, Lonigan, & Wilson	GS, IC, LM	4, 4, 4
227	2013	Bellagamba, Laghi, Lonigro, Pace, & Longobardi	RC	0
228	2013	Benson, Sabbagh, Carlson, & Zelazo	BD, GS	0, 0
229	2013	Bernier, Beauchamp, Bouvette-Turcot, Carlson, & Carrier	BS, SH	1, 1
230	2013	Brown, Ackerman, & Moore	BD, DN	0, 0
231	2013	Carranza, González-Salinas, & Ato	IC	0
232	2013	Caughy, Mills, Owen, & Hurst	GS, SH	0, 0
233	2013	Choe, Olson, & Sameroff	GR, IC, SH, SI, TO	1, 1, 1, 1, 1
234	2013	Choe, Lane, Grabell, & Olson	TO	1
235	2013	Cipriano-Essel, Skowron, Stifter, & Teti	DN, SH	2, 2
236	2013	Clark, Sheffield, Wiebe, & Espy	SH	0
237	2013	Eisenberg, Edwards, Spinrad, Sallquist, Eggum, & Reiser	BD, GS, IC, KT	0, 0, 1, 0
238	2013	Esposito, Baker-Ward, & Mueller	DN	0
239	2013	Evans & Lee	RC, SH	0, 0
240	2013	Kim, Nordling, Yoon, Boldt, & Kochanska	DN, GS, GR, TO	0, 0, 0, 2
241	2013	Kraybill	BD, DN, SH	0, 0, 0
242	2013	Kraybill & Bell	BD, DN	2, 2
243	2013	Kryski, Dougherty, Dyson, Olino, Laptook, Klein, & Hayden	TO	2
244	2013	Lane, Wellman, Olson, Miller, Wang, & Tardif	DN, GS, HG	0, 0, 0
245	2013	Leve, DeGarmo, Bridgett, Neiderhiser, Shaw, Harold, Natsuaki, Reiss	SH	0
246	2013	Lunkenheimer, Albrecht, & Kemp	TO	0
247	2013	Macaulay & Ford	DN, GS, HG	2, 2, 2
248	2013	Moran, Lengua, & Zalewski	BD, DN	1, 1
249	2013	Orta, Corapci, Yagmurlu, & Aksan	BD, DN	2, 2
250	2013	Pauli-Pott, Dalir, Mingebach, Roller, & Becker	BD, DN, TO	0, 0, 0
251	2013	Pellicano	HG	3
252	2013	Riggs, Jolley, & Simpson	BD	0
253	2013	Smith, Blake, & Harris	BD, DN	4, 4
254	2013	Smith, Kryski, Sheikh, Singh, & Hayden	TO	2
255	2013	Sulik, Eisenberg, Silva, Spinrad, & Kupfer	BD, KT	1, 0
256	2013	Thompson, Lengua, Zalewski, & Moran	BD, DN	1, 1
257	2013	Watson & Bell	DN, HG, IC, LM	0, 0, 0, 0
258	2013	Wyss	RC	0
259	2014	Allan & Lonigan	BD, DN, GS, KT, SH	4, 4, 4, 4, 4
260	2014	Calderon, Jambaqué, Bonnet, & Angeard	DN, HG	0, 0

(*continued on next page*)

Table 2 (*continued*)

Numeric index	Study year	Study authors	Inhibitory control measures	Exclusion criterion
261	2014	Causey & Bjorklund	BD, DN	0, 0
262	2014	Chien	DN, SI	0, 0
263	2014	Choe, Olson, & Sameroff	IC, TO	1, 1
264	2014	Choe, Shaw, Brennan, Dishion, & Wilson	IC	0
265	2014	Cuevas & Bell	BD, DN	0, 0
266	2014	Cuevas, Deater-Deckard, Kim-Spoon, Wang, Morasch, & Bell	BD, DN	0, 0
267	2014	Cuevas, Deater-Deckard, Kim-Spoon, Watson, Morasch, & Bell	BD, DN	1, 1
268	2014	Cutuli, Herbers, Lafavor, Ahumada, Masten, & Oberg	SI	1
269	2014	Earhart & Roberts	DN, HG	0, 0
270	2014	Fizke, Barthel, Peters, & Rakoczy	BD, DN, GS	0, 0, 0
271	2014	Herbers, Cutuli, Monn, Narayan, & Masten	SI	1
272	2014	Herbers, Cutuli, Supkoff, Narayan, & Masten	SI	1
273	2014	Gandolfi, Viterbori, Traverso, & Usai	BD, DN, RC, TO	0, 0, 0, 0
274	2014	Lengua, Kiff, Moran, Zalewski, Thompson, Cortes, & Ruberry	BD, DN	0, 0
275	2014	Lerner & Lonigan	BD, DN, HG, KT	0, 0, 0, 0
276	2014	Marshall & Drew	BD, SI	0, 0
277	2014	Mayer, Abelson, & Lopez-Duran	GR, IC, SH, SI, TO	1, 1, 1, 1, 1
278	2014	McClelland, Cameron, Duncan, Bowles, Acock, Miao, & Pratt	DN, SI	0, 0
279	2014	Pauli-Pott, Roller, Heinzel-Gutenbrunner, Mingebach, Dalir, & Becker	BD, DN, TO	1, 1, 1
280	2014	Skowron, Cipriano-Essel, Gatzke-Kopp, Teti, & Ammerman	DN, SH	0, 0
281	2014	Utendale, Nuselovici, Saint-Pierre, Hubert, Chochol, & Hastings	DN	0
282	2014	van Tilborg, Segers, van Balkom, & Verhoeven	GS	0
283	2014	Verkerk	TO	0
284	2015	Cheie, Veraksa, Zinchenko, Gorovaya, & Visu-Petra	KT	0
285	2015	Di Norcia, Pecora, Bombi, Baumgartner, & Laghi	RC	0
286	2015	Lengua, Moran, Zalewski, Ruberry, Kiff, & Thompson	BD, DN	1, 1
287	2015	Petersen, Bates, & Staples	BD, GS, SH	0, 0, 0
288	2015	Wiebe, Clark, De Jong, Chevalier, Espy, & Wakschlag	SH	0

BD = Bear/Dragon, BS = Baby Stroop, DN = Day/Night, GR = Green-Red Signs, GS = Grass/Snow, HG = Hand Game, IC = Inhibitory Control subscale of CBQ, KT = Knock-Tap, LM = Less is More, RC = Reverse Categorization, SH = Shape Stroop, SI = Simon Says, SC = Spatial Conflict, TO = Tower.

Note: "Inhibitory Control Measures" column lists the inhibitory control measures from Table 1 that were examined by each study. "Exclusion Criterion" denotes the reason for exclusion (the exclusion criterion number), if the study was excluded from the meta-analysis for each of the respective inhibitory control measures (0 if the study was included). Study exclusion criteria were: 1 = having identical data to that of another included study, 2 = not reporting the information necessary to determine POM (after contacting authors), 3 = including only a clinical sample, and 4 = providing a single mean score for a sample of children whose ages spanned more than 3 years.

systematic missingness bias, there was one exception to exclusion criterion 2: if a study reported that the task showed a floor or ceiling effect but did not report the actual POM, we imputed a POM of 10% or 90%, respectively.

For one task, Tower, widely discrepant scoring procedures were used across studies (e.g., some used bonus points for self-correction or careful placement of blocks and negative points for knocking down the tower). To make the POM estimates for Tower conceptually comparable across studies, we excluded studies (exclusion criterion 2) that altered the basic scoring metric (number of trials the child yields to the experimenter) by adding or subtracting points for other behaviors. We describe these limitations of the Tower task in greater detail later.

Table 3
Studies examining each inhibitory control measure (or similar variants).

Measure	No. of included studies	No. of excluded studies	Numeric index of included studies (from Table 2) examining measure	Excluded studies
Baby Stroop	4	7	95, 142, 143, 179	59, 75, 113, 129, 183, 199, 229
Bear/Dragon	50	28	4, 10, 17, 20, 22, 28, 34, 39, 44, 48, 49, 57, 70, 71, 80, 86, 92, 97, 107, 109, 126, 127, 132, 137, 162, 164, 167, 172, 173, 181, 193, 194, 205, 207, 225, 228, 230, 237, 241, 250, 252, 261, 265, 266, 270, 273, 274, 275, 276, 287	25, 32, 40, 45, 50, 58, 66, 81, 96, 108, 117, 176, 183, 187, 191, 199, 208, 216, 242, 248, 249, 253, 255, 256, 259, 267, 279, 286
Grass/Snow	31	13	20, 48, 54, 57, 97, 108, 121, 132, 135, 137, 139, 155, 164, 167, 172, 186, 196, 201, 203, 213, 217, 219, 225, 228, 232, 237, 240, 244, 270, 282, 287	44, 58, 66, 81, 96, 117, 134, 141, 212, 218, 225, 247, 259
Green-Red Signs	2	10	12, 240	25, 32, 40, 96, 117, 134, 155, 195, 233, 277
Day/Night	85	31	6, 8, 20, 29, 36, 38, 42, 44, 52, 54, 55, 56, 57, 65, 67, 68, 70, 71, 76, 82, 83, 85, 86, 91, 93, 95, 97, 98, 101, 103, 105, 107, 115, 118, 120, 121, 125, 127, 130, 131, 133, 136, 138, 139, 152, 153, 173, 174, 179, 180, 182, 183, 184, 186, 194, 196, 199, 201, 204, 207, 213, 217, 219, 221, 225, 230, 238, 240, 241, 244, 250, 257, 260, 261, 262, 265, 266, 269, 270, 273, 274, 275, 278, 280, 281	21, 27, 35, 53, 60, 62, 64, 66, 69, 77, 81, 84, 89, 106, 113, 123, 165, 191, 215, 224, 235, 242, 247, 248, 249, 253, 256, 259, 261, 279, 286
Hand Game	26	13	9, 13, 14, 15, 48, 51, 57, 70, 71, 92, 95, 100, 103, 115, 121, 135, 140, 144, 186, 207, 216, 244, 257, 260, 269, 275	30, 37, 47, 66, 78, 113, 153, 159, 208, 212, 218, 247, 251
Inhibitory Control (CBQ)	28	21	5, 10, 12, 19, 23, 25, 26, 48, 52, 55, 63, 73, 79, 107, 109, 119, 122, 147, 148, 164, 174, 177, 196, 203, 209, 231, 257, 264	7, 18, 20, 27, 31, 39, 41, 46, 61, 88, 106, 111, 151, 170, 185, 191, 226, 233, 237, 263, 277
Knock-Tap	16	11	24, 33, 74, 99, 104, 109, 156, 157, 164, 184, 193, 200, 237, 255, 275, 284	11, 47, 53, 60, 69, 72, 89, 124, 160, 215, 259
Less is More	11	2	57, 58, 132, 135, 137, 141, 154, 161, 166, 189, 257	212, 226
Reverse Categorization	13	4	48, 57, 87, 116, 167, 188, 205, 216, 227, 239, 258, 273, 285	128, 168, 198, 208
Shape Stroop	22	11	12, 19, 48, 57, 87, 142, 143, 155, 177, 188, 197, 199, 208, 216, 232, 236, 239, 241, 245, 280, 287, 288	32, 40, 96, 128, 183, 195, 229, 233, 235, 259, 277
Simon Says	15	20	1, 2, 3, 12, 57, 90, 109, 111, 145, 155, 158, 214, 262, 276, 278	25, 31, 32, 41, 96, 134, 146, 150, 153, 171, 178, 190, 195, 210, 218, 233, 268, 271, 272, 277
Spatial Conflict	9	3	16, 18, 20, 43, 57, 94, 112, 169, 221	165, 192, 224
Tower	15	27	20, 22, 32, 44, 57, 63, 102, 110, 173, 177, 223, 246, 250, 273, 283	10, 18, 19, 25, 40, 48, 58, 81, 96, 114, 117, 121, 149, 160, 170, 185, 206, 211, 220, 233, 234, 240, 243, 254, 263, 277, 279

Note: See Table 2 for reasons why, for each measure, studies were excluded from the meta-analysis.

To include as many studies as possible, we contacted authors of studies that did not report POM requesting the mean and possible minimum and maximum. We identified 288 studies that included on average 1.83 inhibitory control measures that we considered, yielding 528 study-by-measure combinations that met our inclusion criteria.[3] Ninety (31%) studies did not report POM estimates or the descriptive statistics necessary – mean and possible minimum and maximum – to calculate POM estimates for one or more inhibitory control measures (158, 30% study-by-measure combinations). We contacted authors of these studies to obtain the necessary descriptive statistics to calculate POM. After authors' responses to our requests, 61 (12%) study-by-measure combinations were excluded for identical data, 107 (20%) were excluded for lacking the necessary descriptive statistics to determine POM, 6 (1%) were excluded for including only a clinical sample, and 27 (5%) were excluded for providing a single average for a sample of children whose ages spanned more than 3 years. The meta-analysis included the remaining 198 (68%) studies and 327 (62%) study-by-measure combinations.

Because some studies provided different POM estimates for multiple subgroups (e.g., 3- versus 5-year-olds), we calculated average POM at each age for each subgroup separately. We calculated an average of the subgroups' POM at each age, weighted by each subgroup's sample size, with higher POM scores reflecting more optimal performance. We used the age ranges of each subgroup for determining the ages to which each subgroup's POM score applied. If subgroup age ranges were not reported, age ranges were estimated from the interval including ±1 standard deviation of the subgroup's mean age. Our estimates of the useful age ranges were intentionally conservative. We assume a task is useful at a particular age if the POM at that age is between 20% and 80% (to prevent floor and ceiling effects). The useful age range was selected as the age range during which the POM was greater than 20% and less than 80%, after which 20% POM was last surpassed and before which 80% was first reached. The inhibitory control measures and their useful age ranges are in Table 1. Based on the measures' useful age ranges, we grouped the measures into early, middle, and late measures. The average POM at each age is depicted in Figs. 1, 2, and 3 for the early, middle, and late measures, respectively.

Several patterns are notable. First, in general and as expected, performance on inhibitory control measures improved with age. Second, performance on some tasks reached plateaus at later ages (beginning in some cases around 4 years of age). Researchers should therefore be thoughtful about the inhibitory control tasks they use, especially at later ages (4–8 years), and might consider ways to avoid ceiling effects to make the tasks usable for a longer developmental span (discussed later). Third, unexpectedly, the Shape Stroop, Bear/Dragon, and Tower tasks showed a pronounced drop in performance from about 25 to 30 months of age. We are hesitant to infer that inhibitory control abilities declined with development over this span. Speculatively, the drop in performance could owe to differences in systematic missingness rates. Assuming children's performance remains relatively stable over this span, decreases in systematic missingness with age (because of lower noncompliance, better verbal comprehension, etc.) could result in more of the least-skilled children completing the task at later ages, resulting in lower average inhibitory scores.

Fig. 4 depicts the useful age range for each measure. Examination of Fig. 4 suggests that the measures have fairly limited age ranges of utility. Whereas the Inhibitory Control questionnaire subscale was useful for 6 years, behavioral tasks were useful for 2.49 years, on average. However, this does not necessarily mean that measures may not prove useful outside the age ranges suggested – we emphasize that the suggested age ranges are based on the POM and age ranges reported in prior studies with normative populations. Some of the measures' POM did not exceed our 20% threshold for floor effects and our 80% threshold for ceiling effects (e.g., Less is More). This indicates that relatively few studies have examined the extremes of low/high performance, and that collectively, studies have not examined the full range of the relevant ages for many inhibitory control measures. As a result, the actual useful age ranges may extend to younger or older ages than so far have been examined in prior studies. Another limitation is that POM may not fully account for whether a task is useful at a given age,

[3] Our literature search was not exhaustive for the Bear/Dragon and Day/Night tasks. We wanted to have sufficient coverage across the entire age range. We did not do an exhaustive search in the middle and high ends of the age range because we had ample studies (>15) to stabilize POM estimates. We attempted to do an exhaustive search at the low end of the age range in order to estimate the youngest age at which these tasks are useful.

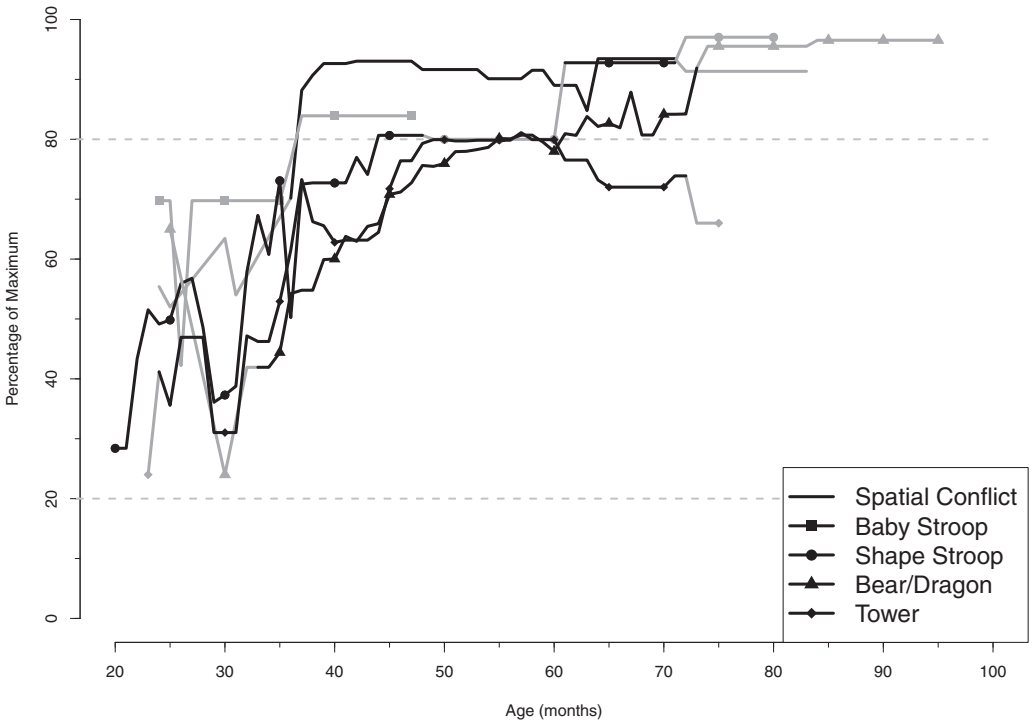

Fig. 1. Children's average percentage-of-maximum score on "early" inhibitory control measures by age. Gray lines represent age ranges with weaker empirical support (i.e., fewer than two studies; line is interpolated if there are no studies at a given age). Dashed gray horizontal lines at 20% and 80% reflect our low and high cutoffs, respectively, for defining the useful age range of each task.

especially at younger ages, because POM does not capture the amount of systematic missingness. Some children may not complete a task at younger ages because they lack a skill required by the task (or for another reason that might correlate with inhibitory control ability, such as noncompliance or poor receptive language skills), resulting in systematic missingness that would artificially inflate the mean score and POM. In these cases, however, the useful age ranges would be even more abbreviated than our estimates. Thus, the suggested age ranges are provisional. Future studies will refine the useful age ranges for inhibitory control and other self-regulation measures. We also recognize that the age ranges we are reporting are based on analysis of the measures' sensitivity to detect individual differences in behavioral accuracy. There are ways to make measures relevant to a wider range of ages, such as by measuring reaction times, increasing rule difficulty, adding rule switches, and decreasing the amount of time a child has to respond.

Comparison of task demands of inhibitory control measures

As a further differentiation of the inhibitory control tasks, we grouped the tasks into four a priori clusters based on the task's demand: (a) perceptual inhibition, (b) performance inhibition, (c) association inhibition, and (d) motivational inhibition. Perceptual inhibition included tasks that involve the inhibition of perceptual information, such as Baby Stroop, Shape Stroop, and Spatial Conflict that require inhibition involving one of the following dimensions: size (Baby Stroop, Shape Stroop) and location (Spatial Conflict). Performance inhibition included tasks that use a cue (person or puppet) to indicate whether or not to inhibit a prepotent behavioral response, such as the activation/go trials versus inhibition/no-go trials of Bear/Dragon and Simon Says. Association inhibition included tasks

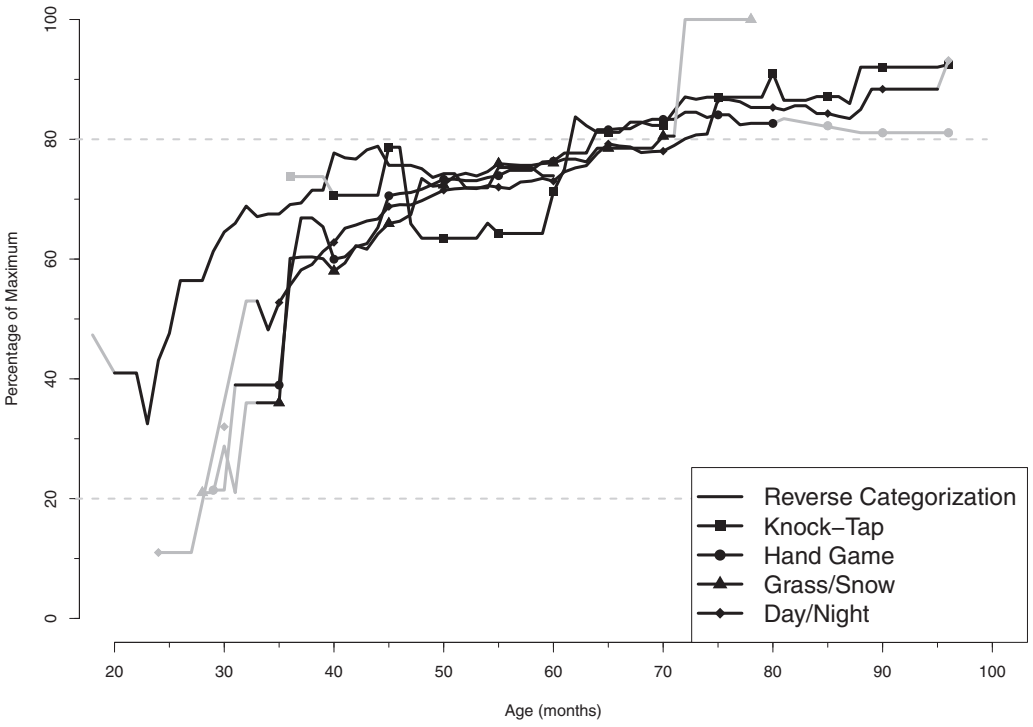

Fig. 2. Children's average percentage-of-maximum score on "middle" inhibitory control measures by age. Gray lines represent age ranges with weaker empirical support (i.e., fewer than two studies; line is interpolated if there are no studies at a given age). Dashed gray horizontal lines at 20% and 80% reflect our low and high cutoffs, respectively, for defining the useful age range of each task.

similar to performance inhibition tasks but required inhibition of a prepotent response and the generation of a competing response on all trials, such as Day/Night, Grass/Snow, Green-Red Signs, Hand Game, Knock-Tap, and Reverse Categorization. Motivational (or "hot") inhibitory control tasks included tasks with a stronger motivational component, such as Less is More (correct choices yield more winnings) and Tower (children tend to prefer to take more turns). Other researchers have distinguished perceptual inhibition (Martin-Rhee & Bialystok, 2008) and "cold" versus "hot" inhibition (Zelazo & Carlson, 2012). However, we know of no groups that have distinguished tasks conceptually based on whether inhibition is required on a subset of trials and the generation of a competing response on the other subset of trials (i.e., performance inhibition) or whether inhibition and the generation of a competing response are required on all trials (i.e., association inhibition). We see the distinction between performance and association inhibition as speculative but possibly useful. We recognize that all measures tap multiple aspects of inhibitory control (e.g., association inhibition tasks also require perceptual inhibition), so our categorization, though a useful early step, should not imply mutually exclusive clusters. Nevertheless, although grouping the tasks into clusters of common task demands is an oversimplification, doing so allows us to compare the developmental rates of different aspects of inhibitory control.

The task demands within and across categories of inhibitory control appear to differ in their useful age ranges and rates of growth. These can inform our understanding of the construct of inhibitory control and the task demands that elicit greater levels of inhibition. The perceptual inhibition tasks generally showed the youngest useful age range, suggesting that perceptual inhibition develops earlier than other aspects of inhibitory control. Among the perceptual inhibition tasks, tasks involving multiple sets of stimuli (e.g., fruit, banana, and apple in Shape Stroop) may be more difficult than tasks

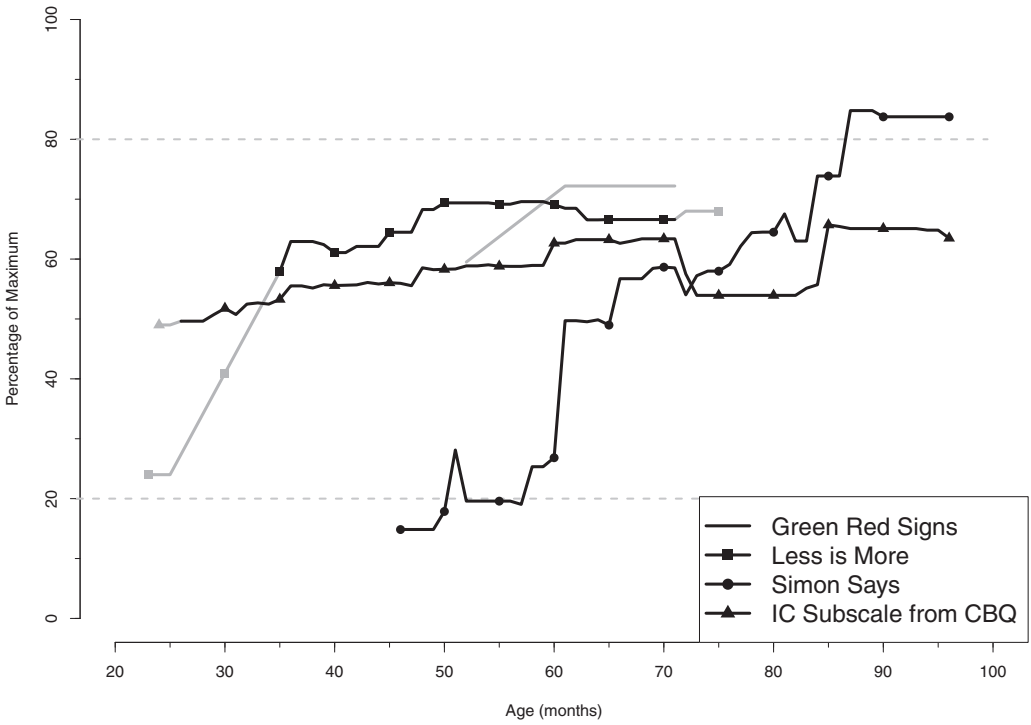

Fig. 3. Children's average percentage-of-maximum score on "late" inhibitory control measures by age. Gray lines represent age ranges with weaker empirical support (i.e., fewer than two studies; line is interpolated if there are no studies at a given age). Dashed gray horizontal lines at 20% and 80% reflect our low and high cutoffs, respectively, for defining the useful age range of each task.

involving one set of stimuli (e.g., cup/spoon in Baby Stroop). Among the performance inhibition tasks, Simon Says was more challenging, i.e., was useful at later ages than Bear/Dragon. This may be because Simon Says typically involves a single cue source (the same person provides cues for both inhibition and activation trials), whereas Bear/Dragon involves two different cue sources (bear for activation trials and dragon for inhibition trials), consistent with findings of Marshall and Drew (2014) who experimentally tested variations of Simon Says with one versus two cue sources. In addition, the nature of the cue source may make Simon Says more difficult because inhibiting a response to an adult may be more difficult than inhibiting a response to an inanimate object, possibly because children are more eager to comply with adult requests.

Among association inhibition tasks, Day/Night and Grass/Snow are conceptually similar and require the inhibition of a pre-established prepotent response; in each, the two stimuli have a widely known association (e.g., day is the opposite of night). Providing construct validation of their conceptual similarity, Day/Night and Grass/Snow had similar useful age ranges. They were somewhat more challenging than other association inhibition tasks including Hand Game and Knock-Tap that used an arbitrary association (e.g., tap when the experimenter knocks) rather than a pre-established association (say "day" when the experimenter points to a picture of a moon). The motivational inhibitory control task, Less is More, showed a later useful age range than other task groupings, suggesting that motivational inhibitory control tasks may tap later-developing affective processes. This was not the case with the Tower task, however, which may reflect methodological limitations of the task (discussed later). In sum, more affective/motivational components, multiple sets of stimuli, a single cue source of information reversal, an adult cue source, and inhibiting a pre-established prepotent response appear to require greater levels of inhibitory control. Further, some aspects of inhibitory control such as

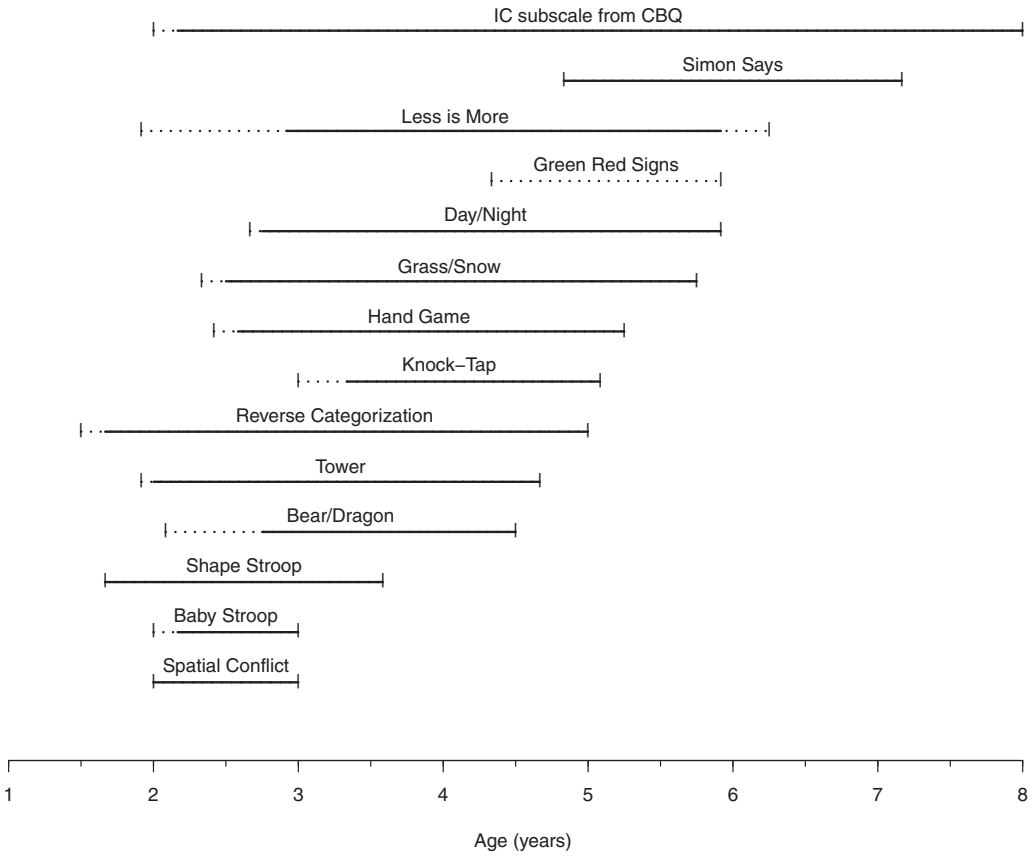

Fig. 4. Depiction of useful age ranges of inhibitory control measures based on age ranges from Table 1 and Fig. 1. Dotted lines represent age ranges with weaker empirical support (fewer than two studies). CBQ = Children's Behavior Questionnaire.

perceptual inhibition may develop earlier than other aspects of inhibitory control such as performance and association inhibition, which may develop earlier than motivational inhibition.

The questionnaire measure of inhibitory control, from the CBQ, showed comparatively little sensitivity to mean-level developmental change (see Fig. 3). The Inhibitory Control subscale of the CBQ would be useful when one's goal is to examine temperamental individual differences in inhibitory control rather than group-level developmental change in inhibitory control over time. Although studies used the Inhibitory Control subscale of the CBQ across 2–8 years of age (and it was useful across this span according to our definition), it was only designed for children aged 3–7 years. Thus, consistent with the heterotypic continuity of inhibitory control, we suggest using the CBQ only for 3- to 7-year-old children for whom it was designed.

Methodological limitations of prior studies

We also noted common limitations across a number of studies. First, many studies included a wide range of ages in the same subgroup. A wide age range in a subgroup might render study inferences less valid because most of the tasks appear to have a fairly limited age range of utility, and as we describe later, the same task may not have construct validity at different ages. Second, many study-by-measure combinations ($N = 160$, 31%) did not report the basic descriptive statistics necessary in the manuscript (mean and possible minimum and maximum) to determine POM for the inhibitory control

measure. Many studies reported descriptive statistics only on the composite measures, rather than on the individual tasks. We suggest that future studies report unstandardized descriptive statistics on behavioral performance (mean, standard deviation, and possible and observed minimum and maximum) even when using a composite of multiple measures (e.g., as was done by Cuevas & Bell, 2014). These basic descriptives are necessary for understanding the metric of task scores and for understanding the sample's performance on this metric. Third, many studies dichotomized task performance into pass/fail. While sometimes appreciating the simplicity of pass/fail, we advise reporting on a continuous metric as possible to retain data precision and have higher statistical power to detect developmental change and associations with other variables.

Fourth, we noted obscure and inconsistent scoring procedures across studies of the same task. For instance, in the Tower task, children are asked to take turns placing blocks to build a tower with the experimenter. The experimenter waits to take a turn until the child signals that he or she is giving the experimenter a turn, with the dependent variable often being the number of turns the child yields to the experimenter. The assumption is that the "optimal score" would be to yield half of the turns to the experimenter. Thus, for 20 turns, an optimal score would be the child yielding 10 turns to the experimenter. This optimal score (10) is not the maximum possible score (20), however, making the interpretation of the metric difficult. Moreover, some studies further complicate this metric by adding bonus points for self-correction and careful placement, and subtracting points for knocking over the tower. Combining all of these measures puts the final score on a metric with no clear optimal/maximum score. While sometimes appreciating creative scoring procedures, for the sake of allowing better cross-study comparisons, we advise reporting clear scoring procedures with a meaningful range of possible values.

Implication of heterotypic continuity for measurement

In addition to empirical considerations for the measurement of inhibitory control, there are theoretical and psychometric considerations as well. Theoretical discussions of the development of self-regulation often invoke the concept of heterotypic continuity (Kopp, 1982; Raffaelli, Crockett, & Shen, 2005). Heterotypic continuity is defined as the "continuity of an inferred genotypic attribute presumed to underlie diverse phenotypic behaviors" (Caspi, 2006, pp. 349–50) or as "the manifestation of the same underlying process through different behavioral presentations at different developmental periods" (Cicchetti & Rogosch, 2002, p. 13). In other words, inhibitory control and self-regulated behavior more generally, although continuous in its underlying purpose or function, may appear different in toddlers than it does in older children and adolescents. Self-regulation and inhibitory control change over development from early reliance on external sources for control to ability for internal, self-initiated forms of control (Berger, 2011; Calkins & Howse, 2004; Kopp, 1982). Acceptance of the notion of heterotypic continuity has two main implications for measurement.

The first implication is that the same measure may or may not reflect the same construct at different ages (Widaman, Ferrer, & Conger, 2010). Across broad ranges of development, a given measure may not be *developmentally appropriate* or construct-relevant at all ages. Similarly, even if a measure taps the same construct at different ages, it may not be sensitive enough to detect variability at each age (Carlson, 2005) – it may not be *developmentally sensitive* across different ages. If the underlying construct changes in its manifestation depending on the stage of development, scores on the same measure across time may reflect differences in the meaning of the measure, rather than real change in the construct.

For example, our meta-analysis indicated that the Bear/Dragon task is most useful (with stronger confidence) from 33 to 54 months of age, and may therefore have a different meaning outside that developmental window. Anecdotally, in testing children longitudinally on a variant of the Bear/Dragon task (Bird/Alligator, methods described in detail elsewhere; Petersen, Bates, & Staples, 2015), we have observed that children at 30 months of age who inhibited responses on the no-go trials also tended to inhibit on the go trials (i.e., performance on go and no-go trials was inversely correlated: $r[137] = -.67$). However, this was not as strongly the case at 36 ($r[133] = -.22$) and 42 ($r[127] = .12$) months (Fisher r-to-z values significant at $p < .001$ level). Thus, there appears to be a developmental change in the meaning of the Bear/Dragon task, where performance on the no-go trials has a

different meaning at 30 months than at later ages. If this is the case, then scores on the task at 30 months would not be conceptually comparable with scores at 42 months. This developmental shift may be consistent with our previous findings that the no-go score on the Bird/Alligator task was associated with another inhibitory control measure, Shape Stroop, at 42 months but not at 30 months (Petersen et al., 2015). We plan to examine whether no-go inhibition on the Bird/Alligator task reflects more motivational or affective inhibition at 30 months and more executive inhibition at later ages. Some researchers have tried to ensure that higher inhibition scores reflect better performance on both go and no-go trials by multiplying the go and no-go scores (Eisenberg et al., 2013). However, if the meaning of the no-go trials changes with development, the composite would change in meaning, too. In sum, to make inferences regarding developmental change, we must use equivalently-functioning and construct-valid measures across time.

The second implication is that describing the heterotypic continuity of inhibitory control may require using different measures across time (Widaman et al., 2010) because children are expected to demonstrate different behaviors at different ages for the same underlying construct of self-regulation (Bates & Novosad, 2005). When theoretically sound, it is preferable to maintain the same measures over time to ensure comparability (Willett, Singer, & Martin, 1998). However, there are also occasions when studies of development benefit from different measures over time (McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009). Ideally, longitudinal assessment of inhibitory control should accommodate changes in the manifestation of inhibitory control by using different measures (when necessary) of the same construct over time, with careful consideration of developmental appropriateness, developmental sensitivity, and construct validity across the time frame of study. In the following, we consider these methodological aims in detail.

Developmental appropriateness and sensitivity

By *developmentally appropriate*, we refer to measures that are consistent with the capacity of most children of a particular developmental level and with our understanding of the construct at that point in development. A measure could be considered developmentally appropriate if its scores cluster around a value close to the center of the plausible range. If a task is not developmentally appropriate for a given age, a score on the task at that age is meaningless for describing development in terms of the construct of interest. By *developmentally sensitive*, we refer to measures that yield enough variability to be useful for assessing individual differences within age and to compare scores across ages. A measure could be considered developmentally sensitive if there are large inter-individual differences around the mean (and not ceiling or floor effects). Developmental appropriateness and sensitivity help ensure proper measurement of each child's inhibitory control and adequate spread between individuals. It is common for behavioral tasks to be developmentally appropriate and sensitive at one age, but too easy for participants at later ages, resulting in ceiling effects. It is also likely that a given task would be too hard for younger children (although this is not represented in publications, as seen in Figs. 1–3). Ceiling and floor effects prevent the researcher from detecting meaningful variability in the construct and result in restricted range, which increases Type II error by reducing power to detect change and associations with other variables. One way to prevent ceiling and floor effects may be to use a more precise metric for the dependent variable (e.g., reaction time in milliseconds): a continuous interval or ratio metric as opposed to an ordinal or nominal metric. Even if no differences may be observed in behavioral accuracy because of ceiling effects, investigation of the same task with a dependent variable of a higher precision metric may detect developmental differences that make the task useful across a wider age range.

Developmental appropriateness and sensitivity depend, however, on the population of interest. For instance, the assessment of clinically significant deficits in inhibitory control in a high-risk sample should discriminate at lower ability levels (i.e., lower difficulty in item response theory; Krueger et al., 2004) rather than only at normative levels. Thus, clinical assessment may involve tasks that are mastered by most typically-developing children of that age and that would yield ceiling effects if used in a normative sample (Carlson, 2005).

Behavioral tasks tend to be developmentally appropriate and sensitive for restricted ages of children, as demonstrated earlier (see Fig. 4), so most inhibitory control tasks are not valid or useful across

a wide age range. In order to cover a wider age range, researchers can use different behavioral tasks at different ages to ensure construct validity, developmental appropriateness, and developmental sensitivity for each age studied. Researchers have noted the importance of investigating heterotypic continuity with changes in measures (Schulenberg & Maslowsky, 2009; Widaman et al., 2010).

Changing measures across time

Using different tasks poses a problem from a longitudinal perspective. One goal in longitudinal studies is to describe continuity and change. Developmental inferences are strengthened by establishing measurement invariance[4] (equivalent measures), ensuring that differences over time reflect change in the phenomenon of interest, not simply changes in measurement. As explained, inhibitory control changes in its manifestation with development and most behavioral tasks have restricted age windows of usefulness, so longitudinal assessment often requires different behavioral tasks for different ages. The use of different measures at different ages violates measurement invariance in the strict sense (same measures, same meaning) and calls into question the comparability of scores across time. Yet in order to maintain construct validity invariance, i.e., heterotypic continuity, theory requires that we accommodate measures to account for changes in the manifestation of a construct.

Potential for using the same measure across time

In light of the restricted time frame of usefulness in behavioral tasks, researchers may be inclined to implement other assessment techniques that would allow comparability across time. There are ways to make behavioral tasks relevant to a wider range of ages, such as by increasing rule difficulty, adding rule switches, and decreasing the amount of time a child has to respond (for a discussion, see Carlson, Faja, & Beck, 2015). However, as discussed earlier, the same task could lack construct validity invariance if it does not reflect the heterotypic continuity of the construct. Another possibility is the use of informant ratings, commonly completed by a parent, teacher, or observer. However, some of the same problems with longitudinal behavioral assessment are present with longitudinal informant ratings. Given that inhibitory control changes in its manifestation with development, in order for a questionnaire to have construct validity, its items – each of which reflects explicitly or implicitly a task or social context – would have to change with age (see, e.g., the widely used Rothbart temperament questionnaires and the Achenbach System of Empirically Based Assessment; Achenbach, 2009; Rothbart, Ahadi, Hershey, & Fisher, 2001). For more information on how to equate questionnaire measures with different item sets, the reader is referred to examples using integrative data analysis (Curran & Hussong, 2009; Curran et al., 2008), proportional scoring metrics (Petersen, Bates, Dodge, Lansford, & Pettit, 2015), and other scaling techniques (Kolen & Brennan, 2014). In practice, however, many investigators use questionnaires across multiple ages, including ages for which the questionnaire was not initially designed. Although the individuals' scores may be on the same *mathematical* metric from year to year, for example, if they were *z*-scores, the metrics may not be comparable across ages because the questionnaire may not assess the same construct at each age due to changes in the manifestation of the construct (Widaman et al., 2010). The scores may not be on the same *conceptual* metric.

Another approach for longitudinal measures may be the use of measures with a higher precision metric, including reaction time and psychophysiological techniques, which purportedly place all individuals on the same metric for equivalent comparison within and across ages. One example is the use of event-related potentials (ERPs) as measured by electroencephalography. Many ERP components are present across wide age ranges, suggesting that amplitudes and latencies of ERP components

[4] The term "measurement invariance" has typically been used to establish the measurement equivalence of the same manifest variables across different groups (e.g., males and females). In this paper, we discuss a method whereby measurement equivalence can be assessed when *different* manifest variables are used in a longitudinal study, which we refer to as "longitudinal factorial invariance." We refer to "longitudinal factorial invariance" when discussing tests of measurement invariance of changing measures across time. We refer to "measurement invariance" when discussing types of measurement invariance to be consistent with prior work on this topic (e.g., Meredith, 1993).

can be compared across ages. However, there are important developmental changes in ERP components (Hoyniak, 2016), related, in part, to changes in brain size and skull thickness (Rueda, Posner, & Rothbart, 2005). Thus, there is question whether developmental changes in ERP components reflect actual changes in cognitive processes over development, and whether ERPs can truly be compared across ages. In addition to physical changes that pose challenges for longitudinal ERP comparisons, the same task may have different task demands (and elicit different cognitive processes) at different ages. Researchers have also questioned the utility of comparing neural activation across time as measured by functional magnetic resonance imaging, because age differences could reflect technical aspects of data acquisition rather than developmental change (Gaillard, Grandin, & Xu, 2001).

In the case of reaction time, there is also reason to be cautious. Anderson, Nettelbeck, and Barlow (1997) found that faster reaction times among 11- than 7-year-old children in a simple stimulus pairing task were related to differences in the speed of response selection rather than speed of information processing, suggesting that differences across time in reaction time may not be meaningful in terms of processing speed. Thus, measures with a higher precision metric such as reaction time and psychophysiological measures may not be on the same conceptual metric across time, and therefore may not be fully comparable across development. Although other measures and methods may be useful, they should supplement but not replace behavioral measures in the assessment of inhibitory control (see, e.g., the importance of behavioral measures when interpreting ERPs; Fox, Hane, & Pérez-Edgar, 2006).

Measuring the development of inhibitory control in a piecewise fashion

Given the complexities of measuring the development of inhibitory control over a long time frame, why not just measure the development of inhibitory control in a shorter, piecewise manner? First, most of the widely used measures of inhibitory control are developmentally appropriate or sensitive for less than 3 years (see Fig. 4), so relying on a fixed set of tasks would severely restrict our ability to see growth over a longer span. Second, heterotypic continuity is a developmental complexity that arises in many different domains. Seeking to understand and predict changes across important developmental periods and transitions is better than ignoring the phenotypic complexities associated with meaningful developmental change. Third, the ultimate goal of developmental psychology is to understand the whole trajectory of an individual's life, and not just transitory outcomes at a particular point in life. Thus, research should strive to build a bridge that spans from childhood to adulthood (Rutter & Sroufe, 2000). The following recommendations seek to improve our ability to build a stable bridge spanning development.

One promising approach to changing measures over time is an overlapping measurement approach, similar to the approaches implemented by Patterson (1993) and Pettit, Keiley, Laird, Bates, and Dodge (2007). A staggered, overlapping measures approach is akin to accelerated longitudinal designs (also known as cohort-sequential or mixed longitudinal designs; Willett et al., 1998) except that the *measures* stagger and overlap, rather than the participant cohorts. The following section describes this approach in further detail.

Recommendations for longitudinal measurement designs

We have conducted longitudinal research, and we are aware of the limitations. So we are seeking to improve our work, and have increasingly focused on seven ways to improve longitudinal research on children's inhibitory control and its implications. Listed in the order that one might approach them in designing research, they include (1) choose the construct of interest; (2) define the age range of interest and frequency of assessment; (3) pick appropriate measures for each age of assessment; (4) use multiple measures at each assessment; (5) use overlapping measures across time points; (6) use structural equation modeling (SEM) with measures as indicators of a latent construct at each age – and in doing so, consider (a) sample size, (b) metric of variables, and (c) scale setting; and (7) test longitudinal factorial invariance. Although some of the suggestions may not be relevant or necessary depending on the research question, they would serve to increase confidence that one is measuring the same construct in a comparable way over time, and therefore, confidence in the resulting

inferences about development. Although the recommendations are provisional, they should be useful especially for developmental studies involving heterotypic continuity and changing measurement over time. The seven steps are described in further detail below.

(1) Choose the construct of interest. To decide what construct (e.g., inhibitory control) to investigate, it is important for the construct to be as unidimensional as possible because multi-dimensional constructs can make it unclear which dimension is driving an effect (Strauss & Smith, 2009). At the same time, the construct should maintain the theoretical essence and expanse of the construct, which means it cannot be *too* narrowly described. For instance, as described earlier, the construct of inhibitory control possibly includes facets such as perceptual, performance, association, and motivational inhibition. Although we could drill down further (e.g., facets of perceptual inhibition), our reading suggests there is theoretical support for examining the higher-order inhibitory control construct (Miyake et al., 2000).

(2) Define the age range and frequency of assessment. The second step is to decide what age range the longitudinal study will encompass, and at what frequency to conduct assessments. The age range and frequency of assessment are critical considerations because they set the boundaries within which one can observe developmental change. In order for development to be observed, it must encompass a construct-appropriate age range, and must be sampled at a high-enough frequency to detect change. The frequency of assessment should be as frequent as, if not more frequent than, the highest frequency of change you seek to observe (Little, Card, Preacher, & McConnell, 2009). In other words, if you theoretically expect and seek to detect observable behavioral change every six months, the frequency of assessment should be at least twice per year, if not more often. Of course, limitations of time, resources, and participant burden place restrictions on the frequency of assessment and number of measurement occasions, but the age range and frequency of assessment set the boundaries on the ability to make inferences regarding developmental change. Repeated measures may involve carryover effects, especially practice effects in the case of behavioral tasks (Hertzog & Nesselroade, 2003). Nevertheless, non-comparable measures are usually a greater threat to the validity of developmental inferences in a longitudinal study than are carryover effects (Willett et al., 1998).

(3) Pick appropriate measures for each age of assessment. The researcher should identify various measures with developmentally-appropriate and -sensitive age ranges that, as a collection, assess the whole age range of interest with the desired frequency of assessment. Moreover, each measure should have construct validity for the same construct (e.g., inhibitory control) at each age the measure is used. In addition, each measure chosen should be as unidimensional as possible. Having unidimensional tasks can be difficult because behavioral tasks typically tap into multiple complex cognitive processes (e.g., Busemeyer & Stout, 2002). For instance, many inhibitory control tasks measure multiple processes, including working memory (Wolfe & Bell, 2007b). One way to promote the unidimensionality of a construct as measured by tasks with multiple task demands may be in the SEM framework, as discussed later. In summary, it is essential to use measures that are relevant to the same construct and that are appropriate and sensitive for the different developmental periods in which they are used.

(4) Use multiple measures at each age. Having multiple measures (ideally three or more) at each age is ideal for establishing measurement invariance in SEM with different measures across time and for distinguishing true construct variance from measurement error. At least three indicators (measures) reflecting each latent construct at each age are typically necessary for a just-identified (as opposed to an under-identified) model, which gives a greater likelihood of having an identified and usable model (Little, Cunningham, Shahar, & Widaman, 2002). Collectively, the measures should have content validity and should span the relevant content of the construct at a given point in development, consistent with its heterotypic continuity. In the case of inhibitory control, based on findings from our meta-analysis, greater emphasis might be placed on perceptual inhibition earlier in development and on motivational inhibition later in development.

(5) Use overlapping measures across time points. In general, fewer changing measures and greater overlap of measures provide the SEM model with a greater likelihood of being identified (McArdle et al., 2009). Overlapping measures provide a more coherent latent structure with measures that serve as anchors across time to link measurement occasions. The more time points that measures overlap, the more likely one will be able to successfully link measurement occasions and to establish longitudinal factorial invariance (statistical support for the claim that the measures reflect the same construct

at each assessment). This point is analogous to the recommendation of Little, Preacher, Selig, and Card (2007) to ensure that multiple different age cohorts describe a long developmental process by having overlap of at least two time points. For measuring a long span of inhibitory control development, we recommend that (a) each measure is present for at least two consecutive time points and (b) most time points have few changes in measures – one or fewer measurement substitutions at any time point, if possible (for examples using growth models with changing measurement, see the figure on p. 101 of Muthén & Muthén, 2010; fig. 1 in Pettit et al., 2007).

(6) Use SEM with measures as indicators of a latent construct at each age. SEM permits testing longitudinal factorial invariance and partials out measurement error in the estimation of latent constructs. In a reflective SEM model, the latent variable seeks to capture the construct of interest by reflecting the common variance (i.e., covariance) among several measures. The unique variance and measurement error of each of the individual measures are set aside in the form of residuals. Because of the importance of using measures reflecting unidimensional constructs (Strauss & Smith, 2009), SEM may be useful for dissociating construct variance from measurement error and from the contributions of other task demands (e.g., dissociating inhibitory control from working memory). Thus, applying SEM to behavioral tasks may decompose variance from the tasks into unique measurement variance and common latent variance that is a reflection of a construct of interest and more unidimensional. As a result, we would endorse Garon et al.'s (2008) recommendation to use SEM as a framework to create purer, more unidimensional measures of inhibitory control. Finally, because of the capacity of SEM to measure the same latent variable with different indicators across time, we would argue that latent growth curve modeling in SEM can promote the study of constructs whose manifestation changes over time (Pettit et al., 2007). The different measures must assess the same construct at each age in order for the latent factors to be compared in a meaningful way, and for inferences to be meaningful about developmental change. In other words, the measures should have construct validity invariance for the phenomenon of interest, as we describe later. Further considerations are important in using an SEM approach:

(a) Sample size. SEM with latent factors requires a sizeable participant sample, particularly for linking time points with changing measurement. According to Westland's (2010, 2012) heuristic, the minimum sample size for a measurement model in SEM depends on the ratio of indicators to latent variables. He estimates that a sample size of 400, 200, and 100 would be necessary for models with 2, 3, and 4 indicators per latent variable, respectively. However, the minimum sample size to have adequate power to detect an effect depends on the effect size. Estimates from Westland's (2010, 2012) equation for the minimum sample size, based on the effect size of the association between two latent variables, are in Table 4. The minimum sample size necessary for prediction in an SEM model is determined by taking the greater of the two values from (a) the minimum sample size necessary for a measurement model (see above based on formula in Westland, 2010) and (b) Table 4 (see the value corresponding to the number of latent variables in the model and the smallest effect size to observe). Westland notes, however, that the required sample size may exceed these estimates because the minimum sample size depends on many other factors, such as missingness and multicollinearity.

(b) Metric of variables. In addition to issues of sample size, another issue in SEM involves the metric of variables. When combining multiple behavioral tasks with a single latent variable, one issue that arises is whether to standardize the tasks to be on the same metric. In most applications, the indicators of a latent factor do not have to be on the same metric. The issue of metrics must be considered, however, when using different manifest indicators at each time point. When using different measures across time, it is preferable for scores on measures to share a common, comparable metric (e.g., have the same possible minimum and maximum) in order for indicators to be comparable within time and for latent variables to be comparable across time.

There are ways to reach the goal of having items on the same comparable mathematical metric, while still preserving a meaningful mean level (e.g., inhibitory control ability level) that is sensitive to change over time. For example, proportional scoring metrics can be used to compare scores across ages. One option is the previously mentioned (e.g., Fig. 1) percentage-of-maximum (POM) scoring method that takes an individual's score and divides it by the total possible score (Little, 2013). POM rescaling essentially converts each individual's score on a task into a proportion representing the percentage correct (or percentage of the measured phenomenon) on the task. As with any metric, the

Table 4
Minimum sample sizes for adequate statistical power to detect associations of varying effect sizes with latent variables in SEM.

Number of latent variables	Minimum sample size		
	$\|\rho\| = .1$	$\|\rho\| = .2$	$\|\rho\| = .3$
1	87	22	10
2	152	37	16
3	290	69	28
4	387	92	38
5	463	110	45
6	526	124	50
7	579	137	55
8	625	148	60
9	666	157	63
10	703	166	67

$|\rho|$ = the smallest effect (Gini correlation coefficient) to detect on latent variables.
Note: Assuming an alpha-significance criterion of $\alpha = .05$ and power = .80. See Westland (2010, 2012) for the equation for estimating minimum sample size when predicting other variables. The minimum sample size refers to the minimum sample size to detect an association between latent variables. The actual lower bound sample size necessary is the greater of the two values from (a) the minimum sample size necessary for a measurement model (see formula in Westland, 2010) and (b) Table 4 (here).

assumption is that a child with a given POM has the same ability as another child with the same POM, and this may not be the case when different tasks have different levels of difficulty. Because all proportions have the same possible range (0–1), they have greater comparability than the raw metric, and still allow the researcher to observe growth (mean-level change) over time. An advantage of POM rescaling is that, unlike standardization (e.g., age-normed z- or t-scores, which are not recommended for longitudinal data; King, 1986; Willett et al., 1998), it does not distort any of the fundamental statistics of the variable (McArdle, Hamagami, Meredith, & Bradway, 2000). In other words, POM rescaling does not shift the mean or variance relative to other tasks. On the other hand, because POM-rescaled variables maintain their fundamental statistics relative to other variables, different POM-rescaled variables may not be fully on the same mathematical metric. For instance, Grass/Snow appears to be a more challenging task than Shape Stroop (see Fig. 4). If a sample of 3-year-old children has a lower mean score (relative to its possible range) on Grass/Snow than Shape Stroop, POM scores for Grass/Snow and Shape Stroop would not be on the same metric for purposes of defining inhibitory control. Nevertheless, rescaling variables to be on the same mathematical metric (e.g., *z*-scores) does not ensure comparability either. The only way to ensure variables are comparable is for them to be on identical metrics prior to rescaling (Willett et al., 1998), which may not always be feasible with behavioral measures.

Alternatively, one could use the greatest observed score as the denominator rather than the total possible score in POM rescaling (e.g., Velásquez, 2010), which could shift the POM to have greater comparability across measures, but it may obscure mean-level change, as well. Consider a hypothetical study where children are measured with 100 Day/Night trials at 3 and 4 years of age. The mean is 30 and 50 at ages 3 and 4 years, respectively, and the observed maximum is 60 and 100, respectively (whereas the possible maximum would be 100 at both ages). Dividing by the observed maximum to calculate POM would shift the mean POM at each age to be the same (50) even though there is a meaningful mean-level difference. In this case, dividing by the possible maximum would retain this mean-level difference, resulting in a POM of 30 and 50 at ages 3 and 4, respectively. Thus, if dividing scores on a measure by the greatest observed score, it is recommended to divide by the same value over time (e.g., the greatest observed score on a measure across all time points; Little, 2013).

A limitation of POM rescaling is that it results in a different metric from the observed metric, which can make interpretation difficult (Velásquez, 2010). When combining measures with different metrics

in the context of changing measurement over time, however, it remains advantageous to have a "reasonably comparable scale" that affords greater potential for comparability across time (for examples of studies using POM rescaling in the context of growth curves, see McArdle et al., 2000, p. 60; Petersen et al., 2015). Despite its limitations, POM rescaling provides a way to make scores on separate tasks more comparable with a similar metric (proportion correct), yet still maintaining the ability to observe changes in means and variances over time. Thus, for equating changing measurement over time when measures have different metrics, we recommend the use of POM rescaling in most cases rather than standardization.

(c) Scale setting. There are several options for scale setting in SEM. Scale setting is important because it defines the metric of the latent variable and the resulting interpretability of model parameters for describing developmental change. We consider two common approaches: the marker-variable and reference-group methods (Little, Slegers, & Card, 2006). The marker-variable method is the most common scale-setting approach in SEM and is the default method in some software for SEM analyses. In the marker-variable method, the first (reference) indicator at each time of measurement is constrained to be equal with its loading fixed to 1 and its intercept fixed to 0 (Little et al., 2007). Because the metric of the latent factor is based on the metric of the arbitrarily chosen first indicator, however, its interpretation may be difficult (Little et al., 2007), particularly in the context of indicators with different scales.

An alternative approach to the marker-variable method is the reference-group method, in which the latent variable for one "reference" group (baseline group of interest) is standardized and other groups are scaled relative to the metric of the reference group. For a single group in the longitudinal context, the reference-group method uses a reference *time*: the latent variable at only T1 (or a different time of interest) is standardized by fixing its mean to 0 and its variance to 1, and all of the factor loadings and intercepts are freely estimated (Ferrer, Balluerka, & Widaman, 2008; Little et al., 2006). The advantage of the reference-group method is that changes in the means and variances over time are scaled relative to the metric of the latent factor at T1 (Ferrer et al., 2008), and are not age-normed or standardized at each age. Thus, although the metric itself may be difficult to interpret, it provides the researcher with the ability to observe growth in means or variances. For most cases dealing with changing measures over time, we recommend using the reference-group approach because it allows observable change in means and variances over time on a more comparable metric.

(7) Test longitudinal factorial invariance. Equivalence of measures across time can be tested in SEM measurement models. Tests of longitudinal factorial invariance help provide evidence for whether one is measuring the same construct on the same scale across time. Longitudinal factorial invariance is demonstrated when multiple measures reliably and validly assess a construct in the same or a similar way across time (or across groups; Curran & Hussong, 2009). Longitudinal factorial invariance tests the extent to which the structure of interrelations among a set of measures (i.e., the factor structure) is statistically the same or different across time. In order for the measures to reflect the same construct across time, it is expected that the measures share a similar factor structure across time. Thus, longitudinal factorial invariance in the situation where different but overlapping measures have been used has the potential to greatly inform our understanding of heterotypic continuity in development (Nesselroade & Estabrook, 2009). However, longitudinal factorial invariance is not sufficient by itself to claim that equivalent measures reflect the same underlying construct and, as a result, is no substitute for construct validity invariance (Knight & Zerr, 2010), as we describe later. We recommend testing longitudinal factorial invariance whether the same or different manifest variables are used, because evidence of failed longitudinal factorial invariance can be informative of developmental change in the phenomena of interest (Edwards & Wirth, 2009; Widaman et al., 2010).

Setting aside the concept of heterotypic continuity, traditionally, in order to make the case that the same construct is being measured across time, two findings are necessary: (1) The measures have construct validity for the same construct of interest at every age in which a particular measure is used (i.e., they are theoretically and empirically associated with the construct at every age they are implemented). (2) The measures show longitudinal factorial invariance (i.e., measurement invariance across time) – the relations of the measures to the construct do not differ across time, and the construct is measured on the same metric across time. Both theoretical (construct validity invariance) and

Table 5
Longitudinal factorial invariance applied to example longitudinal study.

Type of invariance (weak to strong)	Description	Constraints in Fig. 5
Baseline model	The model (see Fig. 5) to which the four models below are compared using a χ^2 difference test with as many *df* as there are additional constraints. Scale-setting for the baseline model: Standardize the latent factor at T1 ($M = 0$, $SD = 1$).	$\alpha_1 = 0$ $\sigma^2_{11} = 1$
1. Configural invariance	All of the above constraints, in addition to the same pattern of free versus fixed factor loadings over time: (1) Estimate the loading of the first indicator on the latent variable at T1, but fix the corresponding first loadings to be equal across time. (2) Estimate the first intercept, but fix the corresponding first intercepts to be equal across time. (3) Freely estimate the remaining parameters.	(1) $\lambda_{11} = \lambda_{12} = \lambda_{13}$ (2) $\tau_{11} = \tau_{12} = \tau_{13}$
2. Weak factorial invariance	All of the above constraints, in addition to: Fix corresponding factor loadings across time.	$\lambda_{21} = \lambda_{22} = \lambda_{23}$ $\lambda_{41} = \lambda_{42}$ $\lambda_{52} = \lambda_{53}$
3. Strong factorial invariance	All of the above constraints, in addition to: Fix corresponding intercepts across time.	$\tau_{21} = \tau_{22} = \tau_{23}$ $\tau_{41} = \tau_{42}$ $\tau_{52} = \tau_{53}$
4. Strict factorial invariance	All of the above constraints, in addition to: Fix corresponding indicator residual variances across time.	$\theta_{11} = \theta_{12} = \theta_{13}$ $\theta_{21} = \theta_{22} = \theta_{23}$ $\theta_{41} = \theta_{42}$ $\theta_{52} = \theta_{53}$

Note: Suggestions for scale-setting and testing longitudinal factorial invariance with changing measures adapted from Widaman et al. (2010). "Corresponding" refers to same indicator (i.e., measure) across time. In other words, longitudinal constraints of indicators are only within measure.

empirical (longitudinal factorial invariance) approaches are necessary to ensure that the same construct is being measured across time.

If both longitudinal factorial invariance and construct validity are established, one can be more confident that differences across time reflect meaningful change and not simply differences in measurement. Longitudinal factorial invariance is generally tested in confirmatory factor analysis (CFA) using SEM (Widaman et al., 2010). Tests of measurement or factorial invariance are critical whenever making comparisons across groups or time periods, as in the case of longitudinal assessment as exemplified by Odgers et al. (2008). Because the focus of this paper is assessing the same construct with *different* measures, the remaining discussion is framed in terms of longitudinal factorial invariance in CFA that can be tested in these steps:

Steps to test longitudinal factorial invariance:

(a) Set up the CFA. Set up a longitudinal CFA in SEM, and represent the construct of interest (e.g., inhibitory control) by a latent factor at each age that is composed of indicators for all of the measures, including behavioral tasks, questionnaires, etc. A longitudinal CFA requires at least two indicators for every latent factor (Ferrer et al., 2008; Widaman et al., 2010). One of the key advantages of SEM is that it allows correlated residuals. It is important to try correlating the residuals of within-measure indicators across time because they often have common measure-specific variances (Little et al., 2007). For instance, residuals for Shape Stroop should be allowed to correlate with each other across time because the Shape Stroop task has construct-irrelevant qualities that are unique from other inhibitory control tasks (i.e., error variance).

(b) Determine whether the measures show longitudinal factorial invariance. There are four types of longitudinal factorial invariance (see Table 5 for a definition and description of longitudinal constraints as applied to the baseline model in Fig. 5), ordered from weak to strong: (1) configural invariance, (2) weak factorial invariance, (3) strong factorial invariance, and (4) strict factorial invariance. There are specific considerations for testing longitudinal factorial invariance in CFA with changing measures. The changing measures across time can be considered a form of missingness, where all of the measures from any measurement occasion inform the structure of the construct at every time point,

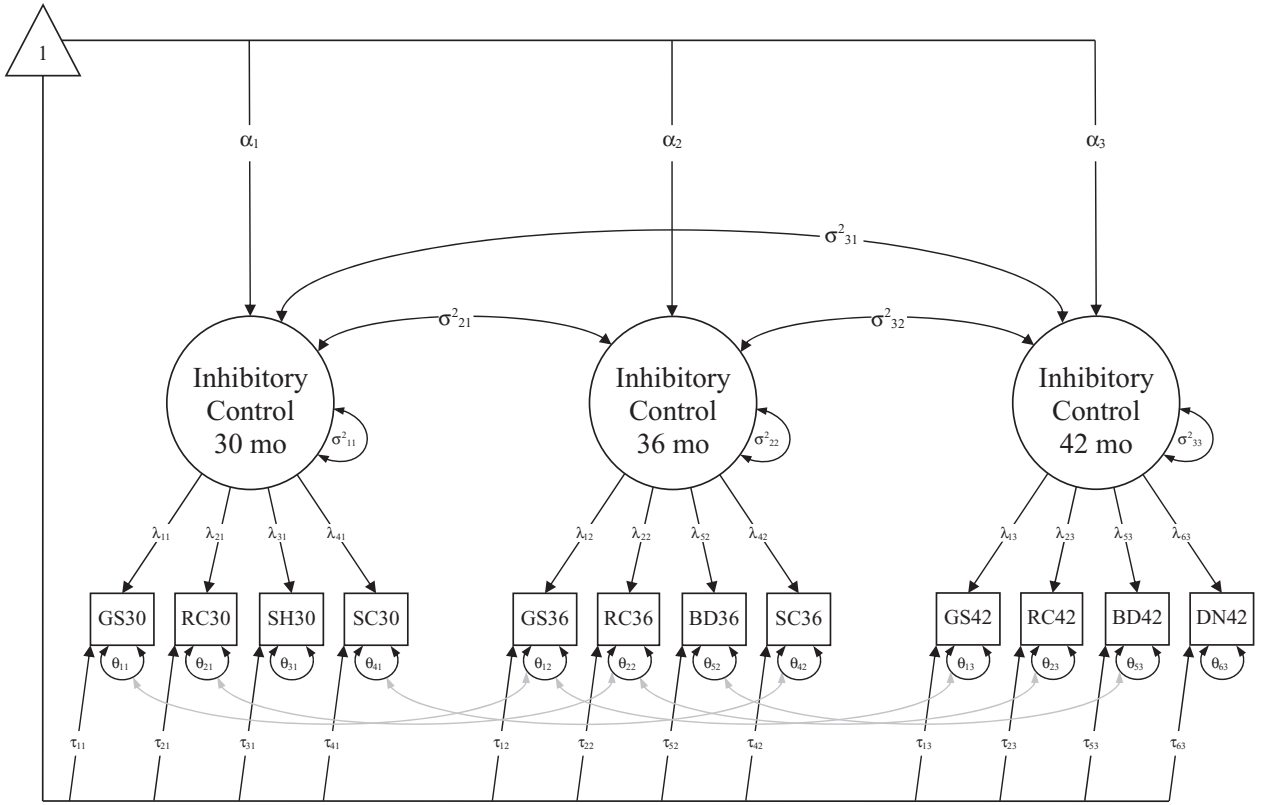

Fig. 5. Example baseline longitudinal CFA model of inhibitory control adapted from Widaman et al. (2010) for use with changing measures over time. In the example model, the four measures at each time point (rectangles) serve as manifest variables (indicators) of the latent variable reflecting inhibitory control at each time point (ovals). The example baseline model is an abridged CFA with only three time points for pictorial simplicity. A full test of longitudinal factorial invariance would incorporate all nine measurement occasions from Table 6. See Table 5 for a list of the constraints added to the baseline model to test longitudinal factorial invariance. Constraints are only made within measure. For factor loadings (λ), residual variances (θ), and intercepts (τ), the first numerical digit refers to the measure (e.g., 2 from λ_{23} = Grass/Snow) and the second digit refers to the occasion of measurement (e.g., 3 from λ_{23} = Time 3: 42 months). For covariances (σ^2), the first numerical digit refers to the occasion of measurement from one end of the arrow and the second digit refers to the occasion of measurement of the other arrow. The covariance of a variable with itself represents that variable's variance. The triangle represents the estimation of means (intercepts), including the means of the latent factors (α) and the intercepts of the indicators (τ). "mo" = child's age in months, BD = Bear/Dragon, DN = Day/Night, GS = Grass/Snow, RC = Reverse Categorization, SH = Shape Stroop, SC = Spatial Conflict.

Table 6
Changing, yet developmentally-sensitive measures of inhibitory control in example longitudinal study.

Age (years)	Measure 1	Measure 2	Measure 3	Measure 4	Facets
2	Spatial Conflict	Shape Stroop	Reverse Categorization	–	Perceptual, association
2.5	Spatial Conflict	Shape Stroop	Reverse Categorization	**Grass/Snow**	Perceptual, association
3	Spatial Conflict	**Bear/Dragon**	Reverse Categorization	Grass/Snow	Perceptual, association, performance
3.5	**Day/Night**	Bear/Dragon	Reverse Categorization	Grass/Snow	Association, performance
4	Day/Night	Bear/Dragon	**Inhibitory Control (CBQ)**	Grass/Snow	Association, performance, temperament
4.5	Day/Night	Bear/Dragon	Inhibitory Control (CBQ)	Grass/Snow	Association, performance, temperament
5	Day/Night	**Simon Says**	Inhibitory Control (CBQ)	Grass/Snow	Association, performance, temperament
5.5	Day/Night	Simon Says	Inhibitory Control (CBQ)	**Less is More**	Association, performance, temperament, motivational
6	–	Simon Says	Inhibitory Control (CBQ)	Less is More	Association, performance, temperament, motivational

Note: Introduction of different, more age-appropriate measures in bold. All measures are conjoined by two or more consecutive time points. No more than one measure is altered between two adjacent time points. Moreover, it would be ideal for a novel task to measure the same facet of inhibitory control as the preceding one that it replaced, and for the facets measured to reflect the heterotypic continuity of the construct.

but with missing measures at various measurement occasions (McArdle et al., 2009). As a result, longitudinal constraints should only be made within measure (see, e.g., Pettit et al., 2007) because all measures are different indicators of the same construct with different patterns of missingness. For example, if Spatial Conflict is measured from ages 2 to 3 years, and Day/Night replaces it after 3 years of age (e.g., see Table 6), Spatial Conflict and Day/Night should not be constrained to have the same factor loadings, intercepts, or residual variances.

Starting with a baseline unconstrained CFA model (similar to the one in Fig. 5), one would then test a sequence of levels of longitudinal factorial invariance by successively adding longitudinal within-measure constraints to the model starting with configural invariance and testing up to strict factorial invariance according to Table 5 or until longitudinal factorial invariance fails. Because the additional constraints result in nested models, the model fit between the baseline model and the model with additional constraints can be compared with the difference in chi-square between the two models with as many degrees of freedom as there are additional constraints (Little et al., 2007). If the additional constraints result in a significant chi-square change (i.e., a significant worsening in model fit), the model with additional longitudinal constraints is considered to be a significantly worse fit for the data, and the model has failed that particular type of measurement invariance. Because the chi-square change statistic is sensitive to sample size, the recommended criteria for evaluating model fit for studies with larger samples are a change in the comparative fit index (ΔCFI) greater than .01 (Cheung & Rensvold, 2002).

If longitudinal factorial invariance holds for all of the types of measurement invariance (see Table 5), the empirical structure of relations suggests that the variables are comparable across time. If longitudinal factorial invariance fails, however, the empirical evidence suggests that the variables may not be comparable across time and that there is a good chance of comparing "apples" to "oranges" when examining change over time (Odgers et al., 2008). Although different researchers use different guidelines to determine measurement invariance, Little et al. (2007) suggested that researchers establish at least partial weak longitudinal factorial invariance to examine covariances between latent variables and partial strong longitudinal factorial invariance to examine mean-level change in latent variables ("partial" invariance refers to invariance with some indicators but failed invariance with others; Ferrer et al., 2008), whereas Widaman et al. (2010) suggested establishing at least strong factorial invariance. For example, to examine whether inhibitory control is related to the development of attention deficit hyperactivity disorder, longitudinal factorial invariance will be considered to have failed if not

achieving at least weak factorial invariance (the same factor loadings across time) for some of the measures of inhibitory control to serve as anchors. For the goal of examining whether mean levels of inhibitory control change as children get older, longitudinal factorial invariance will be considered to have failed if at least strong factorial invariance (the same factor loadings and intercepts/means across time) is not achieved for some of the measures. In our example study of mean-level growth in inhibitory control (described later), we would need to establish the same factor loadings and intercepts for at least one of the four combinations of measures (see Table 6): (1) Spatial Conflict and Day/Night, (2) Shape Stroop, Bear/Dragon, and Simon Says, (3) Reverse Categorization and the Inhibitory Control subscale from the CBQ, and (4) Grass/Snow and Less is More. The more combinations that demonstrate the same factor loadings and intercepts, the more confident we can be in measurement comparability.

(c) Make any necessary adjustments. If longitudinal factorial invariance fails, therefore resulting in the lack of a statistical basis to assert longitudinal equivalence of measures, the researcher has several options that should be guided by theoretical and empirical considerations. To reiterate, longitudinal factorial invariance alone is not sufficient evidence that one has measured the same construct over time; see construct validity below. Theoretically, there are cases in which one might expect changes in a construct's manifestation over time. For example, self-regulation and inhibitory control change over development from early reliance on external sources for control to ability for internal, self-initiated forms of control. These changes in the construct's manifestation necessitate changes in measurement, which may result in failed longitudinal factorial invariance.

Empirically, one should test when (at what ages, Odgers et al., 2008) and with what measures (Millsap, 2010) the assumption of factorial invariance failed. To determine where factorial invariance failed, one can test a piecewise longitudinal CFA, gradually including successive years or tasks until the measurement assumption fails. Alternatively, one can gradually relax parameter constraints for longitudinal factorial invariance to determine with which measures or at what ages longitudinal factorial invariance becomes established (and by extension, where it fails). It will then be up to the researcher to determine, ideally based on additional research, whether the failed longitudinal factorial invariance resulted from change in the manifestation of the construct, changes in the functioning of the measures, or some combination of the two (Knight & Zerr, 2010), because there is no statistical method for choosing between these alternatives. Unfortunately, there are few empirical guidelines for determining the severity of failed factorial invariance (i.e., whether it owes to changes in the functioning of a measure; Knight & Zerr, 2010), although some have proposed using effect sizes or other estimates (Millsap, 2010). If the factorial invariance violations show important differences in the meaning of measures, say at a few of the ages in a series of measures of a construct, it may be necessary to exclude the measures at those ages from the analysis. Alternatively, one might substitute other measures reflecting the latent construct at those ages, as long as the set of measures under consideration remains a representative selection for the construct (Knight & Zerr, 2010).

We know of a few studies that have examined longitudinal factorial invariance of inhibitory control tasks. Studies have typically examined longitudinal factorial invariance of inhibitory control tasks in the context of the broader construct of executive function. Willoughby et al. (2012) found strong longitudinal factorial invariance for variants of the Day/Night and the Spatial Conflict task in relation to a latent variable composite of six executive function tasks from 3 to 5 years of age. The authors found partial strong longitudinal factorial invariance for the executive function composite across the same time frame. Fuhs and Day (2011) found partial strict longitudinal factorial invariance across 43–63 months of age for a latent composite of three executive function tasks that included the Day/Night task. Hughes, Ensor, Wilson, and Graham (2009) found full strong longitudinal factorial invariance from 48 to 72 months of age using a latent composite of three executive function tasks that included the Day/Night task.

In sum, there is some evidence that one can establish *partial* longitudinal factorial invariance with a latent composite that includes inhibitory control tasks over a fairly limited age range, but limited evidence of *full* longitudinal factorial invariance. This is consistent with the notion (but not sufficient to demonstrate) that the same task may reflect the same homotypic construct over a restricted developmental span, and that additional different measures are necessary to capture its heterotypic continuity over time. The prevailing bias for publishing studies may be that they must demonstrate

longitudinal factorial invariance; studies covering lengthy developmental spans, however, may fail traditional criteria for longitudinal factorial invariance and may be less likely to be published. More studies of inhibitory control are needed that test, and possibly fail, longitudinal factorial invariance in order to understand how inhibitory control changes in manifestation with development.

Construct validity invariance

As mentioned before, longitudinal factorial invariance is not sufficient for evaluating whether one is measuring the same construct over time. Construct validity invariance should also be considered. Widaman et al. (2010) described a series of studies on the construct of numerical facility. Whereas second graders used reconstructive, counting strategies for simple addition, complex addition, and subtraction problems, the college students used memory retrieval strategies. Thus, the construct of numerical facility changed in manifestation, resulting in changes in the meaning of the measures, despite having established strict factorial invariance. Thus, to demonstrate that measures at different time points reflect the same processes, evaluating the construct validity of measures at each age as reflections of the same construct of interest, termed *construct validity invariance*, is also important (Knight & Zerr, 2010).

Construct validity includes both content and criterion validity (Strauss & Smith, 2009). There are many ways to explore the construct validity of a set of measures for a given construct. First, the task selection should be based on theory – the tasks should reflect the same construct and the tasks should adequately sample the different facets of the construct (content and face validity). Second, despite theoretically expected heterotypic continuity in the long-term, as in the case of inhibitory control, there should be short-term test–retest reliability (cross-time consistency) of the measures across time. Third, the measures should show convergent validity with each other and discriminant validity with measures of distinct constructs according to Campbell and Fiske's (1959) example of the multi-trait multi-method matrix (for a contemporary longitudinal example, see Grimm, Pianta, & Konold, 2009). The measures should show stronger associations with measures of the same construct and weaker associations with measures of different constructs. For example, measures of inhibitory control should be more strongly associated with each other than with measures of sustained attention (after discarding method variance). Fourth, the measures should show a similar factor structure across time (not necessarily an invariant structure in the case of heterotypic continuity). Fifth, the measures should have high internal consistency (communality or equivalence) as measured by Cronbach's alpha (Cortina, 1993) or McDonald's total omega (Revelle & Zinbarg, 2009). Sixth, the measures should be associated with measures of other constructs in theoretically expected ways (criterion validity) as part of its nomological network, particularly when examining heterotypic continuity (Rutter & Sroufe, 2000). Despite these ways of developing evidence for or against the validity of a set of measures for a given construct, construct validity is inherently indefinite and is a continual process, not a discrete outcome (Strauss & Smith, 2009).

The final decision: whether to test a growth model

Taking into account the above recommendations and considerations, the researcher now must decide whether the totality of theoretical and empirical evidence supports the goal to measure the same construct over time with comparable measures. If researchers find that the theoretical and empirical evidence supports the ability to draw developmental inferences from differences over time in the measurement model, they can now test a growth curve model (or other longitudinal model) with changing measures. Although we do not provide an empirical example of the proposed approach in the current paper, there are previous examples of successful and useful implementations of similar approaches, suggesting that the proposed approach is feasible and defensible. Approaches with changing measures have been successful with diverse constructs such as externalizing problems (e.g., Petersen et al., 2015), intelligence (e.g., McArdle & Grimm, 2011; McArdle et al., 2009), and parental monitoring (e.g., Pettit et al., 2007). For examples of longitudinal growth curve models with changing measures, see Pettit et al. (2007) and Owens and Shaw (2003), or with changing raters, see Odgers et al. (2008), or with changing measures and raters, see Petersen et al. (2015). Testing a growth curve model allows

the researcher to observe how children's inhibitory control develops over time, and how inhibitory control development predicts meaningful developmental outcomes or is predicted by important risk or protective factors. For discussions of different types of longitudinal models and when to use them, see Singer and Willett (2003), McArdle (2009), and Fontaine and Petersen (2016).

Applying the recommendations to a hypothetical example of a longitudinal study of inhibitory control

To illustrate how to measure the development of inhibitory control using the above recommendations, we provide a hypothetical example of longitudinal assessment. We focus on the development of inhibitory control from ages 2 to 6 years, because previous studies have shown extensive development of inhibitory control (operationalized with measures in Table 1) during this time frame (Carlson, 2005; Carlson, Davis, & Leach, 2005; Carlson, Mandell, & Williams, 2004; Carlson & Moses, 2001; Carlson, Moses, & Claxton, 2004; Gerardi-Caulton, 2000; Hughes & Ensor, 2007; Kochanska, Murray, & Coy, 1997; Kochanska et al., 1996, 2000), supported by the findings from our meta-analysis (see Figs. 1–3). Many prior studies of inhibitory control conducted assessments annually or less frequently (e.g., Carlson et al., 2005; Murray & Kochanska, 2002), which may be insensitive to some forms of meaningful developmental change. Because Carlson (2005) noted considerable improvement in inhibitory control with development in a cross-sectional study with 6-month age groups from ages 3 to 5 (their full sample was from ages 2 to 8), we decided, for purposes of the hypothetical example, to propose assessments every six months from ages 2 to 6 years.

After selecting our construct and choosing the age range and frequency of assessment, we chose developmentally-sensitive measures from Table 1 that met our goals, including construct validity at each age, multiple tasks at each age, multiple measurement methods, overlapping time points for measures, at least two consecutive time points for each measure, and no more than one measurement change at any time point. See Table 6 for our selection of measures at each age. Consistent with the heterotypic continuity of inhibitory control, we selected measures that reflect the different facets of inhibitory control that are "coming online" at different points in development. For instance, we selected more perceptual inhibition tasks among the measures at earlier ages and motivational inhibition tasks among the measures at later ages, consistent with findings from our meta-analysis that perceptual inhibition may develop before other facets of inhibitory control. The measures of inhibitory control, however, are not unidimensional and reflect other processes including working memory. There have been attempts to dissociate working memory from inhibitory control in behavioral tasks (e.g., Carlson, Moses, & Breton, 2002; Raaijmakers et al., 2008; Senn, Espy, & Kaufmann, 2004), with studies commonly finding working memory and inhibitory control as correlated but separate factors. Thus, researchers have argued that task demands in behavioral tasks of self-regulation are dissociable (Carlson, 2005). As a result, a plausible next step would be to combine the measures in an SEM framework to create more unidimensional measures of inhibitory control.

In SEM, we would test longitudinal factorial invariance in CFA. First, we would start with the baseline model (see Fig. 5 for an abridged model with only three time points). Second, we would test models with various longitudinal constraints in sequence (see Table 5) to determine where our factorial invariance holds and where it fails. Following determination of longitudinal factorial invariance, evaluation of violations, and any changes to establish factorial invariance while ensuring construct validity invariance, we would then turn our changing measures into a longitudinal growth curve model.

Limitations

Even if researchers follow the empirical recommendations in the present paper, there is no guarantee that the measurement model will pass tests of longitudinal factorial invariance or will reflect the same construct over time. Establishing longitudinal factorial invariance may be difficult with changing measurement over time, especially with common methodological limitations such as high levels of missingness (McArdle et al., 2009). In all likelihood, given adequate construct validity of a set of measures, using more measures over a longer consecutive period of time (albeit within developmentally-appropriate and -sensitive age ranges) with fewer changes will be more likely to provide a meaningful

longitudinal model. Moreover, as discussed earlier, establishing longitudinal factorial invariance does not necessarily mean the same construct has been assessed over time, with or without changing measures. Thus, researchers should be cautious when conducting longitudinal assessments to ensure that their measures have construct validity for the construct of interest at each assessment occasion in which they are used and that the measures are developmentally appropriate and sensitive for the same time frame.

In the current paper, we proposed using an overlapping measurement model in SEM to measure the heterotypic continuity of inhibitory control. We also recognize that there are other possibly useful approaches for linking changing measurement over time (especially with dichotomous or polytomous items) and testing longitudinal factorial invariance, including integrative data analysis (Curran et al., 2008) and IRT (Kolen & Brennan, 2014; McArdle et al., 2009; Meade, Lautenschlager, & Hecht, 2005; Millsap, 2010). The developmental cascade model is another approach to longitudinal measurement where a construct changes in manifestation over time as a function of changing influences that lead to other risks (Cox, Mills-Koonce, Propper, & Gariépy, 2010; Dodge, Greenberg, Malone, & Conduct Problems Prevention Research Group, 2008; Dodge et al., 2009). A discussion of these alternative approaches is beyond the scope of the current paper. For a review of other techniques for dealing with growth modeling in the context of changing measurement, see McArdle et al. (2009).

Future directions

Future research should examine how much overlap is necessary for longitudinal comparability, with methods such as Monte Carlo analysis (Pettit et al., 2007). Moreover, future studies should examine failed longitudinal factorial invariance to establish guidelines for determining when failed factorial invariance reflects change in the manifestation of a construct versus change in the meaning of a measure. Future studies might also develop measures that maintain construct validity invariance across lengthy spans of development (for attempts, see Bauer & Zelazo, 2014; Carlson & Zelazo, 2014; Weintraub et al., 2013). Future work should consider how best to measure inhibitory control across multiple levels of analysis, including bio-psycho-social processes, in ways that are consistent with its heterotypic continuity. For instance, children's inhibitory control and self-regulation develop through the internalization of rules in the context of relationships with parents, siblings, caregivers, and teachers. Thus, future work should consider how to account for environmental influences in the heterotypic development of inhibitory control. Future applications of our recommendations might examine when and how inhibitory control develops over time in young children, how growth in inhibitory control is influenced by other risk and protective factors, and how inhibitory control influences other outcomes. Broader applications of the proposed technique for longitudinal measurement with changing measures may be relevant for many other domains involving constructs that change in manifestation over time (e.g., intelligence, McArdle et al., 2009; externalizing problems, Petersen et al., 2015).

Implications

Inhibitory control and related processes are among the most widely studied constructs in early childhood. As researchers place greater emphasis on measuring inhibitory control using multiple measures, we need to understand how these measures function at different ages, and when they are most useful. Our work is an important step toward understanding the heterotypic continuity of inhibitory control and the age ranges during which measures of inhibitory control are most useful. Using measures in a longitudinal context and testing longitudinal factorial invariance will lead researchers to place greater emphasis on the developmental utility of measures and their construct validity. Stronger longitudinal assessment of inhibitory control will be important for making inferences about the development of inhibitory control. At the same time, no empirical approach can compensate for the necessity of an informed theoretical approach to the selection and structure of measures in a longitudinal measurement model. Unfortunately, there is no way to be entirely certain that one is measuring the same thing across time. Thus, it is imperative to "consider all theoretical and empirical evidence that can strengthen our confidence in whether we are assessing the same construct … in a psychometrically equivalent way" (Curran & Hussong, 2009, p. 93). The suggested approach may also allow for greater

coherence across studies so that results about inhibitory control development can be interpreted in meaningful ways. A more theoretically- and empirically-informed approach to longitudinal assessment has the potential to vastly improve our understanding of the construct of inhibitory control from a developmental perspective and its relation to other important predictors and outcomes.

Summary

Current theory suggests that inhibitory control changes in its manifestation over time (heterotypic continuity). The findings from our meta-analysis advance our understanding of the heterotypic continuity of inhibitory control, with perceptual inhibition developing at earlier ages and motivational inhibition possibly developing at later ages. Further, measures of inhibitory control were generally useful for a developmental span of less than 3 years. As a result, in order to properly measure the development of inhibitory control, appropriate changes in measurement should accompany developmental changes in the manifestation of the construct. To deal with changing measures over time, we recommend a longitudinal approach to studying the development of inhibitory control that involves overlapping measurement in an SEM framework to test longitudinal factorial invariance. Although longitudinal factorial invariance should be tested, establishing measurement or factorial invariance is neither sufficient, as seen where construct invariance does not hold despite establishing factorial invariance, nor necessary, as seen in the case of heterotypic continuity, for examining developmental change, because any developmental approach should be guided by construct validity. In combination, a better empirical and theoretical approach to the selection and combination of measures will lead to improved developmental sensitivity of measurement, and will improve our ability to make meaningful developmental inferences.

References

Achenbach, T. M. (2009). *Achenbach System of Empirically Based Assessment (ASEBA): Development, findings, theory, and applications*. University of Vermont, Research Center of Children, Youth & Families.

Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the US and China: Similarities and differences. *European Journal of Personality*, *7*, 359–377. doi:10.1002/per.2410070506.

Aksan, N., & Kochanska, G. (2004). Links between systems of inhibition from infancy to preschool years. *Child Development*, *75*, 1477–1490. doi:10.1111/j.1467-8624.2004.00752.x.

Albertson, K., & Shore, C. (2008). Holding in mind conflicting information: Pretending, working memory, and executive control. *Journal of Cognition and Development*, *9*, 390–410. doi:10.1080/15248370802678240.

Alexander, K. W., Goodman, G. S., Schaaf, J. M., Edelstein, R. S., Quas, J. A., & Shaver, P. R. (2002). The role of attachment and cognitive inhibition in children's memory and suggestibility for a stressful event. *Journal of Experimental Child Psychology*, *83*, 262–290. doi:10.1016/S0022-0965(02)00149-2.

Allan, N. P. (2010). *A confirmatory factor analysis of effortful control: Comparing a two factor model consisting of hot and cool latent variables and a general domain model* (Doctoral dissertation). Tallahassee, FL: Florida State University. <http://purl.flvc.org/fsu/fd/FSU_migr_etd-0170>.

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology*, *47*, 905–915. doi:10.1037/a0023748.

Allan, N. P., & Lonigan, C. J. (2014). Exploring dimensionality of effortful control using hot and cool tasks in a sample of preschool children. *Journal of Experimental Child Psychology*, *122*, 33–47. doi:10.1016/j.jecp.2013.11.013.

Allan, N. P., Lonigan, C. J., & Wilson, S. B. (2013). Psychometric evaluation of the Children's Behavior Questionnaire-Very Short Form in preschool children using parent and teacher report. *Early Childhood Research Quarterly*, *28*, 302–313. doi:10.1016/j.ecresq.2012.07.009.

Anderson, M., Nettelbeck, T., & Barlow, J. (1997). Reaction time measures of speed of processing: Speed of response selection increases with age but speed of stimulus categorization does not. *The British Journal of Developmental Psychology*, *15*, 145–157. doi:10.1111/j.2044-835X.1997.tb00731.x.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, *5*, 115–129. doi:10.1076/chin.5.2.115.3167.

Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science*, *14*, 270–279. doi:10.1111/j.1467-7687.2010.00972.x.

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

Bates, J. E., & Novosad, C. (2005). Measurement of individual difference constructs in child development, or taking aim at moving targets. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 103–122). Blackwell Publishing Ltd.

Bauer, P. J., & Zelazo, P. D. (2014). The National Institutes of Health Toolbox for the Assessment of Neurological and Behavioral Function: A tool for developmental science. *Child Development Perspectives*, *8*, 119–124. doi:10.1111/cdep.12080.

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development*, *12*, 169–193. doi:10.1080/15248372.2011.563485.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking & Reasoning*, *15*, 337–354. doi:10.1080/13546780903135904.
Bell, M. A., & Wolfe, C. D. (2007a). Changes in brain functioning from infancy to early childhood: Evidence from EEG power and coherence working memory tasks. *Developmental Neuropsychology*, *31*, 21–38. doi:10.1207/s15326942dn3101_2.
Bell, M. A., & Wolfe, C. D. (2007b). The cognitive neuroscience of early socioemotional development. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in toddler years: Transitions and transformations* (pp. 345–369). New York: Guilford.
Bellagamba, F., Laghi, F., Lonigro, A., & Pace, C. (2012). Re-enactment of intended acts from a video presentation by 18- and 24-month-old children. *Cognitive Processing*, *13*, 381–386. doi:10.1007/s10339-012-0518-0.
Bellagamba, F., Laghi, F., Lonigro, A., Pace, C. S., & Longobardi, E. (2013). Concurrent relations between inhibitory control, vocabulary and internal state language in 18- and 24-month-old Italian-speaking infants. *The European Journal of Developmental Psychology*, *11*, 420–432. doi:10.1080/17405629.2013.848164.
Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology*, *49*, 1615–1627. doi:10.1037/a0031056.
Berger, A. (2011). *Self-regulation: Brain, cognition, and development*. Washington, DC, USA: American Psychological Association.
Berger, A., Jones, L., Rothbart, M. K., & Posner, M. (2000). Computerized games to study the development of attention in childhood. *Behavior Research Methods, Instruments, & Computers*, *32*, 297–303. doi:10.3758/BF03207798.
Bernier, A., Beauchamp, M. H., Bouvette-Turcot, A.-A., Carlson, S. M., & Carrier, J. (2013). Sleep and cognition in preschool years: Specific links to executive functioning. *Child Development*, *84*, 1542–1553. doi:10.1111/cdev.12063.
Bernier, A., Carlson, S. M., Bordeleau, S., & Carrier, J. (2010). Relations between physiological and cognitive regulatory systems: Infant sleep regulation and subsequent executive functioning. *Child Development*, *81*, 1739–1752. doi:10.1111/j.1467-8624.2010.01507.x.
Bernier, A., Carlson, S. M., Deschênes, M., & Matte-Gagné, C. (2012). Social factors in the development of early executive functioning: A closer look at the caregiving environment. *Developmental Science*, *15*, 12–24. doi:10.1111/j.1467-7687.2011.01093.x.
Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development*, *81*, 326–339. doi:10.1111/j.1467-8624.2009.01397.x.
Bernstein, D. M., Atance, C., Meltzoff, A. N., & Loftus, G. R. (2007). Hindsight bias and developing theories of mind. *Child Development*, *78*, 1374–1394. doi:10.1111/j.1467-8624.2007.01071.x.
Berwid, O. G., Curko Kera, E. A., Marks, D. J., Santra, A., Bender, H. A., & Halperin, J. M. (2005). Sustained attention and response inhibition in young children at risk for Attention Deficit/Hyperactivity Disorder. *Journal of Child Psychology and Psychiatry*, *46*, 1219–1229. doi:10.1111/j.1469-7610.2005.00417.x.
Bibok, M. B. (2007). *Social interactive quality of parent-child scaffolding as a predictor of children's executive function* (Master's thesis). Burnaby, Canada: Simon Fraser University. <http://summit.sfu.ca/item/8154>.
Bibok, M. B., Carpendale, J. I. M., & Müller, U. (2009). Parental scaffolding and the development of executive function. *New Directions for Child and Adolescent Development*, *2009*, 17–34. doi:10.1002/cd.233.
Bigham, S. (2010). Impaired competence for pretense in children with autism: Exploring potential cognitive predictors. *Journal of Autism and Developmental Disorders*, *40*, 30–38. doi:10.1007/s10803-009-0820-6.
Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *The American Psychologist*, *57*, 111–127. doi:10.1037/0003-066x.57.2.111.
Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., et al. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development*, *82*, 1970–1984. doi:10.1111/j.1467-8624.2011.01643.x.
Blair, C., & Peters, R. (2003). Physiological and neurocognitive correlates of adaptive behavior in preschool among children in head start. *Developmental Neuropsychology*, *24*, 479–497. doi:10.1207/S15326942DN2401_04.
Blair, C., Peters, R., & Granger, D. (2004). Physiological and neuropsychological correlates of approach/withdrawal tendencies in preschool: Further examination of the behavioral inhibition system/behavioral activation system scales for young children. *Developmental Psychobiology*, *45*, 113–124. doi:10.1002/dev.20022.
Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–663. doi:10.1111/j.1467-8624.2007.01019.x.
Bohm, B., Smedler, A. C., & Forssberg, H. (2004). Impulse control, working memory and other executive functions in preterm children when starting school. *Acta Paediatrica*, *93*, 1363–1371. doi:10.1080/08035250410021379.
Bolnick, R. R. (2008). *Contributions of children's theory of mind, inhibitory control and social competence to school adjustment* (Doctoral dissertation). Temple, AZ: Arizona State University. ProQuest Dissertations and Theses database, UMI No. 3341317.
Brocki, K. C., Nyberg, L., Thorell, L. B., & Bohlin, G. (2007). Early concurrent and longitudinal symptoms of ADHD and ODD: Relations to different types of inhibitory control and working memory. *Journal of Child Psychology and Psychiatry*, *48*, 1033–1041. doi:10.1111/j.1469-7610.2007.01811.x.
Brooker, R. J. (2008). *Individual differences in toddler's regulatory abilities: Evidence for links among systems* (Doctoral dissertation). State College, PA: Penn State University. <https://etda.libraries.psu.edu/paper/9222/>.
Brown, E. D., Ackerman, B. P., & Moore, C. A. (2013). Family adversity and inhibitory control for economically disadvantaged children: Preschool relations and associations with school readiness. *Journal of Family Psychology*, *27*, 443–452. doi:10.1037/a0032886.
Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, *14*, 253–262. doi:10.1037/1040-3590.14.3.253.
Calderon, J., Angeard, N., Moutier, S., Plumet, M.-H., Jambaqué, I., & Bonnet, D. (2012). Impact of prenatal diagnosis on neurocognitive outcomes in children with transposition of the great arteries. *The Journal of Pediatrics*, *161*, 94–98, e91.
Calderon, J., Jambaqué, I., Bonnet, D., & Angeard, N. (2014). Executive functions development in 5- to 7-year-old children with transposition of the great arteries: A longitudinal study. *Developmental Neuropsychology*, *39*, 365–384. doi:10.1080/87565641.2014.916709.

Calkins, S. D., & Fox, N. A. (2002). Self-regulatory processes in early personality development: A multilevel approach to the study of childhood social withdrawal and aggression. *Development and Psychopathology*, *14*, 477–498. doi:10.1017/S095457940200305X.
Calkins, S. D., & Howse, R. B. (2004). Individual differences in self-regulation: Implications for childhood adjustment. In P. Philippot & R. S. Feldman (Eds.), *The regulation of emotion* (pp. 307–332). Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers.
Cammock, T., & Cairns, E. D. (1979). Concurrent validity of a children's version of the Stroop Color-Word Test: The fruit distraction test. *Perceptual and Motor Skills*, *49*, 611–616. doi:10.2466/pms.1979.49.2.611.
Camp, B. W. (1977). Verbal mediation in young aggressive boys. *Journal of Abnormal Psychology*, *86*, 145–153. doi:10.1037/0021-843X.86.2.145.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. doi:10.1037/h0046016.
Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*, 595–616. doi:10.1207/s15326942dn2802_3.
Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science*, *16*, 609–616. doi:10.1111/j.1467-9280.2005.01583.x.
Carlson, S. M., Faja, S., & Beck, D. M. (2015). 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*.
Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*, 1105–1122. doi:10.1037/0012-1649.40.6.1105.
Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*, 282–298. doi:10.1111/j.1467-7687.2008.00675.x.
Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053. doi:10.1111/1467-8624.00333.
Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73–92. doi:10.1002/icd.298.
Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*, 299–319. doi:10.1016/j.jecp.2004.01.002.
Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development*, *22*, 489–510. doi:10.1016/j.cogdev.2007.08.002.
Carlson, S. M., & Zelazo, P. D. (2014). *Minnesota executive function scale. Test manual*. St. Paul, MN: Reflection Sciences, LLC.
Carranza, J. A., González-Salinas, C., & Ato, E. (2013). A longitudinal study of temperament continuity through IBQ, TBAQ and CBQ. *Infant Behavior and Development*, *36*, 749–761. doi:10.1016/j.infbeh.2013.08.002.
Carroll, D. J., Riggs, K. J., Apperly, I. A., Graham, K., & Geoghegan, C. (2012). How do alternative ways of responding influence 3- and 4-year-olds' performance on tests of executive function and theory of mind? *Journal of Experimental Child Psychology*, *112*, 312–325. doi:10.1016/j.jecp.2012.03.001.
Caspi, A. (2006). Personality development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (6th ed., Vol. 3). Social, emotional, and personality development. Hoboken, NJ, USA: John Wiley & Sons, Inc.
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*, 839–855. doi:10.1016/j.jecp.2013.07.017.
Causey, K. B., & Bjorklund, D. F. (2014). Prospective memory in preschool children: Influences of agency, incentive, and underlying cognitive mechanisms. *Journal of Experimental Child Psychology*, *127*, 36–51. doi:10.1016/j.jecp.2014.01.020.
Chang, H., Olson, S. L., Sameroff, A. J., & Sexton, H. R. (2011). Child effortful control as a mediator of parenting practices on externalizing behavior: Evidence for a sex-differentiated pathway across the transition from preschool to school. *Journal of Abnormal Child Psychology*, *39*, 71–81. doi:10.1007/s10802-010-9437-7.
Chasiotis, A., Kiessling, F., Hofer, J., & Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. *International Journal of Behavioral Development*, *30*, 249–260. doi:10.1177/0165025406066759.
Chasiotis, A., Kiessling, F., Winter, V., & Hofer, J. (2006). Sensory motor inhibition as a prerequisite for theory-of-mind: A comparison of clinical and normal preschoolers differing in sensory motor abilities. *International Journal of Behavioral Development*, *30*, 178–190. doi:10.1177/0165025406063637.
Cheie, L., Veraksa, A., Zinchenko, Y., Gorovaya, A., & Visu-Petra, L. (2015). A cross-cultural investigation of inhibitory control, generative fluency, and anxiety symptoms in Romanian and Russian preschoolers. *Child Neuropsychology*, *21*, 121–149. doi:10.1080/09297049.2013.879111.
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233–255. doi:10.1207/s15328007sem0902_5.
Chien, R. (2014). *Effects of sleep schedule on training of executive function skills* (Unpublished doctoral dissertation). Bloomington, IN, USA: Indiana University.
Choe, D. E., Lane, J. D., Grabell, A. S., & Olson, S. L. (2013). Developmental precursors of young school-age children's hostile attribution bias. *Developmental Psychology*, *49*, 2245–2256. doi:10.1037/a0032293.
Choe, D. E., Olson, S. L., & Sameroff, A. J. (2013). Effects of early maternal distress and parenting on the development of children's self-regulation and externalizing behavior. *Development and Psychopathology*, *25*, 437–453. doi:10.1017/S0954579412001162.
Choe, D. E., Olson, S. L., & Sameroff, A. J. (2014). Effortful control moderates bidirectional effects between children's externalizing behavior and their mothers' depressive symptoms. *Child Development*, *85*, 643–685. doi:10.1111/cdev.12123.
Choe, D. E., Shaw, D. S., Brennan, L. M., Dishion, T. J., & Wilson, M. N. (2014). Inhibitory control as a mediator of bidirectional effects between early oppositional behavior and maternal depression. *Development and Psychopathology*, *26*(4 Pt. 1), 1129–1147. doi:10.1017/S0954579414000613.
Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, *70*, 6–20. doi:10.1037/0022-006X.70.1.6.

Cipriano-Essel, E., Skowron, E. A., Stifter, C. A., & Teti, D. M. (2013). Heterogeneity in maltreated and non-maltreated preschool children's inhibitory control: The interplay between parenting quality and child temperament. *Infant and Child Development*, *22*, 501–522. doi:10.1002/icd.1801.

Clark, C. A. C., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*, 662–677. doi:10.1111/j.1467-8624.2012.01854.x.

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *The British Journal of Developmental Psychology*, *18*, 279–295. doi:10.1348/026151000165698.

Conway, A., & Stifter, C. A. (2012). Longitudinal antecedents of executive function in preschoolers. *Child Development*, *83*, 1022–1036. doi:10.1111/j.1467-8624.2012.01756.x.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98–104. doi:10.1037/0021-9010.78.1.98.

Cox, M. J., Mills-Koonce, R., Propper, C., & Gariépy, J.-L. (2010). Systems theory and cascades in developmental psychopathology. *Development and Psychopathology*, *22*, 497–506. doi:10.1017/s0954579410000234.

Cromer, L. D., Stevens, C., DePrince, A. P., & Pears, K. (2006). The relationship between executive attention and dissociation in children. *Journal of Trauma & Dissociation*, *7*, 135–153. doi:10.1300/J229v07n04_08.

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

Cuevas, K., Deater-Deckard, K., Kim-Spoon, J., Wang, Z., Morasch, K. C., & Bell, M. A. (2014). A longitudinal intergenerational analysis of executive functions during early childhood. *The British Journal of Developmental Psychology*, *32*, 50–64. doi:10.1111/bjdp.12021.

Cuevas, K., Deater-Deckard, K., Kim-Spoon, J., Watson, A. J., Morasch, K. C., & Bell, M. A. (2014). What's mom got to do with it? Contributions of maternal executive function and caregiving to the development of executive function across early childhood. *Developmental Science*, *17*, 224–238. doi:10.1111/desc.12073.

Cuevas, K., Hubble, M., & Bell, M. A. (2012). Early childhood predictors of post-kindergarten executive function: Behavior, parent report, and psychophysiology. *Early Education and Development*, *23*, 59–73. doi:10.1080/10409289.2011.611441.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81.

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., et al. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*, 365. doi:10.1037/0012-1649.44.2.365.

Cutuli, J. J. (2011). *Context, cortisol, and executive functions among children experiencing homelessness* (Doctoral dissertation). Minneapolis, MN: University of Minnesota. <http://purl.umn.edu/116142>.

Cutuli, J. J., Herbers, J. E., Lafavor, T. L., Ahumada, S. M., Masten, A. S., & Oberg, C. N. (2014). Asthma and adaptive functioning among homeless kindergarten-aged children in emergency housing. *Journal of Health Care for the Poor and Underserved*, *25*, 717–730. doi:10.1353/hpu.2014.0099.

Cutuli, J. J., Herbers, J. E., Rinaldi, M., Masten, A. S., & Oberg, C. N. (2010). Asthma and behavior in homeless 4- to 7-year-olds. *Pediatrics*, *125*, 145–151. doi:10.1542/peds.2009-0103.

Cutuli, J. J., Wiik, K. L., Herbers, J. E., Gunnar, M. R., & Masten, A. S. (2010). Cortisol function among early school-aged homeless children. *Psychoneuroendocrinology*, *35*, 833–845. doi:10.1016/j.psyneuen.2009.11.008.

Dale, K. L., & Baumeister, R. F. (1999). Self-regulation and psychopathology. In R. M. Kowalski & M. R. Leary (Eds.), *The social psychology of emotional and behavioral problems: Interfaces of social and clinical psychology* (pp. 139–166). Washington, DC, USA: American Psychological Association.

Dalen, L., Sonuga-Barke, E. J., Hall, M., & Remington, B. (2004). Inhibitory deficits, delay aversion and preschool AD/HD: Implications for the dual pathway model. *Neural Plasticity*, *11*, 1–11.

Deák, G. O., & Narasimham, G. (2003). Is perseveration caused by inhibition failure? Evidence from preschool children's inferences about word meanings. *Journal of Experimental Child Psychology*, *86*, 194–222. doi:10.1016/j.jecp.2003.08.001.

Dennis, T. A. (2006). Emotional self-regulation in preschoolers: The interplay of child approach reactivity, parenting, and control capacities. *Developmental Psychology*, *42*, 84–97. doi:10.1037/0012-1649.42.1.84.

Dennis, T. A., Hong, M., & Solomon, B. (2010). Do the associations between exuberance and emotion regulation depend on effortful control? *International Journal of Behavioral Development*, *34*, 462–472. doi:10.1177/0165025409355514.

Di Norcia, A., Pecora, G., Bombi, A., Baumgartner, E., & Laghi, F. (2015). Hot and cool inhibitory control in Italian toddlers: Associations with social competence and behavioral problems. *Journal of Child and Family Studies*, *24*, 909–914. doi:10.1007/s10826-014-9901-z.

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. *Annals of the New York Academy of Sciences*, *608*, 637–676. doi:10.1111/j.1749-6632.1990.tb48913.x.

Diamond, A. (2001). A model system for studying the role of dopamine in the prefrontal cortex during early development in humans: Early and continuously treated phenylketonuria. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 433–472). Cambridge, MA: Massachusetts Institute of Technology.

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). New York, NY, USA: Oxford University Press.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*, 352–362. doi:10.1037/0012-1649.38.3.352.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental Psychobiology*, *29*, 315–334. doi:10.1002/(sici)1098-2302(199605)29:4<315::aid-dev2>3.0.co;2-t.

Dodge, K. A., Greenberg, M. T., Malone, P. S., & Conduct Problems Prevention Research Group (2008). Testing an idealized dynamic cascade model of the development of serious violence in adolescence. *Child Development*, *79*, 1907–1927. doi:10.1111/j.1467-8624.2008.01233.x.

Dodge, K. A., Malone, P. S., Lansford, J. E., Miller, S., Bates, J. E., & Pettit, G. S. (2009). *A dynamic cascade model of the development of substance-use onset* (Vol. 74). Boston, MA, USA: Wiley-Blackwell.
Domitrovich, C., Cortes, R., & Greenberg, M. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention*, *28*, 67–91. doi:10.1007/s10935-007-0081-0.
Drayton, S., Turley-Ames, K. J., & Guajardo, N. R. (2011). Counterfactual thinking and false belief: The role of executive function. *Journal of Experimental Child Psychology*, *108*, 532–548. doi:10.1016/j.jecp.2010.09.007.
Duncan, R. J. (2012). *Measures of executive function: Convergent validity and links to academic achievement in preschool* (Master's thesis). Corvallis, OR: Oregon State University. <http://hdl.handle.net/1957/30264>.
Durston, S., Thomas, K. M., Yang, Y., Ulug, A. M., Zimmerman, R. D., & Casey, B. J. (2002). A neural basis for the development of inhibitory control. *Developmental Science*, *5*, F9–F16. doi:10.1111/1467-7687.00235.
Duvall, S. W. (2012). *The relationship of multimodal executive function measurement and associated neuroanatomical factors in preschoolers born very low birth weight and full term* (Doctoral dissertation). Albuquerque, NM: University of New Mexico. <http://hdl.handle.net/1928/21028>.
Dyson, M. W., Olino, T. M., Durbin, C. E., Goldsmith, H. H., & Klein, D. N. (2012). The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion (Washington, D.C.)*, *12*, 44–57. doi:10.1037/a0025023.
Earhart, B., & Roberts, K. (2014). The role of executive function in children's source monitoring with varying retrieval strategies. *Frontiers in Psychology*, *5*, 1–12. doi:10.3389/fpsyg.2014.00405.
Edwards, M. C., & Wirth, R. J. (2009). Measurement and the study of change. *Research in Human Development*, *6*, 74–96. doi:10.1080/15427600902911163.
Eisenberg, N., Edwards, A., Spinrad, T. L., Sallquist, J., Eggum, N. D., & Reiser, M. (2013). Are effortful and reactive control unique constructs in young children? *Developmental Psychology*, *49*, 2082–2094. doi:10.1037/a0031745.
Eisenberg, N., Vidmar, M., Spinrad, T. L., Eggum, N. D., Edwards, A., Gaertner, B., et al. (2010). Mothers' teaching strategies and children's effortful control: A longitudinal study. *Developmental Psychology*, *46*, 1294–1308. doi:10.1037/a0020236.
Esposito, A. G., Baker-Ward, L., & Mueller, S. T. (2013). Interference suppression vs. response inhibition: An explanation for the absence of a bilingual advantage in preschoolers' Stroop task performance. *Cognitive Development*, *28*, 354–363. doi:10.1016/j.cogdev.2013.09.002.
Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental Psychology*, *49*, 1958–1963. doi:10.1037/a0031409.
Evans, A. D., Xu, F., & Lee, K. (2011). When all signs point to you: Lies told in the face of evidence. *Developmental Psychology*, *47*, 39–49. doi:10.1037/a0020787.
Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Journal of Applied Developmental Psychology*, *24*, 51–73. doi:10.1016/S0193-3973(03)00024-8.
Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. Methodology. *European Journal of Research Methods for the Behavioral and Social Sciences*, *4*, 22–36. doi:10.1027/1614-2241.4.1.22.
Fizke, E., Barthel, D., Peters, T., & Rakoczy, H. (2014). Executive function plays a role in coordinating different perspectives, particularly when one's own perspective is involved. *Cognition*, *130*, 315–334. doi:10.1016/j.cognition.2013.11.017.
Floyd, R. G., & Kirby, E. A. (2001). Psychometric properties of measures of behavioral inhibition with preschool-age children: Implications for assessment of children at risk for ADHD. *Journal of Attention Disorders*, *5*, 79–91. doi:10.1177/108705470100500202.
Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development*, *16*, 53–69. doi:10.1002/icd.500.
Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, *7*, 103–115. doi:10.1111/j.1467-7687.2004.00326.x.
Fontaine, N. M. G., & Petersen, I. T. (2016). Developmental trajectories of psychopathology: An overview of approaches and applications. In L. Centifanti & D. Williams (Eds.), *International handbook of developmental psychopathology*. Wiley-Blackwell. In press.
Ford, R. M., Driscoll, T., Shum, D., & Macaulay, C. E. (2012). Executive and theory-of-mind contributions to event-based prospective memory in children: Exploring the self-projection hypothesis. *Journal of Experimental Child Psychology*, *111*, 468–489. doi:10.1016/j.jecp.2011.10.006.
Fox, N. A., Hane, A. A., & Pérez-Edgar, K. (2006). Psychophysiological methods for the study of developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology* (2nd ed., Vol. 2, pp. 381–426). Hoboken, NJ, USA: John Wiley & Sons Inc.
Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology*, *47*, 404–416. doi:10.1037/a0021065.
Gaillard, W. D., Grandin, C. B., & Xu, B. (2001). Developmental aspects of pediatric fMRI: Considerations for image acquisition, analysis, and interpretation. *Neuroimage*, *13*, 239–249. doi:10.1006/nimg.2000.0681.
Gandolfi, E., Viterbori, P., Traverso, L., & Usai, M. C. (2014). Inhibitory processes in toddlers: A latent-variable approach. *Frontiers in Psychology*, *5*, 1–11. doi:10.3389/fpsyg.2014.00381.
Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60. doi:10.1037/0033-2909.134.1.31.
Geng, F., Hu, Y., Wang, Y., & Chen, F. (2011). Two types of behavioral inhibition: Relations to effortful control and attention in school children. *Journal of Research in Personality*, *45*, 662–669. doi:10.1016/j.jrp.2011.09.005.
Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24–36 months of age. *Developmental Science*, *3*, 397–404. doi:10.1111/1467-7687.00134.
Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2–7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153. doi:10.1016/0010-0277(94)90068-x.
Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review: Behavior rating inventory of executive function. *Child Neuropsychology*, *6*, 235–238. doi:10.1076/chin.6.3.235.3152.

Gleason, M. M., Fox, N. A., Drury, S., Smyke, A., Egger, H. L., Nelson, C. A., III, et al. (2011). Validity of evidence-derived criteria for reactive attachment disorder: Indiscriminately social/disinhibited and emotionally withdrawn/inhibited types. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*, 216–231, e213. doi:10.1016/j.jaac.2010.12.012.
González, C., Fuentes, L. J., Carranza, J. A., & Estévez, A. F. (2001). Temperament and attention in the self-regulation of 7-year-old children. *Personality and Individual Differences*, *30*, 931–946. doi:10.1016/S0191-8869(00)00084-2.
Grimm, K. J., Pianta, R. C., & Konold, T. (2009). Longitudinal multitrait-multimethod models for developmental research. *Multivariate Behavioral Research*, *44*, 233–258. doi:10.1080/00273170902794230.
Gusdorf, L. M. A., Karreman, A., van Aken, M. A. G., Deković, M., & van Tuijl, C. (2011). The structure of effortful control in preschoolers and its relation to externalizing problems. *The British Journal of Developmental Psychology*, *29*, 612–634. doi:10.1348/026151010x526542.
Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Journal of Cognition and Development*, *4*, 275–298.
Hammond, S. I., Müller, U., Carpendale, J. I. M., Bibok, M. B., & Liebermann-Finestone, D. P. (2012). The effects of parental scaffolding on preschoolers' executive function. *Developmental Psychology*, *48*, 271–281. doi:10.1037/a0025519.
Hardaway, C. R., Wilson, M. N., Shaw, D. S., & Dishion, T. J. (2012). Family functioning and externalizing behaviour among low-income children: Self-regulation as a mediator. *Infant and Child Development*, *21*, 67–84. doi:10.1002/icd.765.
Hayden, E. P., Klein, D. N., Sheikh, H. I., Olino, T. M., Dougherty, L. R., Dyson, M. W., et al. (2010). The serotonin transporter promoter polymorphism and childhood positive and negative emotionality. *Emotion (Washington, D.C.)*, *10*, 696–702. doi:10.1037/a0019527.
Herbers, J. E., Cutuli, J. J., Lafavor, T. L., Vrieze, D., Leibel, C., Obradović, J., et al. (2011). Direct and indirect effects of parenting on the academic functioning of young homeless children. *Early Education and Development*, *22*, 77–104. doi:10.1080/10409280903507261.
Herbers, J. E., Cutuli, J. J., Monn, A. R., Narayan, A. J., & Masten, A. S. (2014). Trauma, adversity, and parent–child relationships among young children experiencing homelessness. *Journal of Abnormal Child Psychology*, *42*, 1167–1174. doi:10.1007/s10802-014-9868-7.
Herbers, J. E., Cutuli, J. J., Supkoff, L. M., Narayan, A. J., & Masten, A. S. (2014). Parenting and coregulation: Adaptive systems for competence in children experiencing homelessness. *American Journal of Orthopsychiatry*, *84*, 420–430. doi:10.1037/h0099843.
Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, *18*, 639–657. doi:10.1037/0882-7974.18.4.639.
Hinnant, J. B., & O'Brien, M. (2007). Cognitive and emotional control and perspective taking and their relations to empathy in 5-year-old children. *The Journal of Genetic Psychology*, *168*, 301–322. doi:10.3200/GNTP.168.3.301-322.
Hoff Esbjørn, B., Hansen, B. M., Greisen, G., & Mortensen, E. L. (2006). Intellectual development in a Danish cohort of prematurely born preschool children: Specific or general difficulties? *Journal of Developmental and Behavioral Pediatrics: JDBP*, *27*, 477–484.
Holmboe, K., Pasco Fearon, R. M., Csibra, G., Tucker, L. A., & Johnson, M. H. (2008). Freeze-frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology*, *100*, 89–114. doi:10.1016/j.jecp.2007.09.004.
Hoyniak, C. P. (2016). *Development of the neurophysiological correlates of response inhibition from early childhood to early adolescence: A systematic review and meta-analysis of the N2 component*. Under review.
Hrabok, M., Kerns, K. A., & Müller, U. (2007). The vigilance, orienting, and executive attention networks in 4-year-old children. *Child Neuropsychology*, *13*, 408–421. doi:10.1080/13825580600977224.
Hughes, C. H. (1996). Control of action and thought: Normal development and dysfunction in autism: A research note. *Journal of Child Psychology and Psychiatry*, *37*, 229–236. doi:10.1111/j.1469-7610.1996.tb01396.x.
Hughes, C. H. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *The British Journal of Developmental Psychology*, *16*, 233–253. doi:10.1111/j.2044-835X.1998.tb00921.x.
Hughes, C. H. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology*, *34*, 1326–1339. doi:10.1037/0012-1649.34.6.1326.
Hughes, C. H., Dunn, J., & White, A. (1998). Trick or treat?: Uneven understanding of mind and emotion and executive dysfunction in 'hard-to-manage' preschoolers. *Journal of Child Psychology and Psychiatry*, *39*, 981. doi:10.1111/1469-7610.00401.
Hughes, C. H., & Ensor, R. A. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology*, *28*, 645–668. doi:10.1207/s15326942dn2802_5.
Hughes, C. H., & Ensor, R. A. (2006). Behavioural problems in 2-year-olds: Links with individual differences in theory of mind, executive function and harsh parenting. *Journal of Child Psychology & Psychiatry*, *47*, 488–497. doi:10.1111/j.1469-7610.2005.01519.x.
Hughes, C. H., & Ensor, R. A. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*, 1447–1459. doi:10.1037/0012-1649.43.6.1447.
Hughes, C. H., & Ensor, R. A. (2008). Does executive function matter for preschoolers' problem behaviors? *Journal of Abnormal Child Psychology*, *36*, 1–14. doi:10.1007/s10802-007-9107-6.
Hughes, C. H., & Ensor, R. A. (2009a). How do families help or hinder the emergence of early executive function? *New Directions for Child and Adolescent Development*, *2009*, 35–50. doi:10.1002/cd.234.
Hughes, C. H., & Ensor, R. A. (2009b). Independence and interplay between maternal and child risk factors for preschool problem behaviors? *International Journal of Behavioral Development*, *33*, 312–322. doi:10.1177/0165025408101274.
Hughes, C. H., Ensor, R. A., & Marks, A. (2011). Individual differences in false belief understanding are stable from 3 to 6 years of age and predict children's mental state talk with school friends. *Journal of Experimental Child Psychology*, *108*, 96–112. doi:10.1016/j.jecp.2010.07.012.
Hughes, C. H., Ensor, R. A., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, *35*, 20–36. doi:10.1080/87565640903325691.
Hughes, C. H., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. *Developmental Psychology*, *29*, 498–510. doi:10.1037/0012-1649.29.3.498.
Huyder, V. (2010). *The role of executive functioning and perspective-taking in facilitating children's socially competent behaviours* (Master's thesis). Waterloo, Canada: University of Waterloo. <http://hdl.handle.net/10012/5423>.

Huyder, V., & Nilsen, E. S. (2012). A dyadic data analysis of executive functioning and children's socially competent behaviours. *Journal of Applied Developmental Psychology*, *33*, 197–208. doi:10.1016/j.appdev.2012.05.002.
Ito, J. M. B. (2009). *The phenomenology of role play: Are children "thinking-as-if" or "behaving as if"?* (Doctoral thesis). Kingston, Canada: Queen's University. <http://hdl.handle.net/1974/1744>.
Jahromi, L. B. (2004). *Individual differences in inhibitory control: Relations across domains and predictions to theory of mind* (Doctoral dissertation). State College, PA: Penn State University. <https://etda.libraries.psu.edu/paper/6266/>.
Jennings, K. D., Sandberg, I., Kelley, S. A., Valdes, L., Yaggi, K., Abrew, A., et al. (2008). Understanding of self and maternal warmth predict later self-regulation in toddlers. *International Journal of Behavioral Development*, *32*, 108–118. doi:10.1177/0165025407087209.
Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science*, *6*, 498–504. doi:10.1111/1467-7687.00307.
Joseph, R. M., McGrath, L. M., & Tager-Flusberg, H. (2005). Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Developmental Neuropsychology*, *27*, 361–378. doi:10.1207/s15326942dn2703_4.
Joseph, R. M., & Tager-Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology*, *16*, 137–155. doi:10.1017/S095457940404444X.
Karpinski, A. C., Scullin, M. H., & Montgomery-Downs, H. E. (2008). Risk for sleep-disordered breathing and executive function in preschoolers. *Sleep Medicine*, *9*, 418–424. doi:10.1016/j.sleep.2007.06.004.
Kegel, C. A. T. (2011). *Differential susceptibility in education: Interaction between genes, regulatory skills, and computer games* (Doctoral thesis). Leiden, Netherlands: Universiteit Leiden. <https://openaccess.leidenuniv.nl/handle/1887/17974>.
Kegel, C. A. T., van der Kooy-Hofland, V. A. C., & Bus, A. G. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences*, *19*, 549–554. doi:10.1016/j.lindif.2009.07.002.
Kesek, A. (2010). *Direct and indirect approaches to emotion regulation in children* (Doctoral dissertation). Minneapolis, MN: University of Minnesota. <http://purl.umn.edu/98741>.
Kim, S., & Kochanska, G. (2012). Child temperament moderates effects of parent–child mutuality on self-regulation: A relationship-based path for emotionally negative infants. *Child Development*, *83*, 1275–1289. doi:10.1111/j.1467-8624.2012.01778.x.
Kim, S., Nordling, J., Yoon, J., Boldt, L., & Kochanska, G. (2013). Effortful control in "hot" and "cool" tasks differentially predicts children's behavior problems and academic performance. *Journal of Abnormal Child Psychology*, *41*, 43–56. doi:10.1007/s10802-012-9661-4.
King, A. C. (2011). Development of inhibition as a function of the presence of a supernatural agent. *The Journal of Genetic Psychology*, *172*, 414–432. doi:10.1080/00221325.2011.554921.
King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, *30*, 666–687.
Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology*, *20*, 407–428. doi:10.1207/S15326942DN2001_6.
Kloo, D., Perner, J., & Giritzer, T. (2010). Object-based set-shifting in preschoolers: Relations to theory of mind. In B. W. Sokol, U. Muller, J. Carpendale, A. R. Young, & G. Iarocci (Eds.), *Self- and social-regulation: Exploring the relations between social interaction, social cognition, and the development of executive functions* (pp. 193–217). Oxford, England: Oxford University Press.
Kloo, D., Perner, J., Kerschhuber, A., Dabernig, S., & Aichhorn, M. (2008). Sorting between dimensions: Conditions of cognitive flexibility in preschoolers. *Journal of Experimental Child Psychology*, *100*, 115–134. doi:10.1016/j.jecp.2007.12.003.
Knight, G. P., & Zerr, A. A. (2010). Informed theory and measurement equivalence in child development research. *Child Development Perspectives*, *4*, 25–30. doi:10.1111/j.1750-8606.2009.00112.x.
Kochanska, G., Aksan, N., Penney, S. J., & Doobay, A. F. (2007). Early positive emotionality as a heterogenous trait: Implications for children's self-regulation. *Journal of Personality and Social Psychology*, *93*, 1054–1066. doi:10.1037/0022-3514.93.6.1054.
Kochanska, G., Aksan, N., Prisco, T. R., & Adams, E. E. (2008). Mother–child and father–child mutually responsive orientation in the first 2 years and children's outcomes at preschool age: Mechanisms of influence. *Child Development*, *79*, 30–44. doi:10.1111/j.1467-8624.2007.01109.x.
Kochanska, G., Barry, R. A., Jimenez, N. B., Hollatz, A. L., & Woodard, J. (2009). Guilt and effortful control: Two mechanisms that prevent disruptive developmental trajectories. *Journal of Personality and Social Psychology*, *97*, 322–333. doi:10.1037/a0015471.
Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development*, *72*, 1091–1111. doi:10.1111/1467-8624.00336.
Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality*, *71*, 1087–1112. doi:10.1111/1467-6494.7106008.
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*, 263–277. doi:10.1111/j.1467-8624.1997.tb01939.x.
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*, 490–507. doi:10.2307/1131828.
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*, 220–232. doi:10.1037/0012-1649.36.2.220.
Kolen, M. J., & Brennan, R. L. (2014). *Test equating, linking, and scaling: Methods and practices* (3rd ed.). New York, NY, USA: Springer.
Kolnik, S. (2010). *Predicting school readiness: Executive functions, problem behaviors and theory of mind in preschoolers* (Doctoral dissertation). Coral Gables, FL: University of Miami. <http://scholarlyrepository.miami.edu/oa_dissertations/416/>.
Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology*, *18*, 199–214. doi:10.1037/0012-1649.18.2.199.
Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY: A developmental neuropsychological assessment*. Psychological Corporation.
Korkman, M., Liikanen, A., & Fellman, V. (1996). Neuropsychological consequences of very low birth weight and asphyxia at term: Follow-up until school-age. *Journal of Clinical and Experimental Neuropsychology*, *18*, 220–233.
Kraybill, J. H. (2013). *A latent factor analysis of preschool executive functions: Investigations of antecedents and outcomes* (Doctoral dissertation). Blacksburg, VA: Virginia Polytechnic Institute and State University. <http://hdl.handle.net/10919/25334>.

Kraybill, J. H., & Bell, M. A. (2013). Infancy predictors of preschool and post-kindergarten executive function. *Developmental Psychobiology*, *55*, 530–538. doi:10.1002/dev.21057.
Krueger, R. F., Nichol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., Iacono, W. G., et al. (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment*, *16*, 107–119. doi:10.1037/1040-3590.16.2.107.
Kryski, K. R., Dougherty, L. R., Dyson, M. W., Olino, T. M., Laptook, R. S., Klein, D. N., et al. (2013). Effortful control and parenting: Associations with HPA axis reactivity in early childhood. *Developmental Science*, *16*, 531–541. doi:10.1111/desc.12050.
Lackner, C. (2009). *Preliminary investigations of dopaminergic contributions to preschoolers' theory of mind development* (Master's thesis). Kingston, Canada: Queen's University. <http://hdl.handle.net/1974/5143>.
Lackner, C. L., Bowman, L. C., & Sabbagh, M. A. (2010). Dopaminergic functioning and preschoolers' theory of mind. *Neuropsychologia*, *48*, 1767–1774. doi:10.1016/j.neuropsychologia.2010.02.027.
Lackner, C., Sabbagh, M. A., Hallinan, E., Liu, X., & Holden, J. J. A. (2012). Dopamine receptor D4 gene variation predicts preschoolers' developing theory of mind. *Developmental Science*, *15*, 272–280. doi:10.1111/j.1467-7687.2011.01124.x.
Lagattuta, K. H., Sayfan, L., & Monsour, M. (2011). A new measure for assessing executive function across a wide age range: Children and adults find happy-sad more difficult than day-night. *Developmental Science*, *14*, 481–489. doi:10.1111/j.1467-7687.2010.00994.x.
Lahat, A., Degnan, K. A., White, L. K., McDermott, J. M., Henderson, H. A., Lejuez, C. W., et al. (2012). Temperamental exuberance and executive function predict propensity for risk taking in childhood. *Development and Psychopathology*, *24*, 847–856. doi:10.1017/S0954579412000405.
Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J., & Kerr, D. C. R. (2010). Theory of mind and emotion understanding predict moral development in early childhood. *The British Journal of Developmental Psychology*, *28*, 871–889. doi:10.1348/026151009X483056.
Lane, J. D., Wellman, H. M., Olson, S. L., Miller, A. L., Wang, L., & Tardif, T. (2013). Relations between temperament and theory of mind development in the United States and China: Biological and behavioral correlates of preschoolers' false-belief understanding. *Developmental Psychology*, *49*, 825–836. doi:10.1037/a0028825.
Lang, B., & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *The British Journal of Developmental Psychology*, *20*, 67–76. doi:10.1348/026151002166325.
LaVoie, J. C., Anderson, K., Fraze, B., & Johnson, K. (1981). Modeling, tuition, and sanction effects on self-control at different ages. *Journal of Experimental Child Psychology*, *31*, 446–455. doi:10.1016/0022-0965(81)90029-1.
Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, *84*, 1933–1953. doi:10.1111/cdev.12096.
Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *The British Journal of Educational Psychology*, *82*, 82–99. doi:10.1111/j.2044-8279.2010.02016.x.
Leerkes, E. M., Paradise, M., O'Brien, M., Calkins, S. D., & Lange, G. (2008). Emotion and cognition processes in preschool children. *Merrill-Palmer Quarterly*, *54*, 102–124. doi:10.1353/mpq.2008.0009.
Lengua, L. J. (2002). The contribution of emotionality and self-regulation to the understanding of children's response to multiple risk. *Child Development*, *73*, 144–161. doi:10.1111/1467-8624.00397.
Lengua, L. J. (2003). Associations among emotionality, self-regulation, adjustment problems, and positive adjustment in middle childhood. *Journal of Applied Developmental Psychology*, *24*, 595–618. doi:10.1016/j.appdev.2003.08.002.
Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of Applied Developmental Psychology*, *28*, 40–55. doi:10.1016/j.appdev.2006.10.001.
Lengua, L. J., Kiff, C., Moran, L., Zalewski, M., Thompson, S., Cortes, R., et al. (2014). Parenting mediates the effects of income and cumulative risk on the development of effortful control. *Social Development (Oxford, England)*, *23*, 631–649. doi:10.1111/sode.12071.
Lengua, L. J., & Kovacs, E. A. (2005). Bidirectional associations between temperament and parenting and the prediction of adjustment problems in middle childhood. *Journal of Applied Developmental Psychology*, *26*, 21–38. doi:10.1016/j.appdev.2004.10.001.
Lengua, L. J., Moran, L., Zalewski, M., Ruberry, E., Kiff, C., & Thompson, S. (2015). Relations of growth in effortful control to family income, cumulative risk, and adjustment in preschool-age children. *Journal of Abnormal Child Psychology*, *43*, 705–720. doi:10.1007/s10802-014-9941-2.
Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment*, *36*, 626–639. doi:10.1007/s10862-014-9424-3.
Leve, L. D., DeGarmo, D. S., Bridgett, D. J., Neiderhiser, J. M., Shaw, D. S., Harold, G. T., et al. (2013). Using an adoption design to separate genetic, prenatal, and temperament influences on toddler executive function. *Developmental Psychology*, *49*, 1045–1057. doi:10.1037/a0029390.
Lewis, E. E., Dozier, M., Ackerman, J., & Sepulveda-Kozakowski, S. (2007). The effect of placement instability on adopted children's inhibitory control abilities and oppositional behavior. *Developmental Psychology*, *43*, 1415–1427. doi:10.1037/0012-1649.43.6.1415.
Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY, USA: The Guilford Press.
Little, T. D., Card, N. A., Preacher, K. J., & McConnell, E. (2009). Modeling longitudinal data from research on adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Individual bases of adolescent development*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling : A Multidisciplinary Journal*, *9*, 151–173. doi:10.1207/S15328007SEM0902_1.
Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, *31*, 357–365. doi:10.1177/0165025407077757.
Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling : A Multidisciplinary Journal*, *13*, 59–72. doi:10.1207/s15328007sem1301_3.

Livesey, D., Keen, J., Rouse, J., & White, F. (2006). The relationship between measures of executive function, motor performance and externalising behaviour in 5- and 6-year-old children. *Human Movement Science*, *25*, 50–64. doi:10.1016/j.humov.2005.10.008.

Lunkenheimer, E. S., Albrecht, E. C., & Kemp, C. J. (2013). Dyadic flexibility in early parent-child interactions: Relations with maternal depressive symptoms and child negativity and behaviour problems. *Infant and Child Development*, *22*, 250–269. doi:10.1002/icd.1783.

Lunkenheimer, E. S., Dishion, T. J., Shaw, D. S., Connell, A. M., Gardner, F., Wilson, M. N., et al. (2008). Collateral benefits of the family check-up on early childhood school readiness: Indirect effects of parents' positive behavior support. *Developmental Psychology*, *44*, 1737–1752. doi:10.1037/a0013858.

Luria, A. R. (1966). *Higher cortical functions in man*. Oxford, England: Basic Books.

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

Macaulay, C. E., & Ford, R. M. (2013). Family influences on the cognitive development of profoundly deaf children: Exploring the effects of socioeconomic status and siblings. *Journal of Deaf Studies and Deaf Education*, *18*, 545–562. doi:10.1093/deafed/ent019.

Marlow, N., Hennessy, E. M., Bracewell, M. A., Wolke, D., & The EPICure Study Group (2007). Motor and executive function at 6 years of age after extremely preterm birth. *Pediatrics*, *120*, 793–804. doi:10.1542/peds.2007-0440.

Marshall, P. J., & Drew, A. R. (2014). What makes Simon Says so difficult for young children? *Journal of Experimental Child Psychology*, *126*, 112–119. doi:10.1016/j.jecp.2014.03.011.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. Bilingualism. *Language and Cognition*, *11*, 81–93. doi:10.1017/S1366728907003227.

Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K., et al. (2012). Executive function skills and school success in young children experiencing homelessness. *Educational Researcher (Washington, D.C. : 1972)*, *41*, 375–384. doi:10.3102/0013189x12459883.

Matte-Gagné, C., & Bernier, A. (2011). Prospective relations between maternal autonomy support and child executive functioning: Investigating the mediating role of child language ability. *Journal of Experimental Child Psychology*, *110*, 611–625. doi:10.1016/j.jecp.2011.06.006.

Mayer, S. E., Abelson, J. L., & Lopez-Duran, N. L. (2014). Effortful control and context interact in shaping neuroendocrine stress responses during childhood. *Hormones and Behavior*, *66*, 457–465. doi:10.1016/j.yhbeh.2014.06.019.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. doi:10.1146/annurev.psych.60.110707.163612.

McArdle, J. J., & Grimm, K. J. (2011). An empirical example of change analysis by linking longitudinal item response data from multiple tests. In A. A. Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 71–88). Springer New York.

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*, 126–149. doi:10.1037/a0015857.

McArdle, J. J., Hamagami, F., Meredith, W., & Bradway, K. P. (2000). Modeling the dynamic hypotheses of Gf-Gc theory using longitudinal life-span data. *Learning and Individual Differences*, *12*, 53. doi:10.1016/S1041-6080(00)00036-4.

McClelland, M. M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, *6*, 136–142. doi:10.1111/j.1750-8606.2011.00191.x.

McClelland, M. M., Cameron, C. E., Duncan, R. J., Bowles, R. P., Acock, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: The Head-Toes-Knees-Shoulders task. *Frontiers in Psychology*, *5*, 1–14.

Meade, A. W., Lautenschlager, G. J., & Hecht, J. E. (2005). Establishing measurement equivalence and invariance in longitudinal data with item response theory. *International Journal of Testing*, *5*, 279–300. doi:10.1207/s15327574ijt0503_6.

Melinder, A., Endestad, T. O. R., & Magnussen, S. (2006). Relations between episodic memory, suggestibility, theory of mind, and cognitive inhibition in the preschool child. *Scandinavian Journal of Psychology*, *47*, 485–495. doi:10.1111/j.1467-9450.2006.00542.x.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. doi:10.1007/bf02294825.

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*, 395–423. doi:10.1080/15248372.2011.585478.

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

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A Latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi:10.1006/cogp.1999.0734.

Moffitt, T. E., Arseneault, L., Belsky, D. W., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 2693–2698. doi:10.1073/pnas.1010076108.

Molfese, V. J., Molfese, P. J., Molfese, D. L., Rudasill, K. M., Armstrong, N., & Starkey, G. (2010). Executive function skills of 6–8 year olds: Brain and behavioral evidence and implications for school achievement. *Contemporary Educational Psychology*, *35*, 116–125. doi:10.1016/j.cedpsych.2010.03.004.

Monette, S. (2012). *Fonctions exécutives chez les enfants d'âge préscolaire: Lien avec la réussite scolaire ultérieure et association avec les comportements de type externalizes* (Doctoral thesis). Montréal, Canada: Université du Québec à Montréal. <http://www.archipel.uqam.ca/5080>.

Monette, S., Bigras, M., & Guay, M.-C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology*, *109*, 158–173. doi:10.1016/j.jecp.2011.01.008.

Monks, C. P., Smith, P. K., & Swettenham, J. (2005). Psychological correlates of peer victimisation in preschool: Social cognitive skills, executive function and attachment profiles. *Aggressive Behavior*, *31*, 571–588. doi:10.1002/ab.20099.

Montgomery, D. E., Anderson, M., & Uhl, E. (2008). Interference control in preschoolers: Factors influencing performance on the day–night task. *Infant and Child Development*, *17*, 457–470. doi:10.1002/icd.559.
Moran, L. R., Lengua, L. J., & Zalewski, M. (2013). The interaction between negative emotionality and effortful control in early social-emotional development. *Social Development (Oxford, England)*, *22*, 340–362. doi:10.1111/sode.12025.
Murray, K. T., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology*, *30*, 503–514. doi:10.1023/a:1019821031523.
Muthén, L. K., & Muthén, B. O. (2010). *Mplus short courses. Topic 4: Growth modeling with latent variables*. Handout. <http://www.statmodel.com/download/Topic4-v.pdf>.
Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I. M., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of Experimental Child Psychology*, *111*, 331–348. doi:10.1016/j.jecp.2011.08.014.
Nampijja, M., Apule, B., Lule, S., Akurut, H., Muhangi, L., Elliott, A. M., et al. (2010). Adaptation of Western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *The British Journal of Educational Psychology*, *80*, 15–30. doi:10.1348/000709909X460600.
Nesselroade, J. R., & Estabrook, R. (2009). Factor invariance, measurement, and studying development over the lifespan. In H. B. Bosworth & C. Hertzog (Eds.), *Aging and cognition: Research methodologies and empirical advances* (pp. 39–52). Washington, DC, USA: American Psychological Association.
Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development*, *83*, 1400–1415. doi:10.1111/j.1467-8624.2012.01762.x.
Obradović, J. (2010). Effortful control and adaptive functioning of homeless children: Variable-focused and person-focused analyses. *Journal of Applied Developmental Psychology*, *31*, 109–117. doi:10.1016/j.appdev.2009.09.004.
Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., et al. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, *20*, 673–716. doi:10.1017/S0954579408000333.
Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development*, *79*, 80–99. doi:10.1111/j.1467-8624.2007.01112.x.
Olson, S. L., Lopez-Duran, N., Lunkenheimer, E. S., Chang, H., & Sameroff, A. J. (2011). Individual differences in the development of early peer aggression: Integrating contributions of self-regulation, theory of mind, and parenting. *Development and Psychopathology*, *23*, 253–266. doi:10.1017/S0954579410000775.
Olson, S. L., Sameroff, A. J., Kerr, D. C. R., Lopez, N. L., & Wellman, H. M. (2005). Developmental foundations of externalizing problems in young children: The role of effortful control. *Development and Psychopathology*, *17*, 25–45. doi:10.1017/s0954579405050029.
Olson, S. L., Tardif, T. Z., Miller, A., Felt, B., Grabell, A. S., Kessler, D., et al. (2011). Inhibitory control and harsh discipline as predictors of externalizing problems in young children: A comparative study of U.S., Chinese, and Japanese preschoolers. *Journal of Abnormal Child Psychology*, *39*, 1163–1175. doi:10.1007/s10802-011-9531-5.
Orta, I. M., Corapci, F., Yagmurlu, B., & Aksan, N. (2013). The mediational role of effortful control and emotional dysregulation in the link between maternal responsiveness and Turkish preschoolers' social competency and externalizing symptoms. *Infant and Child Development*, *22*, 459–479. doi:10.1002/icd.1806.
Owens, E. B., & Shaw, D. S. (2003). Predicting growth curves of externalizing behavior across the preschool years. *Journal of Abnormal Child Psychology*, *31*, 575–590. doi:10.1023/a:1026254005632.
Patterson, G. R. (1993). Orderly change in a stable world: The antisocial trait as a chimera. *Journal of Consulting and Clinical Psychology*, *61*, 911–919. doi:10.1037/0022-006X.61.6.911.
Pauli-Pott, U., Dalir, S., Mingebach, T., Roller, A., & Becker, K. (2013). Do different ADHD-related etiological risks involve specific neuropsychological pathways? An analysis of mediation processes by inhibitory control and delay aversion. *Journal of Child Psychology and Psychiatry*, *54*, 800–809. doi:10.1111/jcpp.12059.
Pauli-Pott, U., Roller, A., Heinzel-Gutenbrunner, M., Mingebach, T., Dalir, S., & Becker, K. (2014). Inhibitory control and delay aversion in unaffected preschoolers with a positive family history of attention deficit hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, *55*, 1117–1124. doi:10.1111/jcpp.12230.
Pears, K. C., & Fisher, P. A. (2005). Emotion understanding and theory of mind among maltreated children in foster care: Evidence of deficits. *Development and Psychopathology*, *17*, 47–65. doi:10.1017/S0954579405050030.
Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology*, *43*, 974–990. doi:10.1037/0012-1649.43.4.974.
Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental Psychology*, *46*, 530–544. doi:10.1037/a0018287.
Pellicano, E. (2013). Testing the predictive power of cognitive atypicalities in autistic children: Evidence from a 3-year follow-up study. *Autism Research*, *6*, 258–267. doi:10.1002/aur.1286.
Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: "Weak" central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology*, *18*, 77–98. doi:10.1017/S0954579406060056.
Perner, J., Kain, W., & Barchfeld, P. (2002). Executive control and higher-order theory of mind in children at risk of ADHD. *Infant and Child Development*, *11*, 141–158. doi:10.1002/icd.302.
Peskin, J., & Ardino, V. (2003). Representing the mental world in children's social behavior: Playing hide-and-seek and keeping a secret. *Social Development (Oxford, England)*, *12*, 496–512. doi:10.1111/1467-9507.00245.
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*, 791–818. doi:10.1017/S0954579414000789.
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*, 221–237. doi:10.1017/S0954579414000698.

Pettit, G. S., Keiley, M. K., Laird, R. D., Bates, J. E., & Dodge, K. A. (2007). Predicting the developmental course of mother-reported monitoring across childhood and adolescence from early proactive parenting, child temperament, and parents' worries. *Journal of Family Psychology*, *21*, 206–217. doi:10.1037/0893-3200.21.2.206.

Phillips, J. P., Ruhl, D., Montague, E., Gasparovic, C., Caprihan, A., Ohls, R. K., et al. (2011). Anterior cingulate and frontal lobe white matter spectroscopy in early childhood of former very LBW premature infants. *Pediatric Research*, *69*, 224–229. doi:10.1203/PDR.0b013e3182091d52.

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, *108*, 567–579. doi:10.1016/j.jecp.2010.10.009.

Prado, E. L., Hartini, S., Rahmawati, A., Ismayani, E., Hidayati, A., Hikmah, N., et al. (2010). Test selection, adaptation, and evaluation: A systematic approach to assess nutritional influences on child development in developing countries. *The British Journal of Educational Psychology*, *80*, 31–53. doi:10.1348/000709909X470483.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, *29*, 386–401. doi:10.1016/j.infbeh.2006.01.004.

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*, 102–112. doi: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*, 387–405. doi:10.1002/ICD.582.

Qu, L. (2011). Two is better than one, but mine is better than ours: Preschoolers' executive function during co-play. *Journal of Experimental Child Psychology*, *108*, 549–566. doi:10.1016/j.jecp.2010.08.010.

Qu, L., Audrey, L. S., Jun, L. P., & Qun, N. H. (2010). *The impact of social context on preschoolers' flexibility* (pp. 147–153). International Association for Cross-Cultural Psychology.

Raaijmakers, M. A. J., Smidts, D. P., Sergeant, J. A., Maassen, G. H., Posthumus, J. A., van Engeland, H., et al. (2008). Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. *Journal of Abnormal Child Psychology*, *36*, 1097–1107. doi:10.1007/s10802-008-9235-7.

Raffaelli, M., Crockett, L. J., & Shen, Y. L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *The Journal of Genetic Psychology*, *166*, 54–76. doi:10.3200/GNTP.166.1.54-76.

Rakoczy, H. (2010). Executive function and the development of belief–desire psychology. *Developmental Science*, *13*, 648–661. doi:10.1111/j.1467-7687.2009.00922.x.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism. *Child Development*, *82*, 362–378. doi:10.1111/j.1467-8624.2010.01561.x.

Reck, S. G., & Hund, A. M. (2011). Sustained attention and age predict inhibitory control during early childhood. *Journal of Experimental Child Psychology*, *108*, 504–512. doi:10.1016/j.jecp.2010.07.010.

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly*, *30*, 131–147.

Revelle, W., & Zinbarg, R. (2009). Coefficients alpha, beta, omega, and the glb: Comments on sijtsma. *Psychometrika*, *74*, 145–154. doi:10.1007/s11336-008-9102-z.

Rhoades, B. L., Greenberg, M. T., & Domitrovich, C. E. (2009). The contribution of inhibitory control to preschoolers' social-emotional competence. *Journal of Applied Developmental Psychology*, *30*, 310–320. doi:10.1016/j.appdev.2008.12.012.

Rhoades, B. L., Greenberg, M. T., Lanza, S. T., & Blair, C. (2011). Demographic and familial predictors of early executive function development: Contribution of a person-centered perspective. *Journal of Experimental Child Psychology*, *108*, 638–662. doi:10.1016/j.jecp.2010.08.004.

Riggs, K. J., Jolley, R. P., & Simpson, A. (2013). The role of inhibitory control in the development of human figure drawing in young children. *Journal of Experimental Child Psychology*, *114*, 537–542. doi:10.1016/j.jecp.2012.10.003.

Roberts, K. P., & Powell, M. B. (2005). The relation between inhibitory control and children's eyewitness memory. *Applied Cognitive Psychology*, *19*, 1003–1018. doi:10.1002/acp.1141.

Rochette, É., & Bernier, A. (2014). Parenting, family socioeconomic status, and child executive functioning: A longitudinal study. *Merrill-Palmer Quarterly*, *60*, 431–460.

Roebers, C. M., & Schneider, W. (2005). Individual differences in young children's suggestibility: Relations to event memory, language abilities, working memory, and executive functioning. *Cognitive Development*, *20*, 427–447. doi:10.1016/j.cogdev.2005.05.006.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, *52*, 569–578. doi:10.1111/1467-8624.ep8860473.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly*, *40*, 21–39.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*, 1394–1408. doi:10.1111/1467-8624.00355.

Rothbart, M. K., Ellis, L. K., Rueda, M. R., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of Personality*, *71*, 1113–1144. doi:10.1111/1467-6494.7106009.

Röthlisberger, M., Neuenschwander, R., Cimeli, P., Michel, E., & Roebers, C. M. (2012). Improving executive functions in 5- and 6-year-olds: Evaluation of a small group intervention in prekindergarten and kindergarten children. *Infant and Child Development*, *21*, 411–429. doi:10.1002/icd.752.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology*, *28*, 573–594. doi:10.1111/1467-8624.00392.

Rutter, M., & Sroufe, L. A. (2000). Developmental psychopathology: Concepts and challenges. *Development and Psychopathology*, *12*, 265–296. doi:10.1017/S0954579400003023.

Ryan, A. S., & Nelson, E. B. (2008). Assessing the effect of docosahexaenoic acid on cognitive functions in healthy, preschool children: A randomized, placebo-controlled, double-blind study. *Clinical Pediatrics*, *47*, 355–362. doi:10.1177/0009922807311730.

Sabbagh, M. A., Bowman, L. C., Evraire, L. E., & Ito, J. M. B. (2009). Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*, *80*, 1147–1162. doi:10.1111/j.1467-8624.2009.01322.x.

Sabbagh, M. A., Hopkins, S. F. R., Benson, J. E., & Randall Flanagan, J. (2010). Conceptual change and preschoolers' theory of mind: Evidence from load–force adaptation. *Neural Networks*, *23*, 1043–1050. doi:10.1016/j.neunet.2010.08.007.
Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, *77*, 1034–1049. doi:10.1111/j.1467-8624.2006.00917.x.
Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*, 74–81. doi:10.1111/j.1467-9280.2005.01667.x.
Sage, K. D. (2012). *Attending to action at your own pace: Benefits for knowledge acquisition?* (Doctoral dissertation). Eugene, OR: University of Oregon. <http://hdl.handle.net/1794/12523>.
Santostefano, S. (1988). *Cognitive Control Battery (CCB): Manual*. Western Psychological Services.
Schulenberg, J. E., & Maslowsky, J. (2009). Taking substance use and development seriously: Developmentally distal and proximal influences on adolescence drug use. *Monographs of the Society for Research in Child Development*, *74*, 121–130. doi:10.1111/j.1540-5834.2009.00544.x.
Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology*, *26*, 445–464. doi:10.1207/s15326942dn2601_5.
Shing, Y. L., Lindenberger, U., Diamond, A., Li, S.-C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology*, *35*, 679–697. doi:10.1080/87565641.2010.508546.
Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, *110*, 115–122. doi:10.1016/j.cognition.2008.11.002.
Simonds, J. (2006). *The role of reward sensitivity and response execution in childhood extraversion* (Doctoral dissertation). Eugene, OR: University of Oregon. <http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/pdf/Simonds_Dissertation_2006.pdf>.
Simpson, A., & Riggs, K. J. (2005a). Factors responsible for performance on the day–night task: Response set or semantics? *Developmental Science*, *8*, 360–371. doi:10.1111/j.1467-7687.2005.00424.x.
Simpson, A., & Riggs, K. J. (2005b). Inhibitory and working memory demands of the day–night task in children. *The British Journal of Developmental Psychology*, *23*, 471–486. doi:10.1348/026151005X28712.
Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a "button-press" go/no-go task. *Journal of Experimental Child Psychology*, *94*, 18–26. doi:10.1016/j.jecp.2005.10.003.
Simpson, A., & Riggs, K. J. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology*, *43*, 417–428. doi:10.1037/0012-1649.43.2.417.
Simpson, A., & Riggs, K. J. (2009). What makes responses prepotent for young children? Insights from the grass–snow task. *Infant and Child Development*, *18*, 21–35. doi:10.1002/icd.576.
Simpson, A., Riggs, K. J., & Simon, M. (2004). What makes the windows task difficult for young children: Rule inference or rule use? *Journal of Experimental Child Psychology*, *87*, 155–170. doi:10.1016/j.jecp.2003.11.002.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. USA: Oxford University Press.
Sinn, N., Bryan, J., & Wilson, C. (2008). Cognitive effects of polyunsaturated fatty acids in children with attention deficit hyperactivity disorder symptoms: A randomised controlled trial. *Prostaglandins, Leukotrienes, and Essential Fatty Acids*, *78*, 311–326. doi:10.1016/j.plefa.2008.04.004.
Skowron, E. A., Cipriano-Essel, E., Gatzke-Kopp, L. M., Teti, D. M., & Ammerman, R. T. (2014). Early adversity, RSA, and inhibitory control: Evidence of children's neurobiological sensitivity to social context. *Developmental Psychobiology*, *56*, 964–978. doi:10.1002/dev.21175.
Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLoS ONE*, *8*, 1–11. doi:10.1371/journal.pone.0059510.
Smith, H. J., Kryski, K. R., Sheikh, H. I., Singh, S. M., & Hayden, E. P. (2013). The role of parenting and dopamine D4 receptor gene polymorphisms in children's inhibitory control. *Developmental Science*, *16*, 515–530. doi:10.1111/desc.12046.
Smith, H. J., Sheikh, H. I., Dyson, M. W., Olino, T. M., Laptook, R. S., Durbin, C. E., et al. (2012). Parenting and child DRD4 genotype interact to predict children's early emerging effortful control. *Child Development*, *83*, 1932–1944. doi:10.1111/j.1467-8624.2012.01818.x.
Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*, 173–187. doi:10.1016/j.ecresq.2007.01.002.
Sonuga-Barke, E. J. S., Dalen, L., Daley, D., & Remington, B. (2002). Are planning, working memory, and inhibition associated with individual differences in preschool ADHD symptoms? *Developmental Neuropsychology*, *21*, 255–272. doi:10.1207/S15326942DN2103_3.
Steele, A., Karmiloff-Smith, A., Cornish, K., & Scerif, G. (2012). The multiple subfunctions of attention: Differential developmental gateways to literacy and numeracy. *Child Development*, *83*, 2028–2041. doi:10.1111/j.1467-8624.2012.01809.x.
Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. doi:10.1146/annurev.clinpsy.032408.153639.
Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon Says". *Child Development*, *44*, 849–853. doi:10.1111/1467-8624.ep12116436.
Sulik, M. J. (2011). *Parasympathetic nervous system function, temperament, and adjustment in preschoolers* (Master's thesis). Tempe, AZ: Arizona State University. <http://hdl.handle.net/2286/j6kmvlijjyw>.
Sulik, M. J., Eisenberg, N., Silva, K. M., Spinrad, T. L., & Kupfer, A. (2013). Respiratory sinus arrhythmia, shyness, and effortful control in preschool-age children. *Biological Psychology*, *92*, 241–248. doi:10.1016/j.biopsycho.2012.10.009.
Sulik, M. J., Huerta, S., Zerr, A. A., Eisenberg, N., Spinrad, T. L., Valiente, C., et al. (2010). The factor structure of effortful control and measurement invariance across ethnicity and sex in a high-risk sample. *Journal of Psychopathology and Behavioral Assessment*, *32*, 8–22.

Tager-Flusberg, H., & Joseph, R. M. (2005). Theory of mind, language, and executive functions in autism: A longitudinal perspective. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development. Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 239–258). Mahwah, NJ, USA: Psychology Press.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, *79*, 866–881. doi:10.1111/j.1467-8624.2008.01164.x.

Tardif, T., So, C. W.-C., & Kaciroti, N. (2007). Language and false belief: Evidence for general, not specific, effects in Cantonese-speaking preschoolers. *Developmental Psychology*, *43*, 318–340. doi:10.1037/0012-1649.43.2.318.

Thompson, J. (2011). *The effect of prenatal exposure to methamphetamine on children's executive functioning at age 4* (Master's thesis). Albany, New Zealand. <http://hdl.handle.net/10179/3187>.

Thompson, S. F., Lengua, L. J., Zalewski, M., & Moran, L. (2013). Income and the development of effortful control as predictors of teacher reports of preschool adjustment. *Early Childhood Research Quarterly*, *28*, 784–793. doi:10.1016/j.ecresq.2013.07.006.

Thorell, L. B., & Wåhlstedt, C. (2006). Executive functioning deficits in relation to symptoms of ADHD and/or ODD in preschool children. *Infant and Child Development*, *15*, 503–518. doi:10.1002/icd.475.

Tregay, J., Gilmour, J., & Charman, T. (2009). Childhood rituals and executive functions. *The British Journal of Developmental Psychology*, *27*, 283–296. doi:10.1348/026151008X299737.

Utendale, W. T., Nuselovici, J., Saint-Pierre, A. B., Hubert, M., Chochol, C., & Hastings, P. D. (2014). Associations between inhibitory control, respiratory sinus arrhythmia, and externalizing problems in early childhood. *Developmental Psychobiology*, *56*, 686–699. doi:10.1002/dev.21136.

van der Kooy-Hofland, V. A. C., van der Kooy, J., Bus, A. G., van Ijzendoorn, M. H., & Bonsel, G. J. (2012). Differential susceptibility to early literacy intervention in children with mild perinatal adversities: Short- and long-term effects of a randomized control trial. *Journal of Educational Psychology*, *104*, 337–349. doi:10.1037/a0026984.

van Tilborg, A., Segers, E., van Balkom, H., & Verhoeven, L. (2014). Predictors of early literacy skills in children with intellectual disabilities: A clinical perspective. *Research in Developmental Disabilities*, *35*, 1674–1685. doi:10.1016/j.ridd.2014.03.025.

von Stauffenberg, C., & Campbell, S. B. (2007). Predicting the early developmental course of symptoms of attention deficit hyperactivity disorder. *Journal of Applied Developmental Psychology*, *28*, 536–552. doi:10.1016/j.appdev.2007.06.011.

Velásquez, A. M. (2010). *Classroom context and aggression: Understanding the effects of and change in classroom social structure* (Doctoral thesis). Montreal, Canada: Concordia University. <http://spectrum.library.concordia.ca/7404/>.

Verkerk, G. J. Q. (2014). *Functioning of very preterm born children at preschool age: Follow-up of an early intervention programme* (Doctoral thesis). Amsterdam, Netherlands: Universiteit van Amsterdam. <http://hdl.handle.net/11245/1.403294>.

Visu-Petra, L., Benga, O., & Miclea, M. (2007). Dimensions of attention and executive functioning in 5-to 12-years-old children: Neuropsychological assessment with the NEPSY battery. *Cognition, Brain, Behavior*, *11*, 585–608.

Viterbori, P., Gandolfi, E., & Usai, M. C. (2012). Executive skills and early language development. *Journal of Applied Psycholinguistics*, *3*, 17–32.

Walker, R. D. (2003). *Inhibitory control and its relation to theory of mind, parental discipline, and parental self-control in African American preschool children* (Master's thesis). Miami, OH: Miami University. <http://rave.ohiolink.edu/etdc/view?acc_num=miami1071171852>.

Watson, A. J., & Bell, M. A. (2013). Individual differences in inhibitory control skills at three years of age. *Developmental Neuropsychology*, *38*, 1–21. doi:10.1080/87565641.2012.718818.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., et al. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*, S54–S64. doi:10.1212/WNL.0b013e3182872ded.

Wellman, H. M., Lane, J. D., LaBounty, J., & Olson, S. L. (2011). Observant, nonaggressive temperament predicts theory-of-mind development. *Developmental Science*, *14*, 319–326. doi:10.1111/j.1467-7687.2010.00977.x.

Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology*, *44*, 618–623. doi:10.1037/0012-1649.44.2.618.

Welsh, M. C., Friedman, S. L., & Spieker, S. J. (2006). Executive functions in developing children: Current conceptualizations and questions for the future. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 167–187). Malden: Blackwell Publishing.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, *9*, 476–487. doi:10.1016/j.elerap.2010.07.003.

Westland, J. C. (2012). Erratum to "Lower bounds on sample size in structural equation modeling" [Electron. Commerce Res. Appl. 9 (6) (2010) 476–487]. *Electronic Commerce Research and Applications*, *11*, 445. doi:10.1016/j.elerap.2012.06.001.

White, L. K., McDermott, J. M., Degnan, K. A., Henderson, H. A., & Fox, N. A. (2011). Behavioral inhibition and anxiety: The moderating roles of inhibitory control and attention shifting. *Journal of Abnormal Child Psychology*, *39*, 735–747. doi:10.1007/s10802-011-9490-x.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*, 10–18. doi:10.1111/j.1750-8606.2009.00110.x.

Wiebe, S. A., Clark, C. A. C., De Jong, D. M., Chevalier, N., Espy, K. A., & Wakschlag, L. (2015). Prenatal tobacco exposure and self-regulation in early childhood: Implications for developmental psychopathology. *Development and Psychopathology*, *27*, 397–409. doi:10.1017/S095457941500005X.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*, 575–587. doi:10.1037/0012-1649.44.2.575.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*, 436–452. doi:10.1016/j.jecp.2010.08.008.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, *10*, 395–426. doi:10.1017/s0954579498001667.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology*, *35*, 205–213. doi:10.1037/0012-1649.35.1.205.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, *22*, 306–317. doi:10.1037/a0018708.

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment*, *24*, 418–431. doi:10.1037/a0025779.

Wolfe, C. D., & Bell, M. A. (2004). Working memory and inhibitory control in early childhood: Contributions from physiology, temperament, and language. *Developmental Psychobiology*, *44*, 68–83. doi:10.1002/dev.10152.

Wolfe, C. D., & Bell, M. A. (2007a). The integration of cognition and emotion during infancy and early childhood: Regulatory processes associated with the development of working memory. *Brain and Cognition*, *65*, 3–13. doi:10.1016/j.bandc.2006.01.009.

Wolfe, C. D., & Bell, M. A. (2007b). Sources of variability in working memory in early childhood: A consideration of age, temperament, language, and brain electrical activity. *Cognitive Development*, *22*, 431–455. doi:10.1016/j.cogdev.2007.08.007.

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: Typical developmental trends. *Journal of Child Psychology and Psychiatry*, *44*, 561–575. doi:10.1111/1469-7610.00145.

Wyss, N. (2013). *Relations among endogenous attention, executive functioning, and global assessment measures in toddlers and preschoolers born full-term and preterm* (Doctoral dissertation). Chicago, IL: Loyola University. <http://ecommons.luc.edu/luc_diss/735/>.

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*, 537–554. doi:10.1002/icd.1759.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. In P. D. Zelazo & P. J. Bauer (Eds.), *Monographs of the Society for Research in Child Development* (Vol. 78, pp. 16–33).

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*, 354–360. doi:10.1111/j.1750-8606.2012.00246.x.

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives*, *6*, 112–121. doi:10.1111/j.1750-8606.2011.00176.x.