

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications

Psychology, Department of

6-2018

The Benefits of Adding a Brief Measure of Simple Reaction Time to The Assessment of Executive Function Skills in Early Childhood

Michael T. Willoughby
RTI International

Clancy B. Blair
New York University

Laura J. Kuhn
University of North Carolina - Chapel Hill

Brooke E. Magnus
Marquette University, brooke.magnus@bc.edu

Follow this and additional works at: https://epublications.marquette.edu/psych_fac

 Part of the [Psychology Commons](#)

Recommended Citation

Willoughby, Michael T.; Blair, Clancy B.; Kuhn, Laura J.; and Magnus, Brooke E., "The Benefits of Adding a Brief Measure of Simple Reaction Time to The Assessment of Executive Function Skills in Early Childhood" (2018). *Psychology Faculty Research and Publications*. 323.
https://epublications.marquette.edu/psych_fac/323

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications/College of Arts and Sciences

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

Journal of Experimental Child Psychology, Vol. 170, (June 2018): XX-XX. [DOI](#). This article is © Elsevier and permission has been granted for this version to appear in [e-Publications@Marquette](#). Elsevier does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Elsevier.

The benefits of adding a brief measure of simple reaction time to the assessment of executive function skills in early childhood

Michael T. Willoughby

RTI International, Research Triangle Park, NC

Clancy B. Blair

Department of Applied Psychology, New York University, New York, NY

Laura J. Kuhn

Frank Porter Graham Child Development Institute, University of North Carolina, Chapel Hill, NC

Brooke E. Magnus

Department of Psychology, Marquette University, Milwaukee, WI

Abstract

Early childhood represents a period of rapid cognitive developmental change in executive function (EF) skills along with a variety of related cognitive processes, including processing speed. This leads to interpretational

challenges in that children's performance on EF tasks reflects more than EF skills per se. We tested whether the inclusion of a brief measure of simple reaction time (SRT) during EF assessments could help to partially address this challenge. Data were drawn from a cross-sectional convenience sample of 830 preschool-aged children. Individual differences in SRT were significantly associated with performance on all tasks (R^2 s = .09–.26); slower performance on the SRT task was associated with poorer performance on each EF task. Age-related differences in individual EF tasks were reduced by approximately one half after accounting for age-related differences in SRT, and EF task scores were less coherent (i.e., less strongly intercorrelated with each other) after the removal of SRT. Age-related differences in EF were smaller (Cohen d s = 1.36 vs. 0.78), and poverty-related differences in EF were larger (Cohen d s = 0.30 vs. 0.46) after accounting for SRT-related variation. Finally, consistent with previous studies, SRT-related differences in fluid reasoning were mediated by EF skills. Results are discussed with respect to using a brief measure of SRT to partially address the problem of measurement impurity at the level of individual EF tasks.

Keywords

Cognitive development, Assessment, Simple reaction time, Executive function, Early childhood, Measurement impurity

Introduction

Executive function (EF) skills are important for organizing information, planning and problem solving, and orchestrating thought and action in support of goal-directed behavior ([Diamond, 2013](#)). EF skills develop gradually from early childhood through early adulthood ([De Luca & Leventer, 2008](#)) and make unique contributions to children's social, behavioral, emotional, and academic development ([Ursache, Blair, & Raver, 2012](#)). There is strong interest among researchers, educators, and clinicians in measuring EF for purposes of basic and applied research as well as intervention-focused efforts.

One of the challenges in the early childhood literature on EF is that children's performance on tasks that purportedly measure EF skills reflects a wide range of cognitive processes, not just EF skills. For example, a variety of perceptual, receptive language, speed of processing, and fine motor skills likely contribute to children's performance on EF tasks. Many of these cognitive processes, which have been collectively referred to as foundational cognitive abilities (FCAs), develop rapidly in early childhood (see [Espy, 2017](#)). Moreover, individual differences in basic information processing abilities that emerge in infancy, including processing speed, contribute to children's performance on EF tasks in early childhood ([Hendry, Jones, & Charman, 2016](#)). Hence, young children's performance on EF tasks likely represents a confluence of EF and non-EF cognitive processes that are developing across the first 5 years of life.

It has become increasingly common for researchers to administer multiple EF tasks to young children and to use confirmatory factor analytic (CFA) models to represent EF ability as that variation that is shared across tasks ([Hughes et al., 2010](#), [Miller et al., 2012](#), [Monette et al., 2015](#), [Wiebe et al., 2011](#)). However, to the extent that non-EF processes contribute to performance across multiple EF tasks, latent variable representations of EF

skills continue to suffer from the problem of conflating EF and non-EF sources of variation. Espy and colleagues recently drew attention to this problem and advised researchers to include measures of FCAs when assessing EF skills ([Clark et al., 2014](#), [Espy, 2017](#)). They also advocated for the use of bifactor (an extension of CFA) models to explicitly remove FCA-related variation from the latent variable that represented EF. Removing FCA-related variation affected the associations between the construct of EF and criterion variables, including attention-deficit/hyperactivity disorder behaviors and socioeconomic status. A similar approach was used by van der Sluis and colleagues. In their study, the removal of rapid naming speed-related variation from EF tasks attenuated the association between the construct of EF and criterion measures of academic achievement ([van der Sluis, de Jong, & van der Leij, 2007](#)). Both studies underscore the value of incorporating measures of non-EF skills as a part of EF assessments.

As elaborated elsewhere, we have concerns about the use of CFA models, including bifactor models, as an approach for representing children's performance across a battery of EF tasks ([Willoughby, 2014](#), [Willoughby et al., 2014](#)). The crux of our concern stems from the repeated observation that children's performance across a battery of EF tasks is often weakly correlated ($r_s \approx .30$). When CFA methods are used to represent individual differences in latent EF skills, most of the observed variation in each task is attributed to the residual error term, which includes systematic variation specific to each task and measurement error, and the resulting estimate of true score ability has uncertain meaning and implausibly strong temporal stability ([Willoughby, Blair, & The Family Life Project Investigators, 2016](#); [Willoughby, Kuhn, Blair, Samek, & List, 2017](#)). Hence, although we agree with the suggestion that non-EF-related cognitive processes should be included as part of routine assessments of EF, we question whether the application of a CFA or bifactor modeling approach to a battery of tasks is the preferred or necessary solution ([Willoughby, 2017](#)). The current study considered a simple alternative approach that adjusts individual EF task scores.

One of the practical challenges of administering both EF and non-EF (e.g., FCAs, rapid naming speed) tasks is that it increases overall assessment time. Although issues of test burden are important for children of all ages, they are particularly salient when working with preschool-aged children, for whom assessment time is typically constrained to 30–45 min per testing occasion. In many contexts, it is simply not feasible to administer multiple EF and non-EF tasks in a single testing occasion, and multiple testing occasions are often too costly to implement.

In situations where the measurement of EF abilities is of primary importance, it is desirable to approximate non-EF abilities using assessments as brief as possible. There are three reasons why simple reaction time (SRT) measures may be ideally suited for this purpose. First, SRT measures index individual differences in general speed of processing, which contribute to—but are distinct from—EF skills. Unlike complex processing speed measures, SRT measures do not make demands on executive processes ([Cepeda, Blackwell, & Munakata, 2013](#)). Second, SRT undergoes a protracted period of development, with prominent changes evident during the early childhood period ([Kail,](#)

[1991](#), [Kail, 2007](#), [Kail and Ferrer, 2007](#)). Developmental improvements in SRT derive from global changes in white matter development ([Chevalier et al., 2015](#), [Scantlebury et al., 2014](#)). SRT may serve as a proxy for global aspects of cognitive development related to white matter development. Third, individual differences in SRT and related processing speed measures that emerge in infancy and toddlerhood predict children's later performance on EF tasks ([Rose et al., 2011](#), [Rose et al., 2012](#)). Moreover, clinical group differences in EF are often partially or completely accounted for by individual differences in processing speed and SRT ([Cassidy et al., 2016](#), [Kalff et al., 2005](#), [Marco et al., 2012](#), [Mulder et al., 2011](#)). Consequently, SRT may be an especially good candidate to serve as a control variable for inclusion in EF assessments of young children.

The computerized administration of EF tasks makes it easy to assess SRT ([Burke et al., 2017](#)). We recently developed a brief (1- to 2-min) SRT task that children complete immediately before EF assessments. We expected that faster (shorter) SRT would be associated with better performance on all EF tasks. The overarching objective of this study was to test the utility of incorporating a brief measure of SRT as part of an EF assessment. This objective was tested in four ways. First, although early childhood is often characterized as the developmental period in which EF changes are most prominent ([Zelazo et al., 2013](#)), some of the observed changes are likely due to age-related variation in non-EF processes, including processing speed ([Kail, 2002](#)). We tested whether age-related variation in EF task performance would be less prominent after the removal of SRT. Second, factor analytic studies are a widely used method to represent true score variation in EF as well as to test questions about changes in the factor structure (i.e., dimensionality) of EF across time (e.g., [Lee, Bull, & Ho, 2013](#)). Factor analytic models impose structure on the covariances (correlations) between tasks. Given that processing speed accounts for some of the shared association between EF tasks ([Rose et al., 2011](#)), we hypothesized that the removal of SRT-related variation from EF task scores would result in weaker correlations between tasks (i.e., EF tasks would become less coherent). Third, the well-known associations among processing speed (which are often measured using SRT measures), working memory (a component of EF), and fluid reasoning have been described as a developmental cascade such that age-related improvements in speed contribute to improved working memory capacity, which in turn contributes to better fluid reasoning ([Fry & Hale, 2000](#)). One way to test the utility of a brief measure of SRT is to test whether this same pattern of associations is evident even among cross-sectional data. We hypothesized that the contributions of SRT to fluid reasoning would be mediated through EF. Finally, poverty is associated with poorer performance on EF tasks and negatively affects the neural substrates that support EF task performance ([Hackman et al., 2015](#), [Johnson et al., 2016](#); [Raver, Blair, Willoughby, & The Family Life Project Investigators, 2013](#)). Although we are not familiar with studies that have considered the effects of poverty on SRT, poverty has been associated with slower reaction times in attentional assessments ([Mezzacappa, 2004](#)). In the absence of clear hypotheses, we conducted an exploratory test of whether the magnitude of the effect of poverty on EF varied after adjustment for SRT.

In sum, the overarching objectives of this study were to examine the potential benefits of including a brief measure of SRT as a part of EF assessment and investigating the extent to which the removal or covarying of SRT affected expected associations between EF and criterion measures. We focused on individual differences in SRT because they can be rapidly assessed, are distinct from EF abilities, have been associated with EF task performance in older children, exhibit developmental change in early childhood, and are associated with general mental ability and performance on a wide variety of cognitive tasks.

Method

Participants and procedure

This study involved children who attended preschools in New York or North Carolina in the eastern United States. A quota-based sampling procedure was used to guide participant recruitment ([Lohr, 1999](#)). Specifically, using data from the 2012 U.S. Census for children up to 5 years of age, we established a distribution of children with respect to household income (i.e., poor <100%, near poor = 100–200%, and not poor >200% of the U.S. federal poverty threshold for a given household size), race (i.e., Caucasian, African American, Asian, Native American, Pacific Islander), and ethnicity (i.e., Hispanic, not Hispanic). We then used this information to generate 180 mutually exclusive cells (i.e., 3 income × 5 race × 2 ethnicity × 3 age × 2 gender) that served to guide child recruitment. Consistent with a quota sampling approach, the intent was not to recruit exact numbers of children into given cells (in fact, all eligible and interested children were enrolled). Rather, the intent was to recruit a large convenience sample of children who exhibited diversity with respect to race, ethnicity, household income level, age, and gender using the expected cell counts as targets. In general, we met or exceeded the target number of children for most race × ethnicity × income level cells, with the exception that White children were underrepresented across all income levels.

As an initial step in recruitment, we approached the directors of center-based preschools that served children 3–5 years of age from a variety of communities. Center directors distributed consent forms to parents of children who were in the target age range. Interested parents who returned their consent forms to preschools were contacted by research staff for a screening phone call. During the call, parents provided demographic information about themselves and their child, including the number of persons living in their household and estimated total household annual income. This income information was used to construct income-to-needs ratios using poverty thresholds provided by the U.S. Census Bureau and to assign children to different income strata for recruitment purposes (income-to-needs 0–1 = poor, 1–2 = near poor, and >2 = not poor). Children who were outside of the target age (i.e., 3.0–5.9 years), who had physical or mental disabilities that precluded them from participating in direct assessments, or who did not speak English were not eligible to participate. Following the recruitment phone call, children participated in a one-time assessment of EF abilities in their preschool. Of the 924 families who completed recruitment phone calls and were eligible to participate, 92% ($n = 846$) were

tested. This study was based on a total of 830 children, from 395 classrooms and 175 centers, who completed at least one of the tasks from the *EF Touch* battery (16 children who were enrolled in the study but did not complete any EF Touch tasks were not included). Consistent with the sampling plan, participating children varied with respect to gender (50% male), age ($M = 4.4$ years, $SD = 0.7$, range = 3.0–5.9), race (60% White, 31% African American, 7% Asian American, 1% Native American, 1% Pacific Islander), ethnicity (20% Hispanic), and household income level (25% poor, 21% near poor, 54% not poor). All study procedures were approved by the institutional review board at the University of North Carolina at Chapel Hill.

Due to the large number of tasks, a planned missing design was used to determine the subset of tasks each child completed. Specifically, 3-year-olds were administered five tasks from the EF Touch battery and two age-appropriate subtasks from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) that measured working memory. Given that 4- and 5-year-olds could tolerate a longer assessment, they completed five tasks from the EF Touch battery, two age-appropriate subtasks from the WPPSI that measured fluid reasoning, and either two additional WPPSI tasks that measured working memory or two tasks that emulated the NIH (National Institutes of Health) Toolbox (Flanker and Dimensional Change Card Sort). The order in which the assigned tasks were administered was counterbalanced across children (e.g., EF Touch, WPPSI, and NIH Toolbox tasks were randomized in terms of whether they were presented first, second, or third within a visit). Moreover, the choice of tasks from the EF Touch battery was also randomized. Individual testing sessions, completed by a single research assistant, lasted approximately 30–45 min for 3-year-olds and 45–60 min for 4- and 5-year-olds. Preschool centers received \$5 for each completed consent form. Parents received \$40 for the completion of the screening phone call. Children received a small gift for participation.

Measures

EF touch

EF Touch is a computerized battery of EF tasks that were designed specifically for use with preschoolers. EF Touch was initially created, administered, and extensively studied in paper-and-pencil formats ([Willoughby and Blair, 2011](#), [Willoughby et al., 2013](#); [Willoughby, Wirth, Blair, & The Family Life Project Investigators, 2012](#)). More recently, the battery has been computerized to improve the efficiency, standardization, and sensitivity of tasks ([Willoughby & Blair, 2016](#)).

The EF Touch program runs in a Windows OS environment and requires two monitors. One standard monitor displays a script to the interviewer. Children respond to stimuli by touching a capacitive touch-screen monitor (15-in. Planar capacitive touch-screen monitors were used in this study). The battery is modular in nature (i.e., any number of tasks can be administered in any desired order), and each EF task takes 3–7 min to complete. Two warm-up tasks (1–2 min each) are typically administered first to acclimate children to using the touch screen (e.g., in an SRT task, a series of bubbles appear on the screen in random

locations and children are instructed to touch the bubbles as fast as they can). Because each of the tasks in the EF Touch program is described and studied in detail elsewhere, only abbreviated descriptions are provided here.

Bubbles

This 30-item task measured SRT. Each item presented a blue bubble for children to touch, with the goal of popping the bubble. Children were instructed to touch the bubble as fast as they could (only touches of the bubble were recognized; touches anywhere else on the screen were not registered). Items were presented for up to 5000 ms, and the time that transpired between stimuli onset and child bubble touching was recorded. The interstimulus interval was not standardized; the speed of items was contingent on children's response time. If a child did not touch the bubble in the designated time (5000 ms), the item was considered inaccurate and reaction time was not recorded. Item responses that were faster than 400 ms were considered too fast to be plausible (e.g., they may reflect a response recorded from the previous screen) and were set to missing. The mean response time across all valid items was used as an index of SRT.

Spatial Conflict Arrows

This 36-item spatial conflict task measured inhibitory control and cognitive flexibility. Two images of buttons appeared on the monitor, one each on the left- and right-most sides of the screen. Children were instructed to touch the button to which an arrow was pointing. Three blocks of 12 arrows were depicted in which arrows appeared above the button to which they were pointing (congruent condition), above the opposite button to which they were pointing (incongruent condition), or in mixed locations. Each item was presented for 3000 ms, and the accuracy and reaction time of responses were recorded.

Silly Sounds Stroop

This 17-item Stroop-like task measured inhibitory control. Each item displayed pictures of a dog and a cat (the left–right placement on the screen varied across trials) and presented the sound of either a dog barking or a cat meowing. Children were instructed to touch the picture of the animal that did not make the sound (e.g., touch the cat when hearing a dog bark). Each item was presented for 3000 ms, and the accuracy and reaction time of responses were recorded. Mean accuracy across all items was used to index performance.

Animal Go/No-Go

This 40-item go/no-go task measured inhibitory control. Individual pictures of animals were presented, and children were instructed to touch a centrally located button on their screen every time they saw an animal (the “go” response) except when that animal was a pig (the “no-go” response). Each item was presented for 3000 ms, and the accuracy and reaction

time of responses were recorded. Mean accuracy across all no-go responses was used to index task performance.

Working Memory Span

This 18-item span task measured working memory. Each item depicted a picture of one or more houses, each of which contained a picture of an animal, a colored dot, or a colored animal. Children verbally labeled the contents of each house. After a brief delay, the houses were displayed again without their contents. Children were asked to recall either the animal or color (of the animal) that was in each house (i.e., the nonrecalled contents served as a distraction). Items were organized into arrays of two-, three-, four-, and six-house trials. The mean accuracy of responses was used to index task performance.

Pick the Picture

This 32-item self-ordered pointing task measured working memory. Children were presented with arrays of pictures that varied in length (i.e., two, three, four, or six pictures per set). For each set, children were initially instructed to touch any picture of their choice. On subsequent trials within that set, the pictures were presented in different locations and children were instructed to pick a picture that had not yet been touched. The mean accuracy of responses was used to index task performance.

Something's the Same

This 30-item task was intended to measure attention shifting and flexible thinking. In the first 20 items, children were presented with two pictures (animals, flowers, etc.) that were described by the interviewer as being similar in color, shape, or size. A third picture was then presented alongside the original two pictures, and children were asked to select which of the original pictures was similar to the new picture along some other dimension (e.g., color, shape, size). In the last 10 items, children were presented with three pictures and were asked to identify two of the pictures that were similar and then a second pair of the same three pictures that were similar in some other way. The mean accuracy of responses was used to index task performance.

Wechsler preschool and primary scale of intelligence

The WPPSI–fourth edition (WPPSI-IV) is an indicator of full scale intelligence quotient intended for assessing cognitive functioning in two age bands: ages 2;6 (years;months) to 3;11 and ages 4;0 to 7;7 ([Wechsler, 2012](#)). The measure was normed on a stratified sample of 1700 children and has demonstrated good internal consistency, with average composite coefficients ranging from .75 to .90. Test–retest reliabilities for the composites were also sufficient (.84–.89) ([Syeda & Climie, 2014](#)). In the current study, we administered up to four subtests (i.e., Zoo Location, Picture Memory, Matrix Reasoning, Picture Concepts) that are used to create the fluid reasoning index and the working memory index.

Analytic plan

Item-level confirmatory factor models were fit to each EF task using *Mplus* Version 7.3. Models were initially fit using a weighted least squares with mean and variance (WLSMV) adjustment estimator to evaluate global model fit. These task-specific models were extended to include a manifest indicator of SRT (i.e., mean performance on the Bubbles task) as a predictor. The statistical significance and magnitude of coefficients associated with SRT informed the questions about the effects of SRT on each task. Clustered standard errors were used to acknowledge the nesting of children in preschool centers. Models were subsequently fit with a full information maximum likelihood (FIML) estimator to evaluate item parameters and obtain factor score estimates of latent ability (for the tasks that consisted of dichotomous indicators, these models were equivalent to two-parameter item response theory models).

The FIML-based factor score estimates of ability represent an estimate of latent EF ability as indexed by each task. These factor scores were regressed on SRT, and the residuals were saved as a measure of task performance that was free of the influence of SRT. Ordinary least squares (OLS) regression models were used because they represented a simple, easily replicated approach for removing SRT-related variation from individual EF task scores. Notably, the simultaneous estimation of factor scores and the regression coefficient of latent EF on SRT does not yield factor scores from which SRT-related variation is fully removed. This was the reason why we adopted the two-step approach (i.e., factor score estimation followed by OLS regression of factor scores on SRT).

Bivariate correlations were computed to ascertain whether individual EF task factor scores (prior to and following the removal of SRT-related variation) were differentially associated with child age. Comparisons of the dependent correlations were evaluated using [Lee and Preacher's \(2013\)](#) online utility that implements the method of [Steiger \(1980\)](#). Principal components analyses were conducted with EF task factor scores to test whether the removal of SRT-related variation reduced the coherence of EF task scores, where coherence was defined by the magnitude of intertask correlations and the percentage of variance accounted for by the first principal component. Linear mixed models were estimated to characterize whether and to what extent age- and poverty-related group differences in an overall EF composite score differed depending on whether it was created based on either adjusted or unadjusted EF task factor scores. Age and poverty categories were used to maintain consistency with the sampling approach as well as to facilitate the presentation of results (effect size comparisons via group comparisons). Finally, a path model was estimated to test whether the effects of SRT on fluid cognition were mediated by an EF composite score.

Results

As summarized in [Table 1](#), nearly complete data were available for the Bubbles (SRT) task because it was used as a way of acclimating children to the touch-screen assessment (i.e.,

touching the monitor to make responses and receiving visual and audible feedback following each registered touch). On average, it took children just over 1 s to respond to each of the SRT items, although there were appreciable individual differences ($M = 1242$ ms, range = 609–2577). Children completed an average of four of six possible EF tasks ($M = 4.4$ tasks, range = 1–6). This reflected a combination of planned missingness and the inability of some children to complete training items or to persist in testing. Children answered between 56% and 81% of the items on each EF task correctly, with the full range of responses evident on all tasks.

Table 1. Sample description.

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	Min	Max
Child age (years)	830	4.4	0.7	3.0	6.0
Number of administered tasks	830	4.3	0.7	1.0	6.0
Bubbles (ms)	812	1242.6	328.7	609.2	2577.2
Arrows+	645	.60	.29	.00	1.00
Silly Sound Stroop+	591	.58	.27	.00	1.00
Animal Go/No-Go+	563	.81	.24	.00	1.00
Working Memory Span+	475	.56	.20	.00	1.00
Pick the Picture+	467	.70	.13	.00	1.00
Something's the Same+	807	.76	.18	.00	1.00
	<i>n</i>	%			
Male	830	50			
Race (White)	502	60			
Race (African American)	257	31			
Race (Asian American)	58	7			
Race (Native American)	8	1			
Race (Pacific Islander)	5	1			
Ethnicity (non-Hispanic)	663	80			
Ethnicity (Hispanic)	167	20			
Poor (INR < 100%)	207	25			
Near poor (INR 100–200%)	172	21			
Not poor (INR > 200%)	451	54			

Note. +, Proportion of items correct; INR, income-to-needs ratio.

Association between SRT and performance on EF tasks

Item-level CFA models were fit to each EF task using a WLSMV estimator. Five of the six tasks had good global model fit, and one task (i.e., Something's the Same) had poor to modest fit (see [Table 2](#)). The latent variances and factor loadings in each model were always statistically significant, and standardized loadings were of moderate to strong magnitude. Each CFA model was reestimated using robust FIML estimation to obtain a

factor score estimate of true score ability that ignored SRT. These FIML models were then reestimated a second time, regressing the latent variable for each task onto the manifest measure of SRT (mean reaction time from the Bubbles task). SRT was moderately to strongly negatively associated with performance on all six EF tasks (p s < .001, R^2 s = .09–.26; see [Table 3](#)). Children with slower (longer) reaction times performed more poorly on each EF task.

Table 2. Global model fit for item-level CFA models for EF tasks that exclude (top) or include (bottom) simple reaction time as a predictor of the latent factor.

Task (construct)	N	Chi-square (df)	CFI	RMSEA (90% CI)
Arrows (IC)	645	221.2 (106) ^{***}	.98	.04 (.033–.049)
Silly Sound Stroop (IC)	591	230.5 (109) ^{***}	.97	.04 (.036–.051)
Animal Go/No-Go (IC)	563	17.2 (20)	1.00	.00 (.000–.031)
Working Memory Span (WM)	475	46.6 (27) [*]	.95	.04 (.019–.058)
Pick the Picture (WM)	467	65.5 (35) ^{**}	.96	.04 (.027–.059)
Something's the Same (AS)	807	412.4 (133) ^{***}	.89	.05 (.045–.057)
Arrows (IC)	628	234.1 (122) ^{***}	.98	.04 (.031–.046)
Silly Sound Stroop (IC)	583	264.7 (125) ^{***}	.95	.04 (.036–.051)
Animal Go/No-Go (IC)	553	23.5 (27)	1.00	.00 (.000–.028)
Working Memory Span (WM)	472	62.0 (35) ^{**}	.93	.04 (.023–.057)
Pick the Picture (WM)	463	70.0 (44) ^{**}	.96	.04 (.019–.051)
Something's the Same (AS)	795	433.3 (151) ^{***}	.88	.05 (.043–.054)

Note. CFI, comparative fit index; RMSEA, root mean square error of approximation; CI, confidence interval; IC, inhibitory control; WM, working memory; AS, attention shifting. All models used a weighted least squares with mean and variance adjustment estimator.

* p < .05.

** p < .01.

*** p < .001.

Table 3. Standardized coefficients relating simple reaction time to EF task performance.

Task	β	95% CI	R^2
Arrows	-.41	-.47 to -.34	.17
Silly Sound Stroop	-.51	-.59 to -.43	.26
Animal Go/No-Go	-.29	-.38 to -.20	.09
Working Memory Span	-.46	-.56 to -.35	.21
Pick the Picture	-.40	-.51 to -.30	.16
Something's the Same	-.38	-.45 to -.30	.14

Note. All $ps < .001$. CI, confidence interval. All models used a robust full information maximum likelihood estimator and involved clustered standard errors to accommodate nesting of children in classrooms.

Age-related changes in individual EF task scores

Child age was strongly negatively correlated with SRT ($r = -.59, p < .0001$); older children demonstrated faster SRT than their younger peers. The bivariate correlations between child age and both unadjusted and adjusted (for SRT) factor scores for each EF task are summarized in [Table 4](#). For all six tasks, age was more strongly associated with unadjusted EF task scores than with adjusted ones (all $ps < .001$; [Table 4](#)). As much as half of the apparent age-related change in EF was due to age-related change in SRT.

Table 4. Age-related variation in EF task performance as a function of scoring method.

Task	N	Factor score	Age correlation (r)	Comparison (z statistic)
Arrows	628	Unadjusted	.41	14.6
		Adjusted	.20	
Animal Go/No-Go	553	Unadjusted	.30	12.6
		Adjusted	.17	
Silly Sounds Stroop	583	Unadjusted	.47	13.6
		Adjusted	.23	
Working Memory Span	472	Unadjusted	.41	12.0
		Adjusted	.22	
Pick the Picture	463	Unadjusted	.31	12.7
		Adjusted	.12	
Something's the Same	795	Unadjusted	.37	16.3
		Adjusted	.20	

Note. All comparisons significant, $p < .001$.

Coherence of individual EF task scores

[Table 5](#) summarizes the pairwise correlations between EF tasks before and after adjustment for SRT-related variation. The average correlation coefficient among the unadjusted (mean $r = .29$) EF task scores was on average 50% larger than the corresponding correlations among the adjusted (mean $r = .20$) task scores. To further examine differences in task coherence, the expectation maximization algorithm was used to impute complete scores for unadjusted and adjusted tasks separately (this addressed missing task data that existed, in part, due to the planned missing design). A principal component analysis (PCA) of unadjusted scores revealed a single dominant factor that accounted for 41% of variation in tasks (first eigenvalue = 2.5; Kaiser's measure of sampling adequacy [MSA] = .78). By comparison, a PCA of adjusted task scores revealed a single dominant factor that accounted for 34% of variation in tasks (first eigenvalue = 2.1;

MSA = .72). EF task scores were less strongly associated after the removal of shared SRT-related variation.

Table 5. Pairwise correlations among EF tasks before and after adjustment for simple reaction time.

	1	2	3	4	5	6
1. Arrows	.93	.27	.42	.22	.26	.32
2. Animal Go/No-Go	.20	.97	.29	.26	.27	.25
3. Silly Sounds Stroop	.31	.22	.88	.30	.39	.40
4. Working Memory Span	.09	.18	.18	.93	.24	.18
5. Pick the Picture	.15	.24	.25	.17	.94	.25
6. Something's the Same	.24	.21	.30	.08	.17	.95

Note. $N_s = 173-807$. Values above and below the diagonal are unadjusted and adjusted (SRT-related variation removed), respectively. Values on the diagonal represent the association between adjusted and unadjusted scores for each task.

Age and poverty group differences in overall EF performance

Consistent with our recent work ([Willoughby et al., 2016](#)), a composite score was created to represent children's performance across the battery of EF tasks. Each composite was created with the mean of either the unadjusted or adjusted factor scores for all the tasks that a given child completed. The unadjusted and adjusted EF composites were standardized to a mean of 0 and a standard deviation of 1 to put them on the same scale.

Two linear mixed models were estimated, one for each EF composite score (based on either unadjusted or adjusted task factor scores), using child age group (3, 4, or 5 years) and household poverty status (poor, near poor, or not poor) as predictors. Random intercepts were included to accommodate the nonindependence of observations (children in classrooms). In both models, child age and household poverty groups were significantly associated with EF composite scores ($p_s < .0001$). Least square means were used in conjunction with the observed standard deviation of composite scores to characterize the magnitude of group differences as effect sizes (see [Fig. 1](#)). Consistent with the results from the individual tasks, age group differences were more pronounced when evaluated using the unadjusted EF composite (Cohen's $d = 1.36$ for 3- vs. 5-year-olds) compared with the adjusted EF composite (Cohen's $d = 0.78$ for 3- vs. 5-year-olds). The reverse was true for poverty such that group differences were more pronounced for the adjusted EF composite (Cohen's $d = 0.45$ for not poor vs. poor) compared with the unadjusted EF composite (Cohen's $d = 0.30$ for not poor vs. poor).

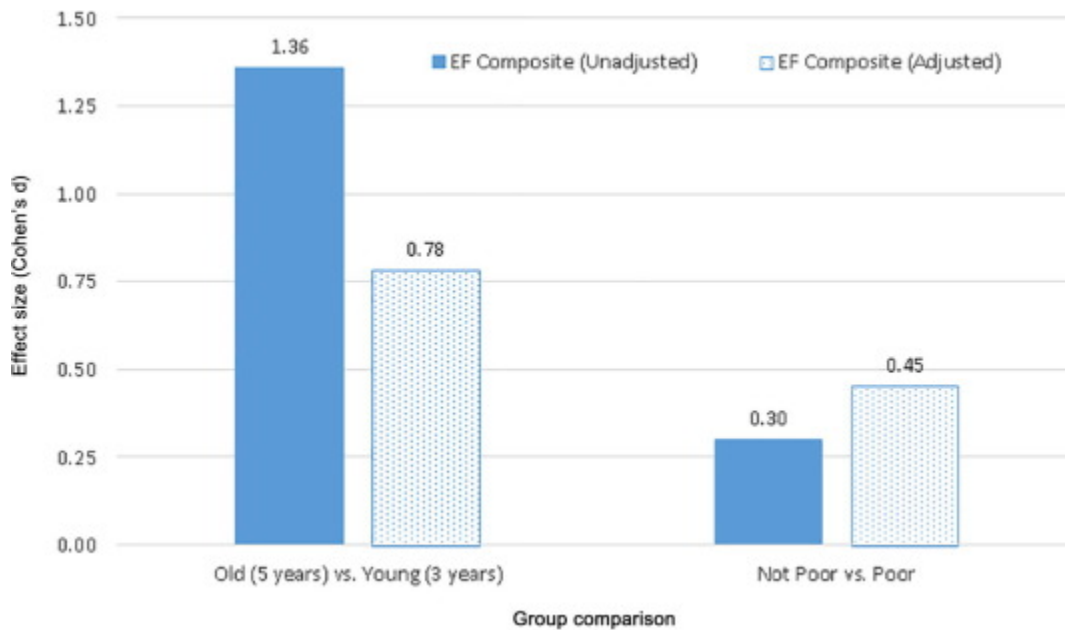


Fig. 1. Group differences in adjusted versus unadjusted EF composite scores.

The results involving poverty status were in the opposite direction of the age group differences (i.e., larger group differences were evident after the removal of SRT from EF scores). A post hoc explanation for this finding was that children who reside in low-income households may have faster SRTs than children who do not live in low-income households, and these faster SRTs may serve a compensatory function when completing EF tasks. To test this post hoc hypothesis, we estimated another mixed linear model in which SRT was regressed on poverty and age groups. Random intercepts were again included to accommodate the nonindependence of observations (children in classrooms). Poverty group was a significant predictor of SRT ($p = .02$) above and beyond the already established age group effect on SRT ($p < .0001$). Least square means, which accounted for age differences, for SRT among children in the poor, near poor, and not poor groups were 1183, 1187, and 1244, respectively. The poor and near poor groups did not differ from each other ($p = .90$), but both groups were faster than the not poor group ($ps = .02$ and $.03$, respectively). These group differences were of small absolute magnitude (poor vs. not poor, Cohen's $d = 0.19$; near poor vs. not poor, Cohen's $d = 0.17$).

Direct and indirect effects of SRT on fluid cognition through EF

A path model was estimated in which the WPPSI fluid reasoning index was simultaneously regressed on SRT and the unadjusted EF score composite, with the EF composite also regressed on SRT (812 children had data that contributed to this analysis). There were direct effects from SRT to EF ($\beta = -.51$, $p < .001$; $R^2 = .26$) or from EF to fluid reasoning ($\beta = .32$, $p < .001$; $R^2 = .12$). Although the direct effect from SRT to fluid reasoning was not statistically significant ($\beta = -.05$, $p = .35$), the indirect effect through EF was statistically significant ($\beta = -.16$, $p < .001$).

Most of the previous studies that have tested the association between processing speed (including SRT) and fluid reasoning focused on working memory, not overall EF, as a potential mediator. To provide data directly comparable to those studies, we created a working memory composite using the two working memory tasks that were part of the overall EF composite (i.e., Pick the Picture and Working Memory Span). The path model above was reestimated using the working memory composite instead of the EF composite (768 children had data that contributed to this analysis). There were direct effects from SRT to working memory ($\beta = -.37, p < .001; R^2 = .14$) and from working memory to fluid reasoning ($\beta = .16, p = .001; R^2 = .06$). Both the direct effect and indirect effect from SRT to fluid reasoning through EF were statistically significant ($\beta = -.14, p = .01$ and $\beta = -.06, p = .001$, respectively).

Discussion

Previous studies have shown SRT to be a central aspect of cognition that is associated with the development of reasoning ability in childhood ([Fry and Hale, 1996](#), [Kail, 2007](#)) and with its decline in later adulthood ([Salthouse, 1996](#)). This reflects the fact that SRT is highly correlated with age and reflects age-related changes in general cognitive ability. This makes SRT an ideal proxy for non-EF skills that contribute to EF task performance. We developed a brief measure of SRT (i.e., the Bubbles task) that was used to orient preschool-aged children to EF Touch. Specifically, the Bubbles task provided a way to introduce children to the idea that they would respond to interviewer questions by touching a monitor, that they would occasionally play games on their own, and that their screen touches were recognized through visual and audible feedback. The primary objective of this study was to ascertain whether the mean reaction time data from the Bubbles task was empirically useful as a proxy for non-EF skills that contribute to EF task performance.

EF has been more frequently studied in early childhood than in middle childhood ([Best, Miller, & Jones, 2009](#)). One of the most commonly offered reasons for this is the widespread belief that EF skills undergo more dramatic improvements in early childhood than in middle childhood, which has sometimes been interpreted as evidence that EF skills are more amenable to change in early childhood than in middle childhood ([Fox et al., 2010](#), [Huston and Bentley, 2010](#)). However, several aspects of cognition other than EF are also rapidly improving in early childhood, including general speed of processing. When we removed SRT from EF task scores, the correlations between EF and age were reduced by half. Most previous studies that characterized changes in EF skills across early childhood, including our own ([Willoughby et al., 2012](#)), failed to covary or otherwise adjust EF scores for SRT or other non-EF skills that contributed to task performance. The results of this study raise the possibility that EF skill development might not proceed any more rapidly in early childhood than in middle childhood. Rather, the appearance of more rapid changes in EF skills may be an artifact of corresponding faster changes in processing speed and reaction time in early childhood than in middle childhood. This is an important direction for future research because it has implications for the ages at which interventions for EF are developed.

One of the unresolved questions in the EF literature is the dimensionality of EF skills. Although the construct of EF is often characterized as undifferentiated (unidimensional) in early childhood and increasingly differentiated in structure (multidimensional) in middle childhood and adolescence, contradictory evidence abounds. The relatively weak associations that exist between individual EF task scores have long been a concern of methodologists ([Rabbitt, 1997](#)), and we have raised concerns that these weak associations undermine conventional efforts to test questions about the dimensionality of EF skills ([Willoughby et al., 2014](#)). In this study, we demonstrated that removing SRT from EF task scores results in even weaker bivariate associations between EF tasks, further compounding our concerns. Numerous reasons for the weak associations among EF tasks have been proposed, ranging from measurement problems, to problems with the EF construct itself, to potential issues relating to state-like, as opposed to trait-like, variation in EF task performance ([Blair, 2016](#)). Theoretical conceptualizations of the construct of EF have not yet resolved the weak empirical associations that exist between tasks that purportedly measure the same skills.

Poverty was more strongly associated with EF after the removal of SRT-related variation. Post hoc results indicated that children from low-income households exhibited faster SRT than did children from higher income households and that this faster SRT partially compensated for true group differences in EF task performance. There are many aspects of children and families that covary with poverty status. At the risk of oversimplifying a highly complex set of associations, poverty may affect both trait-like (e.g., SRT) and state-like (e.g., EF) aspects of cognitive development. The ongoing experience of poverty may be especially detrimental to EF skill development. For example, a growing experimental literature involving adults has indicated that conditions of scarcity, whether induced or an existing condition of participants, lead to reductions in EF, to a narrowing of the focus of attention, and to increased borrowing and risky decision making ([Mani et al., 2013](#), [Shah et al., 2012](#), [Shah et al., 2015](#)). Explanations for these observed associations have to do with the way in which stress affects EF at both the neurobiological ([Arnsten, 2009](#)) and behavioral/psychological ([Schmader & Johns, 2003](#)) levels. The results of this study are consistent with the idea that children who reside in low-income households may experience stressful life conditions that both impair EF skill development and encourage faster reaction time. This is clearly speculative and an idea that requires more systematic study.

As a final check on the validity of including a brief measure of SRT in our standard EF assessment, we estimated a series of path models in which SRT exerted both direct and indirect effects (through EF in general and working memory in particular) on fluid reasoning. Despite the cross-sectional design and limited age span relative to comparison studies, the observed pattern of associations was consistent with a developmental cascade model that aligns with this broader area of research (reviewed by [Fry & Hale, 2000](#)). These results provide additional evidence to support the utility of including even brief measures of SRT as a routine part of EF assessments.

Measurement impurity is a ubiquitous problem in all areas of cognitive assessment, including but not limited to the assessment of EF skills. We do not conceive of measurement impurity as a problem that can be definitively “solved.” Rather, we consider this to be a characteristic of performance-based tasks that may be managed through the explicit measurement of cognitive processes that contribute to EF task performance. Here, we demonstrated that SRT serves as one potential source of measurement impurity. Moreover, we demonstrated an easily replicated methodological approach for removing SRT-related variation from individual tasks rather than at the level of the latent variable, which has been the dominant approach in the literature to date (e.g., via confirmatory factor models). Addressing measurement impurity at the level of individual tasks has pragmatic benefits for researchers. Whereas traditional latent variable approaches require that at least three tasks be available to define each latent construct of interest (i.e., for purposes of model identification), this is not an issue when measurement impurity is addressed at the level of individual tasks. Similarly, whereas traditional latent variable models require relatively large samples for the asymptotic properties of estimators to be realized (e.g., maximum likelihood estimation), the OLS regression method that was employed here has a closed-form solution that is appropriate for use with small samples.

The approach to dealing with measurement impurity that was demonstrated in this study is clearly not applicable to all EF tasks. Some EF tasks are structured such that inferences about EF skills are derived from comparisons of children’s performance across conditions. For example, in the Flanker task, individual differences in EF skills are sometimes inferred from comparisons of children’s performance across congruent versus incongruent trials. Similarly, in task-switching paradigms, scores can be derived based on comparisons across single-task versus mixed-task blocks (i.e., so-called task-switching costs). In these instances, scores that are derived from the performance across conditions are implicitly assumed to account for individual differences in SRT and perhaps related processes (e.g., task engagement).

This study was characterized by at least five limitations. First, although participating children were recruited to ensure variability with respect to race, ethnicity, age, gender, and household income, this was a convenience sample of children who were primarily recruited from center-based preschools. This sampling plan may limit the generalizability of results. Second, although we attempted to make inferences about differences in the magnitude of age-related changes in EF task performance prior to and following the removing of SRT-related variation, these inferences were limited by our reliance on a cross-sectional (vs. longitudinal) design. Third, although we focused on SRT in this study, a variety of cognitive processes contribute to EF task performance. The inclusion of a brief measure of SRT is clearly not a sufficient proxy for all non-EF processes that affect EF task performance. Fourth, we used OLS regression models to remove SRT-related variation from EF tasks. The appeal of this approach includes its simplicity (i.e., facilitates widespread use), robustness to non-normality among EF task scores, and the availability of closed-form solutions that do not require minimum sample sizes. Nonetheless, depending on the nature of the sample and the distributional characteristics of the EF tasks in question, alternative approaches may be more appropriate (e.g., empirical Bayes estimates from hierarchical

generalized linear models). Fifth, it is not clear at what point adjusting EF task scores results in overcorrection. EF skills are interesting because they act on lower order forms of cognition. There would seem to be some risk for covarying cognitive processes out of EF scores that are presumably under the control of EF. This is an idea in need of deeper conceptual consideration (see, e.g., [Cepeda et al., 2013](#), [Dennis et al., 2009](#)).

In sum, this study demonstrated the value of including a brief measure of SRT in standard assessments of EF for preschool-aged children. Practically, this task helped to orient children to a touch-screen-enabled assessment. Empirically, children's SRT was consistently related to EF task performance. Removing SRT-related variation from EF task scores resulted in theoretically important results that challenge widespread assumptions about developmental changes in EF, the coherence between EF tasks, and the magnitude of poverty group differences. We hope that this and related work stimulate greater scrutiny of previous research, including its implications for theory and practice. The continued refinement of measurement issues related to EF skills have implications for our understanding of the developmental causes and course of EF skills as well as their consequences for other aspects of child development.

Acknowledgment

This project was supported by the Institute of Educational Sciences (Grant [R324A120033](#)). The views expressed in this article are those of the authors and do not necessarily represent the opinions or position of the Institute of Educational Sciences.

References

- 1 A.F.T. Arnsten. **Stress signalling pathways that impair prefrontal cortex structure and function**, *Nature Reviews Neuroscience*, 10 (2009), pp. 410-422.
- 2 J.R. Best, P.H. Miller, L.L. Jones. **Executive functions after age 5: Changes and correlates**, *Developmental Review*, 29 (2009), pp. 180-200.
- 3 C. Blair. **Developmental science and executive function**, *Current Directions in Psychological Science*, 25 (2016), pp. 3-7.
- 4 D. Burke, S. Linder, J. Hirsch, T. Dey, D. Kana, S. Ringenbach, ..., J. Alberts. **Characterizing information processing with a mobile device: Measurement of simple and choice reaction time**, *Assessment*, 24 (2017), pp. 885-895.
- 5 A.R. Cassidy, M.T. White, D.R. DeMaso, J.W. Newburger, D.C. Bellinger. **Processing speed, executive function, and academic achievement in children with dextro-**

transposition of the great arteries: Testing a longitudinal developmental cascade model, *Neuropsychology*, 30 (2016), pp. 874-885.

- 6 N.J. Cepeda, K.A. Blackwell, Y. Munakata. **Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control**, *Developmental Science*, 16 (2013), pp. 269-286.
- 7 N. Chevalier, S. Kurth, M.R. Doucette, M. Wiseheart, S.C. Deoni, D.C. Dean 3rd, ..., M.K. LeBourgeois. **Myelination is associated with processing speed in early childhood: Preliminary insights**, *PLoS ONE*, 10 (10) (2015), p. e0139897.
- 8 C.A. Clark, J.M. Nelson, J. Garza, T.D. Sheffield, S.A. Wiebe, K.A. Espy. **Gaining control: Changing relations between executive control and processing speed and their relevance for mathematics achievement over course of the preschool period**, *Frontiers in Psychology*, 5 (2014), [10.3389/fpsyg.2014.00107](https://doi.org/10.3389/fpsyg.2014.00107).
- 9 C.R. De Luca, R.J. Leventer. **Developmental trajectories of executive functions across the life span**, V. Anderson, R. Jacobs, P.J. Anderson (Eds.), *Executive functions and the frontal lobes*, Taylor & Francis, New York (2008), pp. 22-56.
- 10 M. Dennis, D.J. Francis, P.T. Cirino, R. Schachar, M.A. Barnes, J.M. Fletcher. **Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders**, *Journal of the International Neuropsychological Society*, 15 (2009), pp. 331-343.
- 11 A. Diamond. **Executive functions**, *Annual Review of Psychology*, 64 (2013), pp. 135-168.
- 12 Espy, K. (2017). The changing nature of executive control in preschool. *Monographs of the Society for Research in Child Development*, 81(4, Serial No. 323).
- 13 S.E. Fox, P. Levitt, C.A. Nelson 3rd. **How the timing and quality of early experiences influence the development of brain architecture**, *Child Development*, 81 (2010), pp. 28-40.
- 14 A.F. Fry, S. Hale. **Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade**, *Psychological Science*, 7 (1996), pp. 237-241.
- 15 A.F. Fry, S. Hale. **Relationships among processing speed, working memory, and fluid intelligence in children**, *Biological Psychology*, 54 (2000), pp. 1-34.
- 16 D.A. Hackman, R. Gallop, G.W. Evans, M.J. Farah. **Socioeconomic status and executive function: Developmental trajectories and mediation**, *Developmental Science*, 18 (2015), pp. 686-702.

- 17 A. Hendry, E.J.H. Jones, T. Charman. **Executive function in the first three years of life: Precursors, predictors, and patterns**, *Developmental Review*, 42 (2016), pp. 1-33.
- 18 C. Hughes, R. Ensor, A. Wilson, A. Graham. **Tracking executive function across the transition to school: A latent variable approach**, *Developmental Neuropsychology*, 35 (2010), pp. 20-36.
- 19 A.C. Huston, A.C. Bentley. **Human development in societal context**, *Annual Review of Psychology*, 61 (2010), pp. 411-437.
- 20 Johnson, S. B., Riis, J. L., & Noble, K. G. (2016). State of the art review: Poverty and the developing brain. *Pediatrics*. Advance online publication. <http://doi.org/10.1542/peds.2015-3075>.
- 21 R. Kail. **Developmental change in speed of processing during childhood and adolescence**, *Psychological Bulletin*, 109 (1991), pp. 490-501.
- 22 R. Kail. **Developmental change in proactive interference**, *Child Development*, 73 (2002), pp. 1703-1714.
- 23 R.V. Kail. **Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning**, *Psychological Science*, 18 (2007), pp. 312-313.
- 24 R.V. Kail, E. Ferrer. **Processing speed in childhood and adolescence: Longitudinal models for examining developmental change**, *Child Development*, 78 (2007), pp. 1760-1770.
- 25 A.C. Kalff, L.M.J. De Sonneville, P.P.M. Hurks, J.G.M. Hendriksen, M. Kroes, F.J.M. Feron, ..., J. Jolles. **Speed, speed variability, and accuracy of information processing in 5- to 6-year-old children at risk of ADHD**, *Journal of the International Neuropsychological Society*, 11 (2005), pp. 173-183.
- 26 Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [computer software]. Available from: <<http://quantpsy.org>>.
- 27 K. Lee, R. Bull, R.M. Ho. **Developmental changes in executive functioning**, *Child Development*, 84 (2013), pp. 1933-1953.
- 28 S.L. Lohr. **Sampling: Design and analysis**, Duxbury, Pacific Grove, CA (1999).
- 29 A. Mani, S. Mullainathan, E. Shafir, J.Y. Zhao. **Poverty impedes cognitive function**, *Science*, 341 (2013), pp. 976-980.
- 30 E.J. Marco, K.M. Harrell, W.S. Brown, S.S. Hill, R.J. Jeremy, J.H. Kramer, ..., L.K. Paul. **Processing speed delays contribute to executive function deficits in**

individuals with agenesis of the corpus callosum, Journal of the International Neuropsychological Society, 18 (2012), pp. 521-529.

- 31 E. Mezzacappa. **Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children**, Child Development, 75 (2004), pp. 1373-1386.
- 32 M.R. Miller, G.F. Giesbrecht, U. Muller, R.J. McInerney, K.A. Kerns. **A latent variable approach to determining the structure of executive function in preschool children**, Journal of Cognition and Development, 13 (2012), pp. 395-423.
- 33 S. Monette, M. Bigras, M.A. Lafreniere. **Structure of executive functions in typically developing kindergarteners**, Journal of Experimental Child Psychology, 140 (2015), pp. 120-139.
- 34 H. Mulder, N.J. Pitchford, N. Marlow. **Processing speed mediates executive function difficulties in very preterm children in middle childhood**, Journal of the International Neuropsychological Society, 17 (2011), pp. 445-454.
- 35 P. Rabbitt. **Introduction: Methodologies and models in the study of executive function**, P. Rabbitt (Ed.), Methodology of frontal and executive function, Psychology Press, New York (1997), pp. 1-38.
- 36 C.C. Raver, C. Blair, M. Willoughby, The Family Life Project Investigators. **Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility**, Developmental Psychology, 49 (2013), pp. 292-304.
- 37 S.A. Rose, J.F. Feldman, J.J. Jankowski. **Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement**, Developmental Science, 14 (2011), pp. 1161-1175.
- 38 S.A. Rose, J.F. Feldman, J.J. Jankowski. **Implications of infant cognition for executive functions at age 11**, Psychological Science, 23 (2012), pp. 1345-1355.
- 39 T.A. Salthouse. **The processing-speed theory of adult age differences in cognition**, Psychological Review, 103 (1996), pp. 403-428.
- 40 N. Scantlebury, T. Cunningham, C. Dockstader, S. Laughlin, W. Gaetz, C. Rockel, ..., D. Mabbott. **Relations between white matter maturation and reaction time in childhood**, Journal of the International Neuropsychological Society, 20 (2014), pp. 99-112.
- 41 T. Schmader, M. Johns. **Converging evidence that stereotype threat reduces working memory capacity**, Journal of Personality and Social Psychology, 85 (2003), pp. 440-452.

- 42 A.K. Shah, S. Mullainathan, E. Shafir. **Some consequences of having too little**, *Science*, 338 (2012), pp. 682-685.
- 43 A.K. Shah, E. Shafir, S. Mullainathan. **Scarcity frames value**, *Psychological Science*, 26 (2015), pp. 402-412.
- 44 J.H. Steiger. **Tests for comparing elements of a correlation matrix**, *Psychological Bulletin*, 87 (1980), pp. 245-251.
- 45 M.M. Syeda, E.A. Climie. **Test review: Wechsler preschool and primary scale of intelligence**, *Journal of Psychoeducational Assessment*, 32 (2014), pp. 265-272.
- 46 A. Ursache, C. Blair, C.C. Raver. **The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure**, *Child Development Perspectives*, 6 (2012), pp. 122-128.
- 47 S. van der Sluis, P.F. de Jong, A. van der Leij. **Executive functioning in children, and its relations with reasoning, reading, and arithmetic**, *Intelligence*, 35 (2007), pp. 427-449.
- 48 D. Wechsler. **Wechsler Preschool and Primary Scale of Intelligence—fourth edition (WPPSI-IV)**, Pearson, Bloomington, MN (2012).
- 49 S.A. Wiebe, T. Sheffield, J.M. Nelson, C.A.C. Clark, N. Chevalier, K.A. Espy. **The structure of executive function in 3-year-olds**, *Journal of Experimental Child Psychology*, 108 (2011), pp. 436-452.
- 50 Willoughby, M. T. (2014). Formative versus reflective measurement of executive function tasks: Response to commentaries and another perspective. *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 173–178.
- 51 M.T. Willoughby. **Commentary on the changing nature of executive control in preschool**, K. Espy (Ed.), *The changing nature of executive control in preschool*, Vol. 81, Society for Research in Child Development, Ann Arbor, MI (2017), pp. 150-164.
- 52 M. Willoughby, C. Blair. **Test–retest reliability of a new executive function battery for use in early childhood**
Child Neuropsychology, 17 (2011), pp. 564-579.
- 53 M.T. Willoughby, C.B. Blair. **Longitudinal measurement of executive function in preschoolers**, J. Griffin, L. Freund, P. McCardle (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopment, and translational research*, American Psychological Association Press, Washington, DC (2016), pp. 91-113.

- 54 M.T. Willoughby, C.B. Blair, The Family Life Project Investigators. **Measuring executive function in early childhood: A case for formative measurement**, *Psychological Assessment*, 28 (2016), pp. 319-330.
- 55 Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research and Perspectives*, 12(3), 69–95.
- 56 M.T. Willoughby, L.J. Kuhn, C.B. Blair, A. Samek, J.A. List. **The test–retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators**, *Child Neuropsychology*, 23 (2017), pp. 822-837.
- 57 M.T. Willoughby, J. Pek, C.B. Blair. **Measuring executive function in early childhood: A focus on maximal reliability and the derivation of short forms**, *Psychological Assessment*, 25 (2013), pp. 664-670.
- 58 M.T. Willoughby, R.J. Wirth, C.B. Blair, The Family Life Project Investigators. **Executive function in early childhood: Longitudinal measurement invariance and developmental change**, *Psychological Assessment*, 24 (2012), pp. 418-431.
- 59 P.D. Zelazo, J.E. Anderson, J. Richler, K. Wallner-Allen, J.L. Beaumont, S. Weintraub. **II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention**, *Monographs of the Society for Research in Child Development*, 78 (4) (2013), pp. 16-33.