Journal of Educational Psychology
2015, Vol. 107, No. 1, 207–221

0022-0663/15/$12.00 http://dx.doi.org/10.1037/a0037366

Prekindergarten Children's Executive Functioning Skills and Achievement Gains: The Utility of Direct Assessments and Teacher Ratings

Mary Wagner Fuhs
University of Dayton

Dale Clark Farran and Kimberly Turner Nesbitt
Vanderbilt University

An accumulating body of evidence suggests that young children who exhibit greater executive functioning (EF) skills in early childhood also achieve more academically. The goal of the present study was to examine the unique contributions of direct assessments and teacher ratings of children's EF skills at the beginning of prekindergarten (pre-k) to gains in academic achievement over the pre-k year. Data for the current study come from a subsample of children recruited for a large-scale pre-k curriculum intervention. This subsample (n = 719) was restricted to all children who were native English speakers and had at least 1 pretest and posttest score on the assessments. Several important findings emerged. Teacher reports of EF and direct assessments were correlated, particularly when EF direct assessments were modeled as a single component score. When entered into the models simultaneously, *both* teacher ratings and direct assessments significantly predicted academic gains in literacy and mathematics; however, the direct assessments were only marginal in predicting gains in language. EF skills accounted for the largest proportion of variance in mathematics achievement gains. The value of using both types of measures in future research is discussed.

Keywords: executive function, prekindergarten, achievement, teacher ratings, direct assessments

An accumulating body of evidence suggests that young children who exhibit greater self-regulation abilities in early childhood achieve more academically (e.g., Blair & Razza, 2007; Bodovski & Farkas, 2007; Duncan et al., 2007; Li-Grining, Votruba-Drzal, Maldonado-Carreño, & Haas, 2010), have lower rates of hyperactive and disruptive behaviors (e.g., Espy, Sheffield, Wiebe, Clark, & Moehr, 2011; Séguin, Nagin, Assaad, & Tremblay, 2004), and are less likely to commit crimes and engage in delinquent behavior as adolescents or adults (Moffitt et al., 2011). Within the group of studies cited above are ones that capitalized on global ratings of children's self-regulation, including those asking parents and teachers to rate children's self-control, impulsivity, emotion regulation, persistence, and attention (e.g., Duncan et al., 2007; Moffitt et al., 2011). Other researchers, like Blair and Razza (2007), have focused on direct child assessments to capture specific elements of children's self-regulation as they relate to school readiness and academic achievement.

In the current study, we focus on a set of skills within the domain of self-regulation that is typically referred to as executive functioning (EF) or cognitive control skills, including areas such as working memory, inhibitory control, and attention flexibility, and the contributions that both teacher reports and direct assessments of EF make to academic achievement. We examined the associations between children's EF skills and learning in prekindergarten (pre-k). Early childhood is a time when not only are children's EF skills showing rapid improvement (see Carlson, 2005; Garon, Bryson, & Smith, 2008), but also children are beginning to gain exposure to early academic concepts. Hence, this is a key developmental period in which to examine more closely the longitudinal associations between children's EF and early academic skills, providing findings that could inform assessment and intervention efforts in classrooms for young children.

Executive Functioning and Academic Achievement

EF skills include working memory (the ability to keep information active in memory and manipulate it), inhibitory control (the ability to inhibit salient but irrelevant information in favor of relevant information), and attention flexibility (the ability to shift and persist in attention; Garon et al., 2008). One or more of these EF skills have been positively associated with higher academic performance in young children in a variety of studies (e.g., Blair & Razza, 2007; Bodovski & Farkas, 2007; Duncan et al., 2007; Fuhs, Nesbitt, Farran, & Dong, 2014; Li-Grining et al., 2010). It has been hypothesized that stronger EF skills may allow children to meet the demands of an early childhood classroom better by facilitating

This article was published Online First July 28, 2014.

Mary Wagner Fuhs, Department of Psychology, University of Dayton; Dale Clark Farran and Kimberly Turner Nesbitt, Peabody Research Institute, Vanderbilt University.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E090009, awarded to Dale Clark Farran, Mark W. Lipsey, and Sandra J. Wilson. Mary Wagner Fuhs and Kimberly Turner Nesbitt were supported by an Institute of Education Postdoctoral Fellowship (R305B100016), awarded to Dale Clark Farran and Mark W. Lipsey. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education. Special thanks to Deanna Meador and the PRI research team for their management of the data collection for this project.

Correspondence concerning this article should be addressed to Mary Wagner Fuhs, Department of Psychology, University of Dayton, 300 College Park, Dayton, OH 45469. E-mail: mfuhs1@udayton.edu

attention, memory for class rules, and engagement in academic content, all of which may allow them to benefit from an academic environment. EF skills may also work in a more direct way by aiding children's memory for salient information in early mathematics problem solving or increasing their flexibility in maintaining both letter sounds and symbols in memory during early literacy activities.

Much of the previous literature examining associations between EF skills and academic achievement in early childhood does not directly address the extent to which different methodologies for assessing EF skills may address the same construct and will relate to achievement growth. A growing literature focuses on using one or a battery of direct assessments to assess EF, whereas another line of research has addressed the associations of teacher reports of EF and academic skills development. A few studies have included both methodologies. In the following paragraphs, we address prior research on each methodology as well as when they have been used together before addressing gaps in the literature concerning the relative contribution each method makes to our understanding of the association between EF and growth in academic skills in young children.

Teacher Ratings of Executive Functioning Skills

Self-regulation typically serves as an umbrella term that includes cognitive and emotional components (Raver et al., 2012), associated with concepts in the teacher report literature such as "approaches to learning" and "self-control." Much of what we know about associations between self-regulation and academic achievement over time and in older children has derived from the use of these more global teacher report measures. For example, in the Early Childhood Longitudinal Study–Kindergarten, teachers rated children's approaches to learning, which included such behaviors as persistence at tasks, eagerness to learn, attention, learning independence, flexibility, and organization. Teacher ratings of these characteristics predicted mathematics achievement at all grade levels from kindergarten to second grade with the strongest effects for those children whose achievement fell in the bottom quartile (Bodovski, & Farkas, 2007).

Various assessments have been developed to capture self-regulation in young children through parent and/or teacher reports. For example, the Child Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) is considered a temperament scale. Based on Likert-type ratings of young children's emotional and cognitive regulation, the CBQ includes an inhibitory control subscale. This measure is typically defined as an assessment of effortful control, which includes both cognitive and emotion regulation components. Blair and Razza (2007) found that preschool teacher reports of effortful control using the CBQ were related to children's kindergarten mathematics and literacy skills. The coefficients were much weaker for the teacher reports when they were compared to direct child assessments of EF in preschool. The CBQ includes items reflecting both emotion regulation and cognitive regulation, making it difficult to compare directly the contributions of teacher reports of cognitive regulation alone and direct child assessments of targeted EF skills. In addition, Blair and Razza had no preschool measures of achievement; this intriguing study of the associations among these areas does not help us understand the relationship between EF (assessed in different ways) and learning as measured by gains in academic skills across time.

While the CBQ includes both emotion and cognitive regulation, other teacher rating measures focus more specifically on EF skills. A clinically oriented measure, the Behavior Rating Inventory of Executive Function—Preschool Version (BRIEF-P; Gioia, Isquith, Retzlaff, & Espy, 2002), assesses whether children have deficits in particular areas of EF. The BRIEF-P has most commonly been used as a clinical and neuropsychological assessment of executive dysfunction. Other measures such as the Child Behavior Rating Scale (CBRS; Bronson, Tivnan, & Seppanen, 1995) focus more specifically on EF skills as they are manifested in typical classroom behavior in early childhood. Recent work has linked teacher ratings on the CBRS to children's early academic skills development above and beyond a direct child assessment of behavioral regulation (Wanless, McClelland, Acock, Chen, & Chen, 2011). Wanless et al. (2011) focused on the links between EF and achievement across different cultures. The timing of the teacher reports differed such that teachers in the United States did not complete teacher ratings until the middle of the school year. Thus, it is unclear the extent to which the differing timing of the teacher and direct assessments affected their contributions to academic achievement. The CBRS was also used in a kindergarten study that included a direct assessment of EF (Head, Toes, Knees, and Shoulders [HTKS]) as a predictor of achievement (Matthews, Ponitz, & Morrison, 2009); this sample was primarily middle-income and white. While achievement and HTKS were measured at pre- and posttest, teacher CBRS ratings were only collected in the spring and only for about 60% of the sample. Nevertheless, on this subset of children, both spring teacher ratings and fall HTKS scores were related to children's gains in math achievement over the year even when both were in the model.

A differently constructed measure of EF in the classroom is the Work-Related Skills subscale of the Cooper-Farran Behavioral Rating Scale (WRS; Cooper & Farran, 1988, 1991). The CBQ, the BRIEF-P, and the CBRS are similar to each other in that the items in each are rated on a Likert-type scale from, for example, "extremely untrue" to "extremely true" or "often" to "never." The WRS is different in that it assesses children's EF skills in academic learning contexts through the use of behaviorally anchored items related to classroom expectations. For example, one item lists "Listening to Teacher Giving Instructions to Group," and the anchors are "Attends to the teacher without reminders," "Occasionally inattentive; attention is easily regained by a cue from teacher," "Can maintain attending behavior with frequent reminders from the teacher," and "Seems to ignore the teacher; is very distracted and distracting." This type of scale, in contrast to a Likert rating scale, is situationally specific. The "person-situation" debate is a robust one in psychology; a trait approach is useful for predicting behaviors "averaged over many situations, occasions, and responses" (Epstein & O'Brien, 1985, p. 532). In a classroom, however, rating scales can suffer from what has been called a "reference group" problem (Heine, Lehman, Peng, & Greenholtz, 2002). While most clearly evident in cross-cultural work, Heine et al. (2002) argued that the reference group issue applies for any groups that might possibly possess different referents for their ratings, such as teachers. One solution Heine et al. proposed, although not without limitations, is to create items with concrete, objective, response options such as behavioral anchors.

Several studies of classrooms in the United States have demonstrated an association between teacher reports of young children's EF skills assessed by the WRS and their academic achievement (e.g., McClelland, Acock, & Morrison, 2006; McClelland, Morrison, & Holmes, 2000; Speece & Cooper, 1990), but many have focused on the relationship in kindergarten. For example, McClelland et al. (2006) found that children's EF skills, as assessed by the WRS subscale, predicted their academic achievement across domains in kindergarten and continued to predict mathematics and literacy achievement out to second grade after controlling for covariates, including prior achievement. McClelland et al. (2006) also found associations between ratings of EF skills and academic achievement out to sixth grade.

Although the work summarized above points to an association between teacher ratings of children's classroom-related EF and academic achievement, at least two specific questions remain. First, are these associations apparent prior to kindergarten? Children are increasingly likely to experience academic instruction in pre-k classrooms, especially those associated with public schools and a learning agenda. Because the previous research with a more classroom-specific scale has explored these links primarily in kindergarten, it is important to investigate whether teacher reports of children's EF in pre-k classrooms will also capture children's EF skills as they relate to their academic growth. Second, do these associations hold even when accounting for children's performance on direct assessments of EF? Concern is often raised about bias in teacher ratings primarily relating to teacher judgments of behavior and learning problems (e.g., Berg-Nielsen, Solheim, Belsky, & Wichstrom, 2012; Mullola et al., 2012). The assumption is that these ratings will be less valid and reliable than direct assessments of behavior. As direct child measures of EF have been developed and more information has accrued on their reliability and validity, an issue is whether they could substitute for teacher ratings and present equal, or perhaps better, indications of children's learning-related characteristics.

Direct Child Assessments of Executive Function

As previously mentioned, one increasingly common approach to measuring EF skills in young children is to assess them directly with a battery of tasks to tap working memory, inhibitory control, and attention flexibility. For the most part, these tasks were developed first in psychological laboratories. Researchers have found concurrent associations between directly assessed EF skills and children's academic achievement in both literacy and mathematics across different grade levels (e.g., Allan & Lonigan, 2011; Best, Miller, & Naglieri, 2011; Bull, Espy, Wiebe, Sheffield, & Nelson, 2011; Bull & Scerif, 2001; St. Clair-Thompson & Gathercole, 2006). Bull and Scerif (2001) found that several direct assessments of pre-k children's EF were concurrently related to their mathematics skills (see also Bull et al., 2011). In a cross-sectional study, Best et al. (2011) found contemporaneous associations between individually assessed EF measures and achievement at each grade level from age 5 through high school with the strongest, most consistent relationship being with mathematics achievement. Finally, longitudinal research suggests modest correlations between pre-k children's EF skills and their growth in academic skills as well (e.g., Bull, Espy, & Wiebe, 2008; Clark, Pritchard, & Woodward, 2010; Fuhs et al., 2014; McClelland et al., 2007).

Gaps in Extant Literature

While the literature indicates that both direct assessments and teacher ratings of children's EF skills are positively associated with children's academic achievement, several questions remain. First, is there a significant association between teacher reports of EF and direct child assessments in a pre-k sample? McClelland et al. (2007) examined correlations between HTKS as a direct assessment of EF and teacher ratings of children's social and behavioral regulation, but this work has not been extended to the WRS and a broader range of EF direct child assessments.

Second, what are the benefits and unique contributions of these two methods of assessment? A recent review of performance-based measures and ratings of EF found only modest correlations between them when each was used in the same study (Toplak, West, & Stanovich, 2013). The authors concluded that the two types of measures were actually tapping different cognitive levels in the respondent. Direct assessments they argued provide evidence of the individual's available processes, while ratings provide evidence of how those processes may or may not be used in an actual setting. Also, direct assessments provide an understanding of children's EF skills in a controlled or neutral context because they are administered to children individually usually in a quiet space. It is not clear what relationship performance in the controlled setting will have with children's EF skills in an ecological context like a classroom. On the other hand, while teacher ratings of children's behaviors can be particularly beneficial to understand children's EF skills in authentic classroom environments, ratings could be influenced by other aspects of children's abilities and skills besides EF.

Using both types of assessments together in a multimethod approach could yield a more complete understanding of children's EF skills, "providing important and nonredundant information about an individual's efficiency and success in achieving goals" (Toplak et al., 2013, p. 138). Moreover, an understanding of the relation between the two methods of assessing EF is informative for research because (a) utilizing direct child assessments is not always a feasible option for researchers, and (b) teacher reports may not be appropriate as the only means of assessing children's abilities (e.g., in curriculum interventions that focus specifically on developing self regulation).

Current Study

The goal of the present study is to examine the contribution of direct assessments and teacher ratings of children's EF skills at the beginning of pre-k to predict children's gains in academic achievement over the pre-k year. Four research questions were examined: (a) Is children's performance on direct assessments of EF positively correlated with teachers' ratings of their EF skills in the classroom context? (b) Are teacher ratings of children's EF skills at the beginning of pre-k associated with the gains children make in literacy, language, and mathematics across the pre-k year? (c) Is children's performance on direct assessments of EF at the beginning of pre-k associated with the gains they make in literacy, language, and mathematics across the pre-k year? (d) Are direct assessments and teacher ratings of EF, when examined together in a single model, significantly and uniquely associated with children's literacy, language, and mathematics gains?

Method

Participants

Data for the current study are a subsample from a larger sample of children ($N = 1{,}145$) recruited for a large-scale randomized control trial (RCT) to evaluate the effectiveness of the *Tools of the Mind* curriculum (Bodrova & Leong, 2007; Farran, Wilson, & Lipsey, 2013). All assessments were administered in English; therefore, we removed nonnative English speakers ($n = 380$) from the current sample to eliminate confounds due to limited English proficiency. To be consistent with the analytic sample of the RCT, we also removed children who did not have at least one pre- or posttest (primarily due to moving prior to the spring assessments). The analytic sample for the current study, therefore, consisted of 719 English-speaking pre-k students who had at least one complete pretest and posttest measure. There were 695 children ($M_{age} = 54$ months; $SD_{age} = 4$ months) with complete demographic, child assessment, and teacher report data for this study. Children with missing data points did not differ from the analytic sample on any demographic or pretest measure ($p > .05$), and because the cases with missing data constituted less than 5% of the analytic sample, we only used available data for each analysis rather than conducting multiple imputation. Girls were 46% of the sample, and children came from varied racial/ethnic backgrounds (36% Black, 52% White, 5% Hispanic, and 7% other).[1] Sixteen percent of the sample had an individualized education program (IEP), which was included in analyses as a covariate. Although precise socioeconomic status (SES) information was not available due to Family Educational Rights and Privacy Act regulations, all children in this study came from public pre-k programs targeted to low-income families. Therefore, it can be assumed that most, if not all, children in the study were from low-income backgrounds.

Children in the sample were nested in 80 classrooms in 57 schools in six school systems in the Southeastern United States. On average, nine children from each classroom were in the analysis sample. The average number of years of experience teaching pre-k for the teachers in the study was six. Because these data were drawn from an RCT, 32 classrooms were assigned to the *Tools of the Mind* condition and 28 classrooms were assigned to "business as usual," which involved a variety of curricula, but primarily *Creative Curriculum*, *Opening the World of Learning*, and *Building Blocks*. All analyses of main effects for curriculum on individual academic achievement measures and possible interactions between curriculum and demographic characteristics or children's pretests were nonsignificant in the RCT (Farran et al., 2013). Nonetheless, we included condition as a control variable in all of the present analyses as the experimental curriculum could potentially account for variance in academic achievement and EF in this sample.

Measures

Teacher reports. Children's classroom-specific EF skills as well as their more general social skills were assessed using the *Cooper-Farran Behavioral Rating Scale* (CFBRS; Cooper & Farran, 1988, 1991). The CRBRS is an assessment of young children's behaviors at school entry and consists of two subscales: Work-Related Skills (WRS) and Interpersonal Skills (IPS). The WRS subscale rates children's EF skills as they are manifested in the classroom. The WRS consists of 16 items related to children's independent work, compliance and memory for instructions, and ability to complete tasks. The IPS subscale rates children's social skills. The IPS consists of 21 items related to children's ability to engage effectively in interactions with peers and teachers. All CFBRS are items are rated from 1 to 7 using behavioral anchors distinctive to each odd-numbered item. As reported in the manual (Cooper & Farran, 1991), the test–retest reliability after an 8-week delay for the WRS and IPS subscales were .66 and .69, respectively. Interrater reliability on the measure has also been established between teacher and teacher aids; intraclass correlations between the raters were .79 and .78 for the WRS and IPS subscales. The two subscales also indicate high internal consistency (Chronbach's $\alpha > .94$).

Direct child assessments of EF. In the current study, EF was assessed using multiple measures that were chosen to cover the range of EF skills discussed in early childhood literature (see Garon et al., 2008). We included working memory, inhibitory control, and attention flexibility tasks (for additional information on each task, see https://my.vanderbilt.edu/toolsofthemindevaluation/), although each task naturally also tapped other abilities such as motor skills or language ability. This has commonly been called the "task impurity" problem as EF tasks not only tap non-EF skills but also typically tap more than one EF skill (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). To account for this, we were consistent with the methodology of prior research in this area and used a battery of EF tasks to create a composite score, drawing on the common EF variance shared by individual tasks (e.g., Wiebe, Espy, & Charak, 2008). Previous research with pre-k children appears to support a one-factor model of EF (e.g., Fuhs & Day, 2011; Hughes & Ensor, 2011; Wiebe et al., 2008; although see Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012). We did, however, analyze the EF tasks both as individual tasks as well as a composite score to more fully capture their contributions to academic skills.

Visuo-spatial short-term memory and working memory were assessed using the *Corsi Blocks* task (Berch, Krikorian, & Huha, 1998; Corsi, 1972). We chose a visuospatial task instead of a verbal task because previous research has suggested that young children have great difficulty with verbal working memory tasks such as digit span (e.g., Bull et al., 2008). We also were concerned about the potential confounds of using a digit span task to predict mathematics skills because this task also taps digit familiarity. *Corsi Blocks* required children to point to a series of block patterns (block placement modeled after Berch et al., 1998) that became progressively more difficult by increasing the number of points in the pattern. Children were asked to repeat a pattern exactly as presented (Forward) and then to reverse a presented pattern (Backward). The experimenter tapped the blocks at a rate of approximately one block per second, and children received up to two attempts to successfully complete each span length until they scored incorrectly on both trials for a particular span length. Two

[1] Children's ethnicity was provided to us by schools from parent reports the schools collected. However, the reliability of this self-reported information was unclear, and thus, this variable was not used in analyses.

blocks of forward trials were conducted, followed by two blocks of backward trials.

Children received two practice trials prior to assessments in which the child only had to touch the same block as an experimenter to ensure that the child could perform the basic action required of the task. Then, the child received additional practice trials that were identical to test trials but that were followed by feedback. Following practice, the test trials without feedback began. Some have referred to the forward component of the task as a simple working memory task and the backward component as a complex working memory task (e.g., Garon et al., 2008), whereas others have referred to the forward component as a passive working memory tasks and the backward component as an active working memory task (e.g., Passolunghi & Cornoldi, 2008). Across interpretations, however, *Corsi Blocks* has been conceptualized as a working memory task and has been found to be significantly associated with academic skills in older children and adults, and particularly mathematics skills (see Raghubar, Barnes, & Hecht, 2010 for a review). The forward and backward versions of the task have shown high test–retest reliability in children ages 4 to 11 years (r_s = .83 and .82; Alloway, Gathercole, & Pickering, 2006).

Attention flexibility was measured with the *Dimensional Change Card Sort* (DCCS; Zelazo, 2006). The DCCS required children first to sort a set of cards according to one dimension (color) and then according to another (shape). Children were presented with two boxes, one with a red truck on it, and one with a blue star. Each box had a slit in the top for children to sort cards (e.g., blue trucks and red stars). The experimenter first demonstrated the color game on two trials and then conducted a rule check with the child ("Can you show me where the blue ones go in the color game?" and "Can you show me where the red ones go in the color game?"). Children received up to two trials for each rule check. Children were then instructed on each trial, "If it is a blue one, then put it here [pointing to blue star], but if it is a red one, put it here [pointing to red truck]." Children were given six trials and if they completed at least five of six correct, they moved on to the shape game. In the shape game, the rules were given without demonstration with cards, but children still had two opportunities to correctly complete the rule check. The same pass/fail criteria were used for both the color and shape sorts. If children completed both of these games successfully, children moved onto the advanced sort. In this game, children were asked to sort by color if they were presented with a card with a border around it and to sort by shape if the card had no border. Following both demonstration trials and rule checks, children completed 12 advanced sort trials, with nine out of 12 correct counted as passing.

Again, as with the other games, the rules were repeated on each trial. Using scoring suggested by Zelazo (2006), children received a score of 0 if they were unable to successfully sort five of six trials on the first dimension of color, a 1 if they sorted by the first dimension but did not meet the five-out-of-six cutoff criterion for sorting by the dimension of shape, a 2 if they were successful sorting by shape, and a 3 if they also passed the border sort (i.e., correct on at least nine of the 12 trials). Performance on DCCS has shown moderate test–retest reliability with children 36 to 72 months (r = .44; Müller, Kerns, & Konkin, 2012). Larger scores on the DCCS were interpreted as indicating greater attention flexibility.

Copy Design (Osborne, Butler, & Morris, 1984) required children to copy eight simple geometric shapes of increasing difficulty. Tasks of this type are drawing increasing attention (Cameron et al., 2012; Potter, Mashburn, & Grissmer, 2013). Cameron et al. (2012) described the task as requiring children "to process visual information from an external stimulus, invoke a mental representation, and coordinate motor movements to reproduce the image" (p. 1240). Children had two attempts to successfully draw each shape and each attempt was coded to indicate whether the child successfully replicated a design.

This task was used in the British Longitudinal Study analyzed by Duncan et al. (2007) and was one of the stronger long-term predictors of child outcomes. It was also recently used in a large-scale measurement study of children's cognitive self-regulation development (Lipsey et al., 2014). In this longitudinal measurement study, *Copy Design* was significantly correlated with a battery of cognitive self-regulation assessments and showed both construct validity via confirmatory factor analysis (Fuhs & Turner, 2012) and predictive validity for academic achievement (Lipsey et al., 2014). Interrater reliability for *Copy Design* in this study was established by two independent raters double coding 20% of the measures. The kappa coefficients for the 8 shapes ranged from .66 to 1.00 (M_{kappa} = .79). Each shape was scored 0 if coded as incorrect and 1 if coded as correct. Larger scores on the *Copy Design* were interpreted as indicating greater sustained attention.

Inhibitory control was assessed with *Peg Tapping* (PT; Diamond & Taylor, 1996) and *Head Toes Knees Shoulders* (HTKS; Ponitz, McClelland, Matthews, & Morrison, 2009). PT requires children to tap a peg once when an examiner taps it twice and to tap twice when an examiner taps once. Children completed 16 test trials that were scored 0 for incorrect responses and 1 for correct. If the child could not successfully complete practice trials on PT, he or she scored −1 and did not complete the test trials. Performance on PT has shown high test–retest reliability in 5-year-olds (r = .74; Nampijja et al., 2010).

HTKS is another task primarily assessing inhibitory control, although the task likely also taps children's working memory as directions are not repeated on each trial, attention shifting as the rules change during the game, and gross motor coordination. Children were asked to touch their heads when an examiner says "touch your toes" and to touch their toes when an examiner says "touch your head." If children were successful at inhibiting the prepotent response of behaving consistently with the prompt then two new prompts are added, children were then required to touch their knees when an examiner says "touch your shoulders" and vice versa. Children received up to a total of six practice trials and 20 test trials, and each trial was scored with 0 for an incorrect response, 1 for motion toward the incorrect response but ending with a correct response, and 2 for a correct response. Performance on HTKS has shown high test–retest reliability in 4-year-olds (r = .80; Meador, Turner, Lipsey, & Farran, 2013). Larger scores on both PT and HTKS indicated greater ability to inhibit a prepotent response.

Academic achievement. Academic achievement was assessed by administering seven subscales of the Woodcock Johnson III achievement battery (WJ-III; Woodcock, McGrew, Mather, 2001). Literacy skills were assessed with the *Letter-Word Identification* and *Spelling* subtests. *Letter-Word Identification* measures children's ability to identify and pronounce alphabet letters and read words, while *Spelling* measures children's ability to draw simple shapes and write orally presented letters and words. Language skills were assessed with the *Academic Knowledge*, *Oral Comprehension*, and *Picture Vocabulary* subtests. *Academic Knowledge* tests children's factual knowledge of science, social studies, and the humanities; for young children the subtest mainly consists of labeling and identifying pictures, thus heavily relying on vocabulary. *Oral Comprehension* asks children to complete an orally presented passage by providing the appropriate missing word on the basis of semantic and syntactic cues. *Picture Vocabulary* asks children to name objects presented in pictures; it is a test of nouns or knowing the names of things. Mathematics skills were assessed with the *Applied Problems* and *Quantitative Concepts* subtests. *Applied Problems* measures children's ability to solve numerical and spatial problems accompanied by pictures while *Quantitative Concepts* measures children's understanding of number identification, sequencing, shapes, and symbols and in a separate section to manipulate the number line. All analyses were conducted using the item response theory-scaled W-Scores. Standard scores normed with a mean of 100 (SD = 15) can be more interpretable for descriptive purposes and are therefore presented in addition to the W score means in Table 1.

Procedure

Teacher reports were completed in the fall after children had acclimated to the classrooms, about 4–6 weeks. The ratings were collected close to the same time period in which children completed direct assessments. All direct assessments of children were conducted in a quiet area of the building in which they had their prekindergarten program. Data from children's EF assessments in the fall and their academic achievement in both fall and spring were used in analyses. Assessments were administered in a fixed order at each time point. In one child testing session, children completed PT, HTKS, and *Copy Design*, followed by the WJ-III *Oral Comprehension, Applied Problems, Quantitative Concepts,* and *Picture Vocabulary* subtests. The other testing session consisted of the DCCS and *Corsi Blocks*, followed by the WJ-III *Letter-Word Identification, Academic Knowledge,* and *Spelling* subtests. The average interval between fall and spring sessions was 7.38 months (SD = 0.55 months).

Analytic Approach

A series of multilevel models (children nested within classrooms, schools, and systems) was conducted to examine the associations between children's EF and academic achievement gains, using different methodologies to assess EF (teacher report and behavioral assessment). All predictors of interest and outcomes were included as standardized variables so that the parameter estimates could be compared across models. Prior to running conditional models, we first ran a fully unconditional model to determine the percentage of variance in academic achievement outcomes accounted for by the classroom, school, and system levels of our model. We then proceeded to run conditional models to test the associations between EF direct assessments and teacher

Table 1
Descriptive Statistics

Variable	N	M	SD	W score M	W score SD	Skew	t
Corsi Forward T1	717	2.52	1.20			−0.61	
Corsi Backward T1	716	1.17	1.15			0.24	
DCCS T1	717	1.38	0.62			−0.14	
Copy Design T1	717	0.91	1.44			2.30	
HTKS T1	717	11.56	13.71			1.13	
Peg Tapping T1	717	5.31	5.86			0.37	
Work-Related Skills T1	716	4.60	1.13			−0.29	
Interpersonal Skills T1	716	5.18	1.06			−0.81	
Letter Word T1	716	93.23	12.21	318.66	24.57	0.09	
Letter Word T2	700	100.13	11.12	347.91	22.38	−0.23	37.32**
Spelling T1	716	79.15	12.55	336.40	23.52	−0.14	
Spelling T2	700	86.20	15.04	369.31	26.94	−0.26	36.24**
Academic Knowledge T1	716	91.31	12.61	436.21	15.82	−0.62	
Academic Knowledge T2	700	97.15	11.57	449.04	13.34	−0.66	29.20**
Oral Comprehension T1	717	94.55	11.15	445.39	13.42	−0.18	
Oral Comprehension T2	703	99.03	11.65	456.33	13.13	−0.15	27.13**
Picture Vocabulary T1	717	100.67	11.40	462.09	12.78	−2.30	
Picture Vocabulary T2	703	101.03	9.63	468.77	9.93	−2.47	16.93**
Applied Problems T1	717	96.69	12.53	390.27	25.10	−0.77	
Applied Problems T2	703	100.45	11.12	411.54	18.77	−0.88	29.90**
Quantitative Concepts T1	717	87.93	10.69	406.13	12.43	0.68	
Quantitative Concepts T2	703	92.97	12.87	422.50	14.60	−0.03	40.46**

Note. T1/T2 = Time 1/Time 2; DCCS = Dimensional Change Card Sort (Zelazo, 2006); HTKS = Head Toes Knees Shoulders (Ponitz, McClelland, Matthews, & Morrison, 2009). t statistics reported are from independent-samples t tests comparing assessment performance at T2 to T1. For Woodcock Johnson III achievement battery (WJ-III; Woodcock, McGrew, Mather, 2001), standard scores are reported, but W scores are also reported as they were used for tests of skewness and longitudinal analyses including t tests. Cooper-Farran Behavioral Rating Scale (CFBRS; Cooper & Farran, 1988, 1991) Work-Related Skills and Interpersonal Skills scores were based on average ratings on a 7-point scale.
** $p < .01$.

reports and gains in academic skills in literacy, language, and mathematics. Our predictors of interest were entered as fixed effects. A number of covariates were also entered as fixed effects at the child level including pretest scores, pre-post testing interval, age at pretest, gender, and IEP status. Because these data were taken from a large-scale RCT, condition was included as a fixed effect at the school level, although our primary interest was not in evaluating condition but rather in accounting for potential variance in outcomes due to variations in curricula used in classrooms. All multilevel models were run in IBM SPSS (Version 20 Mixed Models) using restricted maximum-likelihood estimation. A sample model equation for the EF direct assessments and teacher reports entered simultaneously to predict language outcomes is presented in the Appendix.

Results

Descriptive Statistics

Descriptive statistics for all variables are presented in Table 1. EF direct assessment scores and teacher ratings from the fall are presented in the top of the table; it is clear that for all the direct assessments variation among the children was great. Children entered pre-k with quite different EF skills. As indicated by the *t* tests in Table 1, children significantly improved their academic skills, with children making particularly large gains on W scores in WJ-III *Letter-Word Identification, Spelling,* and *Quantitative Concepts* subtests. We also examined each variable for evidence of nonnormality; both *Copy Design* and *Picture Vocabulary* showed evidence of skewness (see Table 1). *Copy Design* was positively skewed, and *Picture Vocabulary* was negatively skewed at both time points. Therefore, we performed log transformations of these variables before entering them in analyses. The log transformation reduced the skewness of *Copy Design* to .958, the skewness of *Picture Vocabulary* at Time 1 to −1.464, and the skewness of *Picture Vocabulary* at Time 2 to –1.057. Transformed variables were used in all subsequent analyses.

Data Reduction for EF and Achievement

Due to the nature and complexity of direct assessments of EF in young children, we are presenting results of our analytic models in two ways: (a) for the individual EF tasks entered simultaneously and (b) for the component EF score. We used principal component analysis (PCA) to extract common variance among the EF tasks (*Corsi Blocks*, DCCS, *Copy Design*, HTKS, and *Peg Tapping*) and saved a component score for one set of analyses. Using eigenvalues of >1 as the criterion to determine the number of components, a one-component PCA solution for EF at Time 1 accounted for 41.58% of the variance in the assessments. Component loadings for the EF measures were all above .50. We saved the component scores as a variable, which was standardized with a mean of 0 and a standard deviation of 1 and used in one set of analyses.

PCAs were also conducted to reduce redundancy in the measurement of academic achievement and to ensure that high correlations among the subtests would not jeopardize model specificity. We saved component scores for literacy (WJ-III Letter-Word Identification and Spelling), language (WJ-III *Academic Knowledge, Oral Comprehension,* and *Picture Vocabulary*), and mathematics (WJ-III *Applied Problems* and *Quantitative Concepts*) at each time point. For literacy, we entered the WJ-III *Letter-Word Identification* and *Spelling* subtests into a PCA analysis, and we found a one-factor solution that accounted for 71.21% of the variance in the measures at Time 1. Again, for this and all other component scores, we saved these scores as variables and used them in analyses. At Time 2, the one-factor PCA solution for literacy with the same two measures accounted for 77.81% of the variance in the assessments. For language, the component score explained 72.24% of the variance at Time 1 and 72.53% of the variance at Time 2. The mathematics PCA produced a component that explained 82.16% of the variance in the assessments at Time 1 and 83.54% of the variance at Time 2. Within each PCA, the loadings of each of the measures onto the component were all above .80.

Correlations

Zero-order correlations among demographics, teacher ratings, the EF component score, individual EF direct assessments, and the achievement composites for both fall and spring are presented in Table 2. All academic achievement component scores, teacher ratings, and EF scores were moderately to strongly correlated with each other. Particularly strong correlations emerged between the EF direct assessments and the mathematics composite score at both time points. The EF direct assessment composite and teacher reports of EF (WRS) were strongly correlated, suggesting they are tapping a similar underlying construct. Interestingly the correlations between the EF composite score and entering achievement were somewhat higher than between teacher WRS ratings and entering academic skills, suggesting teachers were rating observed classroom behavior and not just children's skill levels. The correlations were notably weaker between teacher reports of social skills (IPS) and both direct assessments of EF and academic achievement.

Unconditional Multilevel Models

We first ran unconditional models to determine the percentage of variance in academic achievement gains in literacy, language, and mathematics that could be accounted for by the nesting levels of classroom, school, and system. It was especially important to establish the percentage of variance in academic achievement gains that could be accounted for by child-level differences as our predictors of interest were at the child level. If we found, for example, that the largest percentage of variance in achievement outcomes was at the classroom level, we would have little variance left to be explained by child-level predictors. Because our focus was on explaining children's pre-k gains in academic achievement, unconditional models included the covariates of gender, age, experimental condition, IEP status, interval of time that elapsed between pre- and posttest, and pretest achievement scores.

Based on the random parameters reported in the first columns of Tables 3, 4, and 5 (labeled "Unconditional Model"), 89.3% of variance in literacy outcomes, 94.1% of variance in language outcomes, and 92.7% of variance in mathematics outcomes (note that values were rounded) was attributed to child-level differences and could be modeled with our child-level predictors of interest, namely, EF direct assessments and teacher reports. Although the

Table 2
Correlations Among All Study Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Corsi Forward	—														
2. Corsi Backward	.264**	—													
3. DCCS	.284**	.213**	—												
4. Copy Design	.280**	.199**	.176**	—											
5. HTKS	.334**	.235**	.300**	.249**	—										
6. Peg Tapping	.411**	.245**	.336**	.322**	.519**	—									
7. EF Factor	.680**	.518**	.586**	.554**	.717**	.775**	—								
8. WRS Rating	.384**	.257**	.291**	.271**	.350**	.393**	.511**	—							
9. IPS Rating	.196**	.092*	.137**	.074*	.175**	.195**	.232**	.649**	—						
10. Literacy T1	.413**	.231**	.289**	.485**	.351**	.445**	.575**	.413**	.146**	—					
11. Language T1	.393**	.284**	.422**	.250**	.491**	.498**	.617**	.458**	.188**	.492**	—				
12. Math T1	.447**	.324**	.398**	.378**	.523**	.591**	.701**	.493**	.200**	.630**	.713**	—			
13. Literacy T2	.358**	.187**	.236**	.396**	.235**	.349**	.458**	.431**	.171**	.667**	.433**	.542**	—		
14. Language T2	.355**	.275**	.375**	.198**	.417**	.440**	.546**	.462**	.191**	.429**	.810**	.632**	.447**	—	
15. Math T2	.471**	.275**	.381**	.348**	.444**	.528**	.645**	.486**	.190**	.612**	.634**	.755**	.653**	.653**	—
16. Gender	−.044	−.089*	−.111**	−.118**	−.080*	−.016	−.111**	−.227**	−.191**	−.144**	−.066	−.075*	−.214**	−.031	−.051
17. Age	.220**	.108**	.060	.303**	.195**	.227**	.290**	.162**	−.013	.283**	.201**	.294**	.156**	.126**	.187**
18. Condition	.024	−.029	.029	.037	.013	.011	.024	−.058	−.045	.065	.027	.025	.023	−.035	−.013
19. IEP Status	−.177**	−.070	−.112**	−.053	−.093*	−.150**	−.176**	−.222**	−.106**	−.117**	−.228**	−.170**	−.128**	−.200**	−.147**
20. Pre-Post Interval	.074	.061	.114**	.023	.067	.085*	.111**	.094*	.066	.045	.108**	.074	−.003	.166**	.130**

Note. DCCS = Dimensional Change Card Sort (Zelazo, 2006); HTKS = Head Toes Knees Shoulders (Ponitz, McClelland, Matthews, & Morrison, 2009); EF = executive functioning; WRS = Work-Related Skills subscale of the Cooper-Farran Behavioral Rating Scale (Cooper & Farran, 1988, 1991); IPS = Interpersonal Skills subscale of the Cooper-Farran Behavioral Rating Scale; IEP = individualized education program; T1/T2 = Time 1/Time 2.
* $p < .05$. ** $p < .01$.

Table 3

EF Direct Assessments and Teacher Reports Predict End of Prekindergarten Literacy Skills

	Unconditional model		Model 1		Model 2		Model 3		Model 4		Model 5	
Variable	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Fixed Parameters												
Intercept	0.01	0.08	0.00	0.09	0.00	0.09	−0.01	0.08	−0.01	0.09	−0.02	0.09
Gender	-0.12^{**}	0.03	-0.12^{**}	0.03	-0.12^{**}	0.03	-0.09^{**}	0.03	-0.09^{**}	0.03	-0.09^{**}	0.03
Age	−0.04	0.03	-0.08^{**}	0.03	-0.07^{*}	0.03	$-0.06^{\dagger}$	0.03	-0.08^{**}	0.03	-0.07^{*}	0.03
Curriculum Condition	−0.04	0.07	−0.05	0.07	−0.04	0.07	−0.01	0.07	−0.02	0.07	−0.01	0.07
IEP Status	−0.04	0.03	−0.02	0.03	−0.02	0.03	−0.01	0.03	0.00	0.03	−0.01	0.03
Pre-Post Interval	−0.02	0.03	−0.03	0.04	−0.03	0.04	−0.04	0.04	−0.04	0.04	−0.04	0.04
Pretest	0.66^{**}	0.03	0.57^{**}	0.03	0.59^{**}	0.03	0.60^{**}	0.03	0.54^{**}	0.03	0.57^{**}	0.03
Corsi Forward			0.09^{**}	0.03					0.07^{*}	0.03		
Corsi Backward			0.01	0.03					−0.01	0.03		
DCCS			0.04	0.03					0.03	0.03		
HTKS			−0.02	0.03					−0.03	0.03		
Peg Tapping			$0.07^{\dagger}$	0.04					0.05	0.04		
Copy Design			0.09^{**}	0.03					0.09^{**}	0.03		
EF Composite Score					0.16^{**}	0.04					0.10^{**}	0.04
EF Teacher Report							0.18^{**}	0.03	0.15^{**}	0.03	0.15^{**}	0.03
Random Parameters												
Child	0.49^{**}	0.03	0.47^{**}	0.03	0.48^{**}	0.03	0.47^{**}	0.03	0.46^{**}	0.03	0.46^{**}	0.03
Classroom	0.03	0.02	0.02	0.03	0.03	0.04	0.03	0.04	0.02	0.04	0.03	0.04
School	0.00	0.00	0.01	0.04	0.00	0.04	0.01	0.05	0.01	0.04	0.01	0.05
System	0.02	0.02	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.03	0.03	0.03
Psuedo-R^2			0.04		0.03		0.05		0.06		0.06	

Note. EF = executive functioning; IEP = individualized education plan; DCCS = Dimensional Change Card Sort (Zelazo, 2006); HTKS = Head Toes Knees Shoulders (Ponitz, McClelland, Matthews, & Morrison, 2009). Pseudo-R^2 estimates indicate the amount of within-child variability in end of prekindergarten literacy skills explained by the addition of EF measures to the unconditional model.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

percentage of variance accounted for by the different levels of the model varied across academic content areas we decided to account for all levels in all academic content area analytic models to maintain consistency across models and to aid in comparison of effects.

Conditional Multilevel Models

In Tables 3–5, we present the associations between children's fall EF scores and their spring academic achievement in the areas of literacy, language and mathematics after controlling for demographic covariates and children's academic achievement in the fall, in other words the gain in achievement related to initial EF scores. Each table focuses on a different academic area and each contains five models. The first model in each table is the model with the EF direct assessments entered individually into the model to predict gains in academic achievement. The second model (Model 2), shows results when the EF composite score was entered alone. Model 3 in each table depicts the results when the WRS ratings are entered; Model 4 shows the individual direct assessment scores with the addition of the teacher ratings. Finally in each table, Model 5 examines the joint contribution of the EF direct assessment composite and the teacher ratings of EF in predicting gains in achievement across the year.

Multilevel models for EF direct assessments. First we discuss the results from Models 1 and 2 examining the effects for the individual direct assessments entered simultaneously versus the composite score alone. For literacy gains (Table 3), *Corsi Forward*, *Peg Tapping*, and *Copy Design* were significant or marginal predictors. For language gains shown in Table 4, none of the individual EF measures significantly predicted growth except for a marginal effect for *Corsi Backward*. The individual significant EF predictors of mathematics gains as shown in Table 5 were *Corsi Forward, Peg Tapping, Copy Design and DCCS*. HTKS did not predict any achievement content area gains. When examining variance accounted for by adding the group of EF assessments to the conditional model, the pseudo-R^2 estimate of effect size was largest for the mathematics content area.

Model 2 in Tables 3–5 presents the results when EF skills are entered as a PCA composite score. The composite EF score was predictive of children's literacy, language, and mathematics skills in the spring with fall pretest scores entered as a covariate. Again, the pseudo-R^2 estimate of effect size was the largest for mathematics. Thus, Models 1 and 2 indicate that direct assessments of EF both individually and as a composite related to achievement gains over and above children's entering skill levels, but the magnitude of effects varied both by individual EF assessment and by academic content area.

Multilevel models for EF teacher reports. Model 3 tested the association between teachers' ratings of children's EF and spring academic achievement in the areas of literacy, language, and mathematics after controlling for demographic covariates and children's academic achievement in the fall. After accounting for covariates and academic pretest skills, Model 3 in each of Tables 3–5 demonstrates that teachers' fall reports of EF significantly predicted literacy, language, and mathematics outcomes, with the pseudo-R^2 estimate of effects size for literacy achievement being

Table 4
EF Direct Assessments and Teacher Reports Predict End of Prekindergarten Language Skills

Variable	Unconditional model β	Unconditional model *SE*	Model 1 β	Model 1 *SE*	Model 2 β	Model 2 *SE*	Model 3 β	Model 3 *SE*	Model 4 β	Model 4 *SE*	Model 5 β	Model 5 *SE*
Fixed Parameters												
Intercept	0.07	0.04	0.07	0.04	0.07†	0.04	0.06	0.04	0.06	0.04	0.06	0.04
Gender	0.02	0.02	0.03	0.02	0.03	0.02	0.05*	0.02	0.05*	0.02	0.05*	0.02
Age	−0.03	0.02	−0.04	0.02	−0.05	0.03	−0.04†	0.02	−0.04†	0.02	−0.05*	0.02
Curriculum Condition	−0.11*	0.05	−0.11*	0.05	−0.12	0.06	−0.09†	0.05	−0.10†	0.05	−0.10†	0.05
IEP Status	−0.03	0.02	−0.02	0.02	−0.02	0.02	−0.01	0.02	−0.01	0.02	−0.01	0.02
Pre-Post Interval	0.07**	0.03	0.06*	0.03	0.06*	0.03	0.06*	0.02	0.06*	0.03	0.06*	0.03
Pretest	0.80**	0.02	0.75**	0.03	0.75**	0.03	0.75**	0.03	0.73**	0.03	0.73**	0.03
Corsi Forward			0.03	0.03					0.01	0.03		
Corsi Backward			0.04†	0.02					0.03	0.02		
DCCS			0.03	0.02					0.03	0.02		
HTKS			0.01	0.03					0.01	0.03		
Peg Tapping			0.03	0.03					0.02	0.03		
Copy Design			−0.01	0.02					−0.01	0.02		
EF Composite Score					0.09**	0.03					.05†	0.03
EF Teacher Report							0.10**	0.04	0.10**	0.03	.10**	0.03
Random Parameters												
Child	0.32**	0.02	0.32**	0.02	0.32**	0.02	0.31**	0.02	0.31**	0.02	0.32**	0.02
Classroom	0.02†	0.01	0.02†	0.01	0.02†	0.01	0.01	0.01	0.01†	0.01	0.01†	0.01
School	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
System	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Psuedo-R^2			0.01		0.01		0.02		0.02		0.02	

Note. EF = executive functioning; IEP = individualized education plan; DCCS = Dimensional Change Card Sort (Zelazo, 2006); HTKS = Head Toes Knees Shoulders (Ponitz, McClelland, Matthews, & Morrison, 2009). Psuedo-R^2 estimates indicate the amount of within-child variability in end of prekindegarten language skills explained by the addition of EF measures to the unconditional model.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

the largest. Therefore, although EF direct assessments accounted for the most variance in mathematics achievement gains, EF teacher reports accounted for the most variance in literacy achievement gains.

Multilevel models for EF direct assessments and teacher reports together. Finally, we assessed the unique contributions of direct assessments of EF, individually and as a composite, and teacher reports of EF when entered into a model simultaneously (see Tables 3–5, Models 4 and 5). We again ran two separate models, one in which EF direct assessments were entered individually with the teacher ratings and one in which the EF direct assessments composite score was entered with teacher ratings but without the individual EF assessments. After controlling for covariates and academic pretests, teacher reports remained significant predictors of academic achievement when entered simultaneously with EF direct child assessments.

As shown in Table 3, Model 4, for literacy gains both teacher ratings and EF assessments of *Corsi Forward* and *Copy Design* continued to be significant predictors of literacy outcomes. Similarly both teacher ratings and the EF component score were uniquely related to literacy gains (Model 5). For language outcomes presented in Table 4, none of the individual EF measures was significantly associated with outcomes, but Model 5 shows that the EF component score was a marginal predictor when included in the model with teacher reports. Table 5 shows that *Corsi Forward, DCCS, Peg Tapping, Copy Design* (Model 4), and the EF component score (Model 5) were significant or marginal predictors of mathematics outcomes along with teacher ratings.

To summarize, Models 4 and 5 indicate that both EF methods (teacher reports and direct assessments, whether entered individually or as a composite) were significant predictors of children's achievement outcomes in language, literacy, and mathematics. The standardized coefficients for teacher reports were larger for language and literacy, whereas the coefficients for direct assessments were larger for mathematics. The magnitude of the EF effect when both assessment types were entered simultaneously was largest for mathematics outcomes.

Multilevel models for interpersonal skills. Despite the fact that our analyses related to *gains* in achievement by including pretest in the models, one could question whether teacher ratings reflected a general favorable bias toward higher achieving children. If so, that bias should have been reflected in all the ratings teachers provided, both the ones related to EF and the ones related to social interactions. Therefore, we also assessed whether teachers' fall ratings of their children from the other CFBRS scale, the IPS subscale, also accounted for unique variation in academic achievement gains across the pre-k year above and beyond variance accounted for by children's EF direct assessment as well as covariates including the academic skills pretests. Teacher ratings of children's interpersonal skills were not predictive of their academic achievement gains above and beyond the EF direct assessments when entered individually for literacy (β = .04, *SE* = .03, *p* = .171), language (β = .03, *SE* = .02, *p* = .226), or mathematics (β = .003, *SE* = .02, *p* = .901). Teacher ratings of children's interpersonal skills were also nonsignificant predictors of academic achievement gains when entered into models with the EF

Table 5
EF Direct Assessments and Teacher Reports Predict End of Prekindegarten Mathematics Skills

Variable	Unconditional model		Model 1		Model 2		Model 3		Model 4		Model 5	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Fixed Parameters												
Intercept	0.05	0.06	0.05	0.05	0.05	0.05	0.04	0.06	0.04	0.05	0.04	0.05
Gender	0.00	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.04	0.02	0.04	0.02
Age	−0.02	0.03	−0.06*	0.03	−0.05*	0.03	−0.03	0.03	−0.06*	0.03	−0.06*	0.03
Curriculum Condition	−0.06	0.06	−0.08	0.06	−0.07	0.06	−0.04	0.06	−0.06	0.06	−0.06	0.06
IEP Status	−0.02	0.03	0.00	0.02	−0.01	0.02	0.00	0.02	−0.01	0.02	0.01	0.02
Pre-Post Interval	0.05	0.03	0.03	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.03
Pretest	0.74**	0.03	0.61**	0.03	0.60**	0.03	0.67**	0.06	0.60**	0.03	0.57**	0.03
Corsi Forward			0.12**	0.03					0.11**	0.03		
Corsi Backward			0.00	0.03					−0.01	0.02		
DCCS			0.07*	0.03					0.06*	0.03		
HTKS			0.02	0.03					0.02	0.03		
Peg Tapping			0.06†	0.03					0.05†	0.03		
Copy Design			0.07*	0.03					0.06*	0.03		
EF Composite Score					0.23**	0.03					0.20**	0.03
EF Teacher Report							0.14**	0.03	0.10**	0.03	0.10**	0.03
Random Parameters												
Child	0.39**	0.02	0.36**	0.02	0.36**	0.02	0.37**	0.02	0.36**	0.02	0.36**	0.02
Classroom	0.03*	0.01	0.02	0.02	0.03*	0.01	0.03*	0.01	0.02	0.02	0.03*	0.01
School	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
System	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01
Psuedo-R^2			0.06		0.05		0.03		0.07		0.07	

Note. EF = executive functioning; IEP = individualized education plan; DCCS = Dimensional Change Card Sort (Zelazo, 2006); HTKS = Head Toes Knees Shoulders (Ponitz, McClelland, Matthews, & Morrison, 2009). Pseudo-R^2 estimates indicate the amount of within-child variability in end of prekindergarten mathematics skills explained by the addition of EF measures to the unconditional model.
† $p < .10$. * $p < .05$. ** $p < .01$.

composite score—literacy (β = .04, *SE* = .03, *p* = .193), language (β = .03, *SE* = .02, *p* = .221), mathematics (β = .01, *SE* = .03, *p* = .829).

Discussion

In the present study, we compared the unique contributions of teacher reports of EF and direct child assessments of EF when added into a model simultaneously to predict gains in academic skills across the pre-k year. Three important findings emerged. First, children's EF skills both as rated by teachers and as observed in direct child assessments were significantly related to each other. Second, when entered in separate models, direct assessments and teacher reports of children's EF skills at school-entry were significantly related to their academic gains in literacy, language, and mathematics in pre-k above and beyond covariates. Third, both teacher reports of children's EF and direct assessments of EF remained significant predictors of literacy and mathematic gains even when both were entered into a model simultaneously. Direct assessments of EF were only marginally associated with language gains when entered into the model simultaneously with teacher reports.

Teacher-reported EF was strongly related to behavioral assessments of children's EF, which suggests that they may be tapping similar if not identical underlying characteristics; our correlations were stronger than those summarized by Toplak et al. (2013) in clinical populations. In fact, as can be seen in the simple correlations, in some cases the teacher reports of EF were more highly correlated with individual EF direct assessments than some of the correlations among EF direct child assessments themselves. This provides further evidence that these two different methodologies may be tapping into similar skill sets in young children and also speaks to the measurement issues with individual direct assessments. Although previous research has estimated correlations between the HTKS and teacher reports of behavioral regulation in the early childhood classroom (McClelland et al., 2007), the current study extends this research by creating a component score of EF that draws upon the common variance among these measures, and associating it with teacher reports of specific EF behaviors in the classroom. The correlations obtained in the current study were much higher than those obtained by McClelland et al. (2007) using only the HTKS, suggesting the possibility that when the variance unique to each individual assessment is removed, the overlap between teacher reports and child direct assessments of EF may be higher than previously reported. However, the correlations between teacher reports of EF and direct child assessments were not so high as to suggest complete redundancy in measurement. It is quite possible that the teacher report may tap skills that are not being tapped by direct child assessments, lending support to the idea that both methodologies may yield important information about young children's EF skills. Direct assessments may capture children's available cognitive processes and teacher reports may assess how these processes are used in a real-world setting.

We found positive associations between children's entering EF skills, assessed through direct child assessments and teacher reports, and their *gains* in literacy, language, and mathematics skills. When entered into models separately, the effect size estimates of the unique variance accounted for by EF (above and beyond covariates and pretests) were larger for the model of teacher reports predicting literacy and language compared to the models including direct assessments This was not the case for mathematics achievement gains, as the addition of direct assessments of EF to the model accounted for a larger proportion of variance compared to the addition of teacher reports. When examining teacher reports and direct assessments entered together as predictors, effects differed such that after controlling for covariates and pretests, the strongest effect of children's school-entry EF skills was observed for gains in mathematics achievement. Effect sizes were smaller across the board for the variance accounted for by EF above and beyond language pretests and covariates.

While the pseudo-R^2 estimate of effects size indicated that EF skills accounted for 7% of the variance in gains in children's mathematics achievement, it is necessary to interpret the magnitude of this effect based on its practical value for a given field (Cumming, 2014). In the present case, the practical significance of the effects must be interpreted in light of explaining unique variance in *gains* in children's academic achievement *beyond* that which can be explained by initial achievement as well as covariates. Such a conservative approach suggests that even for smaller pseudo-R^2 estimates, the effects remain meaningful for practice.

Several researchers have previously posited that mathematics skills uniquely tax children's inhibitory control, attentional flexibility, and working memory (Blair, Knipe, & Gamson, 2008). Specifically, recent accounts suggest that although children may initially heavily recruit EF resources for academic learning across domains, certain skills, such as literacy, may become more automatic requiring less higher order problem solving compared to mathematics tasks, which likely only increase in their cognitive demands as new mathematical concepts are learned (e.g., Blair et al., 2008; Welsh et al., 2010). Our results are consistent with this account when examining the variance in mathematics achievement gains accounted for by both EF direct child assessments and teacher report simultaneously, such that the effect size values were the largest for mathematics models. However, we found somewhat smaller but still important predictions from EF measures to gains in literacy, suggesting that during the pre-k year at least, these skills are not as automatic as they will become in kindergarten and first grade.

When examining direct child assessments and teacher reports separately we found that teacher reports of EF accounted for more variance in literacy and language achievement gains than did EF direct assessments examined alone. Conversely, the model with EF direct assessments alone accounted for more variance in mathematics achievement gains compared to the model with EF teacher reports alone. Prior work on EF and achievement is largely based on the use of direct child assessments; the current study results suggest that the addition of teacher reports will yield a more comprehensive picture, at least in pre-k. It could be that previous findings were limited by the use of measures of only one methodology, and perhaps that literacy, language, and mathematics skills may all be influenced by EF skills in the pre-k year but in different ways. For example, it could be the case that literacy and language skills are more affected by the types of skills tapped by teacher reports, namely, how children use their EF skills in classroom learning, whereas mathematics skills are more directly affected by the cognitive processes themselves. Future research is necessary to examine these effects beyond pre-k and into early elementary school to determine if differential patterns emerge.

The difference in magnitude of effects of EF skills on literacy and mathematics gains compared to language gains is not necessarily surprising considering the nature of pre-k instruction in which literacy and mathematics skills receive more explicit attention compared to language skills. Thus, the benefits of having greater EF skills that allow children to attend and remain engaged during classroom instruction may be more relevant for early literacy and mathematics skills compared to language skills. Also, in this particular data set, children made substantially less gain in the language across the pre-k year, and the correlations from pre- to posttest were very high, suggesting strong stability and less intraindividual variability to be explained by EF measures.

It is worth commenting on the fact that we did not find HTKS to be a unique predictor of gains in any of the three academic areas, which is in contrast to several recent studies with this measure (e.g., McClelland et al., 2007; Wanless et al., 2011). A big difference in our study compared to the ones cited above is that we used several direct assessment measures and not just HTKS alone. It is apparent from Table 2 and the zero-order correlations that Peg Tapping and HTKS were the most highly correlated among the direct assessment measures (r = .52). Peg Taping generally correlated more highly with the other direct assessments than HTKS did. It is possible that alone HTKS is an important contributor to achievement gains, but its contribution was swamped by the stronger relations between gains and some of the other EF direct assessment measures.

Taken together, these findings illustrate the potential for ecologically valid teacher ratings of EF skills in combination with direct assessments, especially with regard to examining the interrelations between EF skills and academic achievement in field-based research. Importantly, the same pattern of associations was not observed for teacher reports of children's social skills, suggesting that teachers were not simply more positively predisposed to some children than others and also as Duncan et al. (2007) also found, that social skills may be important but perhaps not for academic achievement. Teachers seem to be able to recognize specific types of behaviors in young children that will facilitate their learning in the classroom over the year. Researchers will sometimes use a battery of direct assessments when measuring the various aspects of children's EF but seldom does the battery include teacher reports. Such batteries can require taking children out of their classrooms for one-on-one testing for up to 45 min, which is not always desirable or feasible. Moreover, because traditional direct assessments of EF were initially developed for use in neurological research, many assessments of EF are not situated within the context of typical preschool learning activities. Teacher reports that are easily administered and ecologically valid could considerably enhance our understanding of the interrelations between the development of EF and early academic success, particularly when used in conjunction with direct child assessments.

Limitations

The contributions of this research notwithstanding, study limitations must be acknowledged. While the current findings help us better understand the unique value of various modes of assessing young children's EF skills, the correlational design of the study does not permit causal conclusions regarding the associations between children's EF skills and gains in achievement over the pre-k year. In particular, while we demonstrated that teacher ratings were convergent with traditional direct assessments of EF and explained unique variance in children's literacy, language, and mathematics achievement in conjunction with direct assessments of EF, it is possible that teacher ratings could be capturing other aspects of children's scholastic abilities. It may be that future research could be conducted to further validate the use of teacher ratings by including statistical controls for other aspects of children's scholastic abilities that might be confounded with ratings of EF skills, including general intelligence. Another limitation with the use of the Work-Related Skills subscale of the Cooper-Farran Behavioral Rating Scale is that it does have three items that reference emotion or social context. Thus, it is possible that although the measure primarily assesses EF, it may also capture other additional skills as well. Future work should examine associations between the Work-Related Skills subscale and other assessments of emotion regulation to get a clearer picture of the extent to which the Work-Related Skills subscale might also capture more affectively laden components of self-regulation skills.

Conclusions

Both EF direct assessments and EF teacher reports explained unique variance in children's academic achievement gains in literacy and mathematics. Teacher reports may be an ecologically valid and efficient way of capturing children's EF development in early childhood classrooms when direct assessments are not feasible. If possible, including both teacher ratings and one or more direct assessments would seem to be the best course.

References

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

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77,* 1698–1716.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition, 38,* 317–338. doi:10.1006/brcg.1998.1039

Berg-Nielsen, T., Solheim, E., Belsky, J., & Wichstrom, L. (2012). Preschoolers' psychosocial problems: In the eyes of the beholder? Adding teacher characteristics as determinants of discrepant parent-teacher reports. *Child Psychiatry and Human Development, 43,* 393–413. doi:10.1007/s10578-011-0271-0

Best, J. R., Miller, P. H., & Naglieri, J. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21,* 327–336. doi:10.1016/j.lindif.2011.01.007

Blair, C., Knipe, H., & Gamson, D. (2008). Is there a role for executive functions in the development of mathematics ability? *Mind, Brain, and Education, 2,* 80–89. doi:10.1111/j.1751-228X.2008.00036.x

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

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 108,* 115–130. doi:10.1086/525550

Bodrova, E., & Leong, D. J. (2007). *Tools of the mind: The Vygotskian approach to early childhood education* (2nd ed.). Columbus, OH: Merrill/Prentice Hall.

Bronson, M. B., Tivnan, T., & Seppanen, P. S. (1995). Relations between teacher and classroom activity variables and the classroom behaviors of prekindergarten children in Chapter 1 funded programs. *Journal of Applied Developmental Psychology, 16,* 253–282. doi:10.1016/0193-3973(95)90035-7

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of math achievement at age 7 years. *Developmental Neuropsychology, 33,* 205–228. doi:10.1080/87565640801982312

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: Sources of variation in emergent math achievement. *Developmental Science, 14,* 679–692. doi:10.1111/j.1467-7687.2010.01012.x

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's math ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273–293. doi:10.1207/S15326942DN1903_3

Cameron, C., Brock, L., Murrah, W., Bell, L., Worzalla, S., Grissmer, D., & Morrison, F. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development, 83,* 1229–1244. doi:10.1111/j.1467-8624.2012.01768.x

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

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46,* 1176–1191. doi:10.1037/a0019672

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk in kindergarten. *Early Childhood Research Quarterly, 3,* 1–19. doi:10.1016/0885-2006(88)90026-9

Cooper, D. H., & Farran, D. C. (1991). *The Cooper-Farran Behavioral Rating Scales.* Brandon, VT: Clinical Psychology.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Doctoral dissertation, McGill University). Retrieved from http://digitool.Library.McGill.CA:80/R/-?func=dbin-jump-full&object_id=93903&silo_library=GEN01

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. doi:10.1177/0956797613504966

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the ability to remember what I said and to "do as I say, not as I do." *Developmental Psychobiology, 29,* 315–334.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446. doi:10.1037/0012-1649.43.6.1428

Epstein, S., & O'Brien, E. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin, 98,* 513–537. doi:10.1037/0033-2909.98.3.513

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

Farran, D. C., Wilson, S. J., & Lipsey, M. W. (2013, March). *Effects of a curricular attempt to improve self-regulation and achievement in prekindergarten children*. Paper presented at the 2013 Society for Research in Child Development Conference, Seattle, WA.

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

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic achievement across content areas. *Developmental Psychology*. Advance online publication. doi:10.1037/a0036633

Fuhs, M. W., & Turner, K. A. (2012, February). *Evaluating group and longitudinal measurement equivalence in a battery of cognitive self-regulation measures for preschoolers*. Poster presented at the Society for Research in Child Development 2012 themed meeting: Developmental methodology, Tampa, FL.

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

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology, 8,* 249–257. doi:10.1076/chin.8.4.249.13513

Heine, S., Lehman, D., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology, 82,* 903–918. doi:10.1037/0022-3514.82.6.903

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology, 108,* 663–676. doi:10.1016/j.jecp.2010.06.005

Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology, 46,* 1062–1077. doi:10.1037/a0020066

Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., & Wilson, S. J. (2014). *Learning-related cognitive self-regulation school readiness measures for preschool children: Optimizing predictive validity for achievement*. Manuscript submitted for publication.

Matthews, J. S., Ponitz, C., & Morrison, F. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101,* 689–704. doi:10.1037/a0014240

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21,* 471–490. doi:10.1016/j.ecresq.2006.09.003

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43,* 947–959. doi:10.1037/0012-1649.43.4.947

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly, 15,* 307–329. doi:10.1016/S0885-2006(00)00069-7

Meador, D. N., Turner, K. A., Lipsey, M. W., & Farran, D. C. (2013). *Administering measures from the PRI learning-related cognitive self-regulation study*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

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

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (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., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108,* 2693–2698. doi:10.1073/pnas.1010076108

Müller, U., Kerns, K. A., & Konkin, K. (2012). Test–retest reliability and practice effects of executive function tasks in preschool children. *The Clinical Neuropsychologist, 26,* 271–287. doi:10.1080/13854046.2011.645558

Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., & Keltikangas-Jävinen, L. (2012). Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability. *British Journal of Educational Psychology, 82,* 185–206. doi:10.1111/j.2044-8279.2010.02017.x

Nampijja, M., Apule, B., Lule, S., Akurut, H., Muhangi, L., Elliott, A. M., & Alcock, K. J. (2010). Adaptation of Western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *British Journal of Educational Psychology, 80,* 15–30. doi:10.1348/000709909X460600

Osborne, A. F., Butler, N. R., & Morris, A. C. (1984). *The social life of Britain's five year olds: A report of the Child Health and Education Study*. London, England: Routledge and Kegan Paul.

Passolunghi, M. C., & Cornoldi, C. (2008). Working memory failures in children with arithmetical difficulties. *Child Neuropsychology, 14,* 387–400. doi:10.1080/09297040701566662

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral regulation and its contributions to kindergarten outcomes. *Developmental Psychology, 45,* 605–619. doi:10.1037/a0015365

Potter, D., Mashburn, A., & Grissmer, D. (2013). The family, neuroscience, and academic skills: An interdisciplinary account of social class gaps in children's test scores. *Social Science Research, 42,* 446–464. doi:10.1016/j.ssresearch.2012.09.009

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20,* 110–122. doi:10.1016/j.lindif.2009.10.005

Raver, C. C., Carter, J. S., McCoy, D. C., Roy, A., Ursache, A., & Friedman, A. (2012). Testing models of children's self-regulation within educational contexts: Implications for measurement. *Advances in Child Development and Behavior, 42,* 245–270. doi:10.1016/B978-0-12-394388-0.00007-1

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

Séguin, J. R., Nagin, D., Assaad, J.-M., & Tremblay, R. E. (2004). Cognitive-neuropsychological function in chronic physical aggression and hyperactivity. *Journal of Abnormal Psychology, 113,* 603–613. doi:10.1037/0021-843X.113.4.603

Speece, D. L., & Cooper, D. H. (1990). Ontogeny of school failure: Classification of first-grade children. *American Educational Research Journal, 27,* 119–140. doi:10.3102/00028312027001119

St. Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59,* 745–759. doi:10.1080/17470210500162854

Toplak, M., West, R., & Stanovich, K. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, 54,* 131–143. doi:10.1111/jcpp.12001_2012

Wanless, S. B., McClelland, M. M., Acock, A. C., Chen, F. M., & Chen, J. L. (2011). Behavioral regulation and early academic achievement in Taiwan. *Early Education and Development, 22,* 1–28. doi:10.1080/10409280903493306

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102,* 43–53. doi:10.1037/a0016738

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

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Rolling Meadows, IL: Riverside.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1,* 297–301. doi:10.1038/nprot.2006.46

Appendix

Multilevel Regression Equation for Model 4 (See Model 4 in Table 4)

All multilevel models were run in IBM SPSS (Version 20 Mixed Models) using restricted maximum-likelihood estimation. Provided below is a sample model equation for the EF direct assessments and teacher reports entered simultaneously to predict language outcomes. In the Level 1 equation, the posttest language score (Y) for a child (i) who is in classroom (j) situated in school (k) and system (l) is a function of the intercept of the mean language score (β_{0jkl}) and the fixed effects associated with a vector of demographic covariates, including pretest language score, interval of time elapsed between pretest and posttest, age at pretest, gender, and individualized education plan status ($\Sigma\beta_{1jkl}$). A child's posttest language score is also a function of the child's teacher-reported EF skills (β_{2jkl}), the individual child direct assessments of EF skills (β_{3jkl}, β_{4jkl}, β_{5jkl}, β_{6jkl}, β_{7jkl}, and β_{8jkl}), and the Level 1 random effect of the mean language score for children in each classroom (ε_{ijkl}). In the Level 2 equation, the intercept is a function of the classroom mean language score (γ_{00kl}), the fixed effects of the Level 1 predictors ($\gamma_{1000} \ldots \gamma_{8000}$), and the Level 2 random effect associated with the intercept (η_{0jkl}). At Level 3, the intercept is a function of the school mean language score (π_{000l}), the experimental condition assignment of the school (π_{0011}), and the Level 3 random effect associated with the intercept (ξ_{00kl}). Last at Level 4, the intercept is a function of the system mean language score (μ_{0000}), the Level 3 experimental condition (μ_{0010}), and the Level 4 random effect associated with the intercept (ω_{0001}).

Level 1 (child level):

$$Yijkl = \beta_{0jkl} + \sum \beta_{1jkl}(\text{Covariates})$$
$$+ \beta_{2jkl}(\text{TEACHER REPORT}) + \beta_{3jkl}(\text{CORSI FORWARD})$$
$$+ \beta_{4jkl}(\text{CORSI BACKWARD}) + \beta_{5jkl}(\text{DCCS}) + \beta_{6jkl}(\text{HTKS})$$
$$+ \beta_{7jkl}(\text{PEG TAPPING}) + \beta_{8jkl}(\text{COPY DESIGN}) + \varepsilon_{ijkl}$$

Level 2 (classroom level):

$$\beta_{0jkl} = \gamma_{00kl} + \eta_{0jkl}$$
$$\beta_{1jkl} = \gamma_{1000}$$
$$\beta_{2jkl} = \gamma_{2000}$$
$$\beta_{3jkl} = \gamma_{3000}$$
$$\beta_{4jkl} = \gamma_{4000}$$
$$\beta_{5jkl} = \gamma_{5000}$$
$$\beta_{6jkl} = \gamma_{6000}$$
$$\beta_{7jkl} = \gamma_{7000}$$
$$\beta_{8jkl} = \gamma_{8000}$$

Level 3 (school level):

$$\gamma_{00kl} = \pi_{000l} + \pi_{001l}(\text{CONDITION}) + \xi_{00kl}$$

Level 4 (system level):

$$\pi_{000l} + \mu_{0000} + \omega_{0001}$$
$$\pi_{001l} = \mu_{0010}$$

Received January 28, 2013
Revision received May 21, 2014
Accepted June 11, 2014 ■