**Off-Task Behavior in Kindergarten: Relations to Executive Function and Academic Achievement**

Lillie Moffett and Frederick J. Morrison

University of Michigan

Note: Lillie Moffett's work was supported by the Institute of Education Sciences Predoctoral Fellowship at the University of Michigan [Grant R305B150012].

American Psychological Association

Journal of Educational Psychology


0022-0663/20/$12.00

2020, Vol. 112, No. 5, 938–955
http://dx.doi.org/10.1037/edu0000397

# Off-Task Behavior in Kindergarten: Relations to Executive Function and Academic Achievement

Lillie Moffett and Frederick J. Morrison
University of Michigan

Behavioral self-regulation supports young children's learning and is a strong predictor of later academic achievement. The capacity to manage one's attention and control one's behavior is commonly measured via direct assessments of executive function (EF). However, to understand how EF skills contribute to academic achievement, it is helpful to investigate how EF manifests in the classroom context and in children's overt behavior. The current study observed 172 kindergarteners for a single school day and captured the total proportion of class time children were off-task in the classroom. This behavior was further classified into specific subtypes to assess whether these categorizations differentially predicted components of EF and academic achievement in first grade. Results indicated that children with lower response inhibition spent statistically significantly more time in one type of off-task behavior (i.e., off-task actively engaging with other materials), and children with lower working memory spent significantly more time in another type of off-task behavior (i.e., off-task passively disengaged). Higher proportion of class time spent off-task passively disengaged in kindergarten further statistically significantly predicted fewer gains in reading comprehension in first grade. These findings illustrate the utility of measuring children's EF in a classroom context, and how fine-grained observation systems can shed light on the specific classroom and child processes that influence learning.

***Educational Impact and Implications Statement***

In this study we find that executive functioning can be captured in the kindergarten classroom context by observing subtypes of off-task behavior. The implications of this study suggest that focusing on how children go off-task in the classroom may offer insight into specific response inhibition and working memory deficits kindergarteners may have, and how this behavior may later predict later academic achievement. For example, we find that going off-task with a peer and going off-task by engaging in other, more appealing activities (e.g., playing with instructional materials) is not particularly detrimental to a child's later academic achievement. Whereas going off-task by disengaging altogether (e.g., mind wandering) may be more deleterious for a child's academic outcomes. In effectively identifying these classroom behaviors, efforts can be made to further elucidate the mechanisms by which these behaviors lead to fewer gains in academic achievement and the steps teachers can take to mitigate the consequences of certain types of off-task behaviors in their classroom.

*Keywords:* self-regulation, executive function, off-task behavior, classroom observations



When entering kindergarten, children face increased expectations to regulate their behavior. They must focus attention during instruction, sit for extended periods of time, work on tasks individually, transition from one activity to the next independently, and remember and follow more complex rules and directions. The integration of these skills has been commonly referred to as *behavioral self-regulation*, or the manifestation of executive function (EF) in behavior (McClelland et al., 2007). Past research has demonstrated the importance of EF in early elementary school, and particularly its impact on overall learning and resulting academic achievement (Duncan et al., 2007; McClelland, Acock, & Morrison, 2006; Pianta & Rimm-Kaufman, 2006). However, EF is typically assessed individually and directly, and few studies have related these direct assessments to observations of behavioral self-regulation in the classroom context, which typically focus on observer ratings of task engagement and other "learning-related behaviors" (Bohlmann & Downer, 2016; Chafouleas, McDougal, Riley-Tillman, Panahon, & Hilt, 2005; Griggs, Mikami, & Rimm-Kaufman, 2016; Nesbitt, Farran, & Fuhs, 2015). A separate measure of classroom behavioral self-regulation—namely, duration of time spent off-task—is an operationalization that has been less extensively examined, even though this behavior has also demon-

This article was published Online First July 18, 2019.

Lillie Moffett and Frederick J. Morrison, Department of Psychology, University of Michigan.

Correspondence concerning this article should be addressed to Lillie Moffett, Department of Psychology, University of Michigan, East Hall 3242, 530 Church St., Ann Arbor, MI 48109. E-mail: lillmoff@umich.edu

strated relations with EF and academic achievement (Day, Connor, & McClelland, 2015; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009).

A recent review of EF and self-regulation assessments emphasized the need to measure EF in context (McCoy, 2019). The author highlighted how real-world settings are filled with "distractions, emotions and supports" (p. 1), and thus measurement in context could create a valuable tool for practitioners tasked with addressing behavior issues in the classroom. The purpose of the current study is to take a closer look at quantifying behavioral self-regulation (and its underlying EF components) directly in the classroom by observing children's off-task behavior, and its subtypes. This method of measurement is less prevalent than other classroom observation systems; however, it is possible that quantifying a lack of behavioral self-regulation as a duration of time (ISI; Connor et al., 2009) is preferable to a rating scale that may be more subjective (Observed Engagement in Learning Scale; Rimm-Kaufman et al., 2005) or time-sampling "snapshot" method (Farran & Son-Yarbrough, 2001) that may overlook behaviors occurring throughout a longer observation period. In this article, we explore (1) the operationalization (and further categorization of subtypes) of off-task behavior in kindergarten, (2) whether off-task behavior relates to EF components in kindergarten, and (3) if proportion of time spent off-task predicts academic achievement and EF gains in first grade.

## Observations of Off-Task Behavior

Classroom observations of off-task behavior have been historically used with ADHD populations (Abikoff, Gittelman, & Klein, 1980; Atkins & Pelham, 1991; Gaastra, Groen, Tucha, & Tucha, 2016; Platzman et al., 1992), although there has been increased attention on typically developing children (Austin & Soeda, 2008; Kilian, Hofer, Fries, & Kuhnle, 2010). One of the first studies to conduct an in-depth analysis of off-task behavior in typically developing elementary schoolchildren (Godwin et al., 2016) investigated whether patterns of off-task behavior vary across the school year, across grade levels, between gender, and among subtypes of off-task behavior. Although comprehensive in quantifying the prevalence of off-task behavior across different variables, this study did not explore off-task behavior associations with EF or achievement outcomes, thus it is difficult to ascertain whether off-task behavior is a predictive reflection of poor EF skills, and further predicts student academic outcomes.

Two studies did examine whether off-task behavior (as part of a larger construct of unproductive noninstructional activities) related to student outcomes (Day et al., 2015; McLean, Sparapani, Toste, & Connor, 2016). Both studies found that, in first grade, time spent in noninstructional activities predicted fewer EF gains at the end of the school year but yielded mixed findings on whether time spent in these activities predicted gains in literacy. However, because these studies bundled off-task behavior into a larger construct of unproductive noninstructional activities, it is difficult to tease apart how off-task behavior independently related to EF, and why perhaps off-task behavior relates to literacy achievement in one study, but not in another. In the present study, disentangling which facets of unproductive noninstruction (like off-task behavior) predict academic achievement, will yield a clearer understanding of which aspects of the classroom matter for achievement.

## Classroom Context and Off-Task Behavior

When attempting to capture EF in context (e.g., behavioral self-regulation), it is also important to consider the dynamic interplay between individual differences in EF and the classroom environment (McClelland & Cameron, 2011). An advantage of direct assessments of EF is that environmental influence is minimized; however, when it comes to addressing EF in the classroom, capturing environmental factors is crucial to understanding how these cognitive processes manifest in overt behavior. Some studies have attempted to separate the classroom from the child in analyzing off-task behavior by mean-centering this behavior at the classroom level (Nesbitt et al., 2015) Others have included additional classroom factors in their analyses of behavioral self-regulation in the classroom (Day et al., 2015; Timmons, Pelletier, & Corter, 2015). Findings reveal strong (likely bidirectional) associations between classroom factors and children's EF. Further, when predicting academic achievement, past work has found that children who spend more time in unproductive noninstruction (e.g., class disruptions, waiting for the teacher to get organized, lengthy transitions)—in addition to being more off-task—make fewer literacy gains (Day et al., 2015).

One explanation for this finding is that children demonstrate poor behavioral self-regulation in the classroom, in part, because of poor teacher organization and classroom management (Bohn, Roehrig, & Pressley, 2004; Rimm-Kaufman et al., 2005). This interpretation is supported by past findings that improving teacher's planning/organizing leads to greater growth in EF and literacy (Connor et al., 2010; Pressley et al., 2001). Another explanation, on the other hand, is that children who demonstrate high levels of behavioral self-regulation in the classroom do so because they are well-managed, and the content of instruction is more engaging—potentially masking individual variability in EF. Other classroom factors, such as how children's attention is managed, also influence behavioral self-regulation. For example, children are typically off-task less during play and small-group contexts in kindergarten (Timmons et al., 2015), as well as during whole-class instruction versus when they are working independently (Godwin et al., 2016).

Thus, children's off-task behavior can be conceptualized as a combination of individual differences in underlying EF, as well as teacher's effectiveness in managing behavior, teacher expectations for behavior, the behavior of one's peers, and overall task difficulty during learning opportunities. To address these dynamic elements, the current study measured relevant facets of the classroom environment, such as "unproductive" (teacher disorganization/waiting, disruptions) and "productive" classroom factors (teacher planning/organizing), as well as total time the teacher is managing students' attention (vs. the child working independently; Day et al., 2015).

## Types of Off-Task Behavior

Beyond the broad notion of off-task behavior, some investigations have probed the nuances within the broader category. Similar to other behavior observation systems, measures of off-task (and on-task) behavior can be vague and inconsistent in their definition of what it means to be "off-task" (Gill & Remedios, 2013). Thus, careful classification and specific example behaviors are needed to help elucidate which behaviors are most crucial when examining this construct and its relation to student outcomes. For example, Shapiro (2004) developed the Behavioral Observation of Students in School (BOSS) system, which distinguished between three types of off-task behavior (motor, verbal, and passive); however, there are no published data on the convergent validity of this system, which make it difficult to judge its utility. Despite lack of validity, this coding scheme has been used to investigate differences across these three types of off-task behaviors with a sample of children with ADHD and typically developing children (Junod, Dupaul, Jitendra, Volpe, & Cleary, 2006). They found that children with ADHD engaged in more off-task behavior than did their non-ADHD peers, but both groups of children engaged in far more motor off-task behavior than verbal or passive off-task behavior (measured via frequency counts within 15-s intervals). It is important to note that this study observed children for a total of only 15 min, and because the focus of the study was on differences between children with and without attention-deficit/hyperactivity disorder (ADHD), little attention was given to explaining emergent frequency differences between the three types of off-task behavior. Thus, whether these types of off-task behavior matter for student outcomes remains to be explored.

Other studies have focused more closely on specific subtypes of off-task behavior and emerging profiles of classroom behavioral self-regulation more generally. Godwin et al. (2016) investigated associated sources of off-task behavior (i.e., distracted by self, peer, or items in the environment) and found important gender and context differences across subtypes, suggesting that more fine-grained specification of off-task behavior would be revealing. In their study, females were more likely to be peer-distracted, whereas males were more likely to be environmentally distracted. Peer distractions also occurred more frequently in individual and small-group contexts, whereas self and environmental distractions were more frequent during whole-group instruction. However, this study live-coded behavior for 20-s intervals in a round-robin type fashion, to which each observer either noted whether an off-task behavior was present or not (Baker–Rodrigo Observation Method Protocol [BROMP]; Ocumpaugh, Baker, & Rodrigo, 2012). Although reliable, this observation method might be less comprehensive than a methodology that observes each child continuously throughout a video observation. Godwin et al. (2016) also did not examine whether these types of off-task behaviors differentially predict important student outcomes, such as academic achievement. One investigation that assessed profiles of self-regulation, however, did find evidence for differential predictions to math and reading skills, depending on the profile (Magi, Mannamaa, & Kikas, 2016). This study determined these self-regulation profiles by grouping children within classrooms into low, middle and high self-regulation groups based on their performance in planning and task-persistent assessments. Because the authors did not look at children's behavior in the classroom, it would be valuable to see whether a coding system that captures such "profiles" in the classroom predicts academic achievement as well.

Using Godwin et al.'s (2016) "sources" of off-task behavior (peer, environment and self-distractions) and Shapiro's (2004) off-task subtypes (verbal, motor, and passive) as guidelines, the current study aimed to generate subtypes of off-task behavior and explore whether these "types" differentially relate to EF and academic achievement. For example, it's possible that some types of off-task behavior are more harmful for children's ability to learn and make academic gains than are others. In Godwin et al.'s (2016) study, peer distractions were the most prevalent source of off-task behavior and occurred more frequently when children were working individually and/or in a small group. But we may find that this type of off-task behavior, although frequent, may not relate to children's academic outcomes considering social competence (e.g., engaging in positive peer interactions) is positively related to academic skills (Mashburn et al., 2008). Further, peers are likely one of the most distracting part of a child's classroom environment, so even children with higher self-regulation might not be able to resist the temptation of talking with their peer. Godwin et al. (2016) also noted that peer distractions occur more frequently when children are working independently—a context that might also prove difficult for most children, regardless of their EF capacities. On the other hand, Godwin et al. (2016) found that self and environmental distractions were more prevalent during whole-class instruction—a context where teachers are managing children's attention, so off-task behavior should hypothetically occur less frequently in this context. It's plausible that children with the lowest EF skills, despite teacher attention management, are off-task the most with these self and environmental-distractions. Furthermore, spending more time in these types of off-task behaviors, as opposed to peer interactions may be more deleterious for a child's academic gains.

## Current Study

In summary, past studies have either looked closely at the nature of off-task behavior, or at how off-task behavior relates to classroom variables, EF, or academic achievement; however, no study to date has investigated these questions simultaneously.

The purpose of the current study is to

(1) explore the classification of off-task behavior into three subtypes—off-task nonengaged, off-task engaging in other activity, and off-task interacting with peer—and examine the resulting measurement reliability and relations to individual child differences and classroom factors;

(2) investigate the extent to which subtypes of off-task behavior—versus off-task behavior as a unitary construct—relate to EF components in kindergarten, and further predict gains in EF in first grade; and

(3) determine whether off-task behavior in kindergarten (subtypes of this behavior vs. unitary contrast) predicts gains in academic achievement in first grade, controlling for child baseline characteristics and time spent in classroom factors (e.g., teacher dis/organizational time).

## Method

### Participants

The current sample consists of 172 kindergarteners ($M = 5.76$ years, $SD = 0.38$; 47% female) across three cohorts of data collection ($n = 69$, Year 1; $n = 67$, Year 2; $n = 36$, Year 3). Students were recruited from four elementary schools (totaling 22 classrooms) in Southeastern Michigan and were observed for a single school day in the spring of kindergarten. These students were part of a larger longitudinal study examining the effect of schooling on EF development in elementary schoolchildren, approved by the University of Michigan's Institutional Review Board (HUM00085500). To recruit students, schools sent home a letter to parents, who then decided whether to sign their child up for the study. Students from the four schools represented a wide range of socioeconomic statuses. From the present study, 32 students attended a school where 2% of the student population qualified for free or reduced-price lunch (FRPL), 64 students attended a school where 61% qualified, 42 children attended where 69% qualified, and 34 students attended a school were 72% of the student population qualified for FRPL. Kindergarteners were tested on direct assessments of EF and academic achievement in the fall of kindergarten and first grade, and classroom observations occurred during the spring of kindergarten. Children with an Individualized Education Program ($n = 6$) were included in the study. Children were only initially excluded from further analyses if they were absent from the classroom for more than 25% of the classroom observation period. A total of 19 children across the 3 years were absent from their observation period—this number is not included in the $n = 172$; we omitted these children from the sample because we were only interested in including children with observational data in our analyses.

## Procedure

Parental consent was obtained for all children in the study as well as teacher consent for the classroom observations. Direct assessments of EF and academic achievement were collected in the fall of kindergarten and first grade. Children were individually tested outside of their classroom for a 45-min period, where they were assessed on a battery of EF and academic achievement measures. The order and versions of assessments were counterbalanced. There were two different orderings of assessments and there were also two different versions of each of the assessments. Children were provided a 2-min break between every two assessments where they were given the opportunity to decorate a bookmark with stickers.

Classroom observations were conducted in the spring of kindergarten and the length of recorded videos varied depending on how long students were physically in the classroom during their full-day kindergarten program (i.e., gym, lunch, recess, art, etc., were not observed), and for how long the teacher agreed to be recorded ($M = 3.18$ hr $SD = .24$). Each of the 22 classrooms was observed on a different day, and all observations within a cohort were conducted within 69 days or less. There were an average of 7.8 children observed per classroom. To ensure we did not miss any behaviors from children in our study, two camera angles were used and adjusted accordingly when target children moved about the classroom. Across the three years of data collection and 22 classrooms, 13 teachers participated in the study; we observed seven teachers once, three teachers two times (e.g., a teacher participated in Year 1 and Year 2 of the study), and three teachers three times.

Detailed observation notes were taken in order to clarify potential ambiguities that may arise when coding the videos (e.g., descriptions of the worksheets children were working on, where children went when they left the room, the content of conversations between peers that video audio may not have picked up). The individuals who coded the classroom videos were not the same individuals who administered the direct assessments or conducted the observations.

## Measures

Three tests of executive function were included. Each test is theorized to emphasize a different component of EF (e.g., working memory, response inhibition and attentional control), although it is recognized that all EF tasks include a mixture of all components.

**Working memory.** Children were tested on the Backwards Digit Span subtest of the Wechsler Scales (Wechsler, 1991). In this task, children are asked to repeat number sequences in reverse order, ranging from two to six strings of numbers. Children received a score based on the longest string they could successfully repeat backward. This measure demonstrates acceptable test–retest reliability ($r = .73$; Lipsey et al., 2017).

**Response inhibition.** The heads-toes-knees-shoulders (HTKS) task (Ponitz et al., 2008) was used to test children's ability to inhibit prepotent responses. In this task, children were asked to touch the opposite body part from that of which the experimenter requested they touch (touch head when hear "touch toes" and vice versa), with the task increasing in difficulty over time (more body parts, e.g., shoulders and knees; and a rule switch e.g., toes and shoulders, head and knees). There was a total of 30 trials, where children received a 0 for incorrect responses, 1 for self-corrected responses, and 2 for correct responses. The internal consistency of this measure for the current study is .83 (Cronbach's alpha).

**Attentional control.** The Pair Cancellation subtest from the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001) was used. In this task, children were asked to pay attention to and draw a circle around two consecutive pictures of items (dog, then ball) on a larger worksheet covered with pictures of dogs, balls and cups. W scores were used for analyses. Test–retest reliability for this subtest is $r = .78$ (Mather & Woodcock, 2001).

**Academic achievement.** Math and literacy skills were assessed using the WJ-III (Woodcock et al., 2001). Specifically, W scores from the Applied Problems subtest were used to measure math skills, and W scores from the Letter-Word Identification and Passage Comprehension subtests were used to measure literacy skills. There is well-established reliability on all three of these subtests (Woodcock & Johnson, 1990).

## Observational Measures

**Off-task behavior and productive/unproductive noninstruction.** Off-task behavior was examined by the Individualized Student Instruction (ISI) Coding System (Connor et al., 2009) using The Noldus Observer XT 13 software (Noldus Information Technology, 2013). The ISI system codes duration of time each child is experiencing a given type of instructional activity (e.g., literacy, math etc.) or noninstructional activity (e.g., transitioning, waiting for the teacher to give directions, listening to directions/planning, going off-task etc.). The ISI system also records duration of time children are in a given context (e.g., whole-class, individual, or peer) and who is managing the child's attention (e.g., the child, peers, or the teacher). Although each child was observed using the full ISI coding system, the present study only looked at noninstructional variables—specifically the duration of time children were in "off-task unproductive" behaviors, and other noninstructional activities hypothesized to be related to EF and subsequent academic achievement. Productive noninstructional activities were defined as the teacher providing directions, behavior expectation/rules, or orienting children to an activity (planning/organizing). Unproductive noninstruction (teacher) is defined as activities when the children were waiting for the teacher to get organized (e.g., a lull in instruction), and for time spent in disruptions and students are waiting for the activity or lesson to resume. In the ISI coding system, *off-task behavior*—or *off-task unproductive behavior*—is defined as blatantly not completing the activity he/she was assigned, such as being out of one's seat for no purpose and is only recorded if it lasts at least 15 s. Reliability on the ISI coding system was achieved between eight coders ($\kappa > .80$).

**Off-task behavior subtypes.** In order to determine whether operationalizing the type of off-task behavior makes a difference in predicting individual differences in underlying EF components, a coding system was developed to further specify the off-task behavior. To create this coding system, off-task behaviors were first qualitatively described, and then further separated into categories based on similarities, using Godwin et al.'s (2016) coding system as a template. At the end of this phase, four categories of off-task behavior were identified: off-task nonengaged, off-task interacting with a different activity, off-task interacting with peer, and off-task other. Moderate reliability was established on this measure between two coders; ICC (2, 2) = .71 across all 22 video observations. These codes were also recorded using the Noldus Observer XT software.

***Nonengaged.*** This behavior is defined as a child not interacting with or engaging in the activity he or she is supposed to be (and not interacting with anything else as an alternative). A child's eye-gaze is considered when coding this behavior, although it is not the only indicator. In-seat examples include not looking at the workbook one is supposed to be doing, and instead looking around the room at other children, or some other part of one's desk. However, one is only coded as nonengaged if one is not obviously attempting to interact with the given activity (i.e., pencil has been set down, the workbook has not been opened, child is leaning back in chair). A child gazing away from his paper while still leaning forward, holding pencil and worksheet on desk, was given the benefit of the doubt that he may have still been on-task (e.g., he could be thinking of what to write) and thus was not coded as off-task.

A common out-of-seat example of this behavior occurs at the end of a transition period, when other children have already started an activity, yet some children are still out of their seats wandering and not explicitly engaging in anything else. This behavior might also be observed when children are tasked with doing individual work at their seats, and some children get out of their seats to do something initially (what ISI terms) *off-task productive* (i.e., sharpens their pencil), but instead end up somewhere else (e.g., wandering around the room). Behaviors that fall within this category more broadly also require that the child not be actively interacting/engaging with anything else (e.g., organizing their desk, talking to a peer, playing with a toy) and is, instead, passively disengaged (e.g., they didn't seek out a distraction, per se). In contexts where the task is to attend and listen to the teacher, the child is coded as nonengaged only if he or she is physically turned away from the teacher (head or entire body), because at this point, the child is not overtly following the expectations of how to behave when the teacher is teaching. A child gazing away from the teacher while his or her body is still oriented toward the teacher was not coded as off-task because it was difficult to determine whether the child was still listening, thus the child was given the benefit of the doubt.

***Other activity.*** This behavior is defined as not engaging in the given activity or task at hand (e.g., worksheet at desk, listening to teacher, cooperating with peers), and instead actively engaging in another activity. Examples include organizing the content of one's pencil box when they should be writing, drawing on their worksheet when the teacher is giving a lesson, playing/making a tower of base-ten blocks instead of counting them, and so forth This behavior also includes motor activities such as drumming or banging one's desk or bouncing up and down in one's seat. An out-of-seat example of this is when children are supposed to be at their seat or center table but are instead at a different part of the room playing with toys they are not supposed to be playing with. This can also be observed during transition periods when children get distracted with cleaning up materials from their center area and instead start playing or interacting with materials from a different center area.

***Interacting with peer.*** This behavior is defined as talking or playing with another peer when not instructed to. If the child is talking to his or her peer about the task or related to the activity at hand, this was not coded as off-task unless the behavior expectation of the teacher for the whole class during the given activity was straightforward (i.e., the room is otherwise completely silent, or the teacher explicitly says "no talking"). Discussing other matters unrelated to the task at hand, however, was considered off-task. As soon as a peer interaction becomes more centered on an activity, however (e.g., two children are building with base-ten blocks instead of counting with them), then this is coded as in-seat/other activity, as now the child's attention is more directed toward activity than the peer.

***Other.*** This behavior was coded when a child was off-task but the behavior did not fall into the above three categories. Instances of this type were generally emotional in nature and ignored by the teacher. An example is a child crying at their desk while the rest of the class is engaged in whole-class instruction at the carpet. Another example is when a child is throwing a tantrum in the corner of the room but not explicitly disrupting the rest of the class (e.g., cannot be coded with the "disruptor" code in the ISI coding system because the rest of the children are far enough away to not be distracted). This behavior only occurred for three children in the study so was not included in analyses.

## Analytic Strategy

This study included a total of two literacy assessment variables, one math assessment variable, three executive functioning variables, and six classroom observation variables. W scores for all the WJ-III assessments were used, as this score allows for the direct comparison of the achievement of one student against another, regardless of age. Raw scores for HTKS and Backward Digit Span were used. Since there was variability in total amount of class time observed (2.84 hr to 4.25 hr), percent of total class time was used for all of the classroom variables: off-task behavior (off-task total, off-task nonengaged, off-task other-activity, off-task peer), productive noninstruction (teacher planning/organizing), teacher unproductive noninstruction (disruptions/teacher disorganization). Proportion of time spent in teacher-managed instruction (e.g., whole class instruction, small-group with the teacher, teacher-child one-on-one experiences) was also included. Of importance to note is that off-task behavior, productive noninstruction and unproductive noninstruction are mutually exclusive, but time spent in teacher-managed instruction is not mutually exclusive with off-task behavior. For example, children could go off-task during teacher-managed instruction. We chose to use percent of total class time for all our classroom variables in order to standardize our variables across classrooms of varying observation lengths.

**Preliminary analyses.** Descriptive statistics (see Table 1) revealed that direct assessments fell within the expected ranges for kindergarten and first grade, and children made anticipated gains in these domains. There were also no significant differences between the three cohorts on any variables, thus the three cohorts were combined in further analyses. Bivariate correlations among all study variables are presented in Table 2.

**Missing data.** Small amounts of missing data were present in the direct assessments in kindergarten, but all observational data were complete (see Table 1). No significant relations were detected between missing data on the direct assessments and observation variables in kindergarten. Longitudinal data (first grade academic achievement scores) were available for 121 out of 172 kindergartners. To address this significant amount of missingness, we conducted multiple imputation using chained equations with 200 imputations. The resulting imputed data set was used in all subsequent analyses.

**Analysis.** All regression analyses were conducted controlling for classroom-level covariates (time in productive and unproductive noninstruction activities, and teacher-managed instruction), student-level covariates (age and gender), and a fixed effect for school. To address the nesting of students in classroom, standard errors were clustered at the classroom level. To examine whether kindergarten total off-task behavior versus the three subtypes of off-task behavior added unique variance in explaining differences in first grade EF skills and academic achievement—half of these models included only off-task behavior and covariates, and the rest included off-task behavior subtypes and covariates.

## Results

In an average 190-min observation time period, kindergarteners in this study spent an average of 14 min off-task (8.87% of instructional time), with a range of 0 to 78 min (see Figure 1). An instance of off-task behavior lasted as short as 15 s and as long as 22 min ($M = 1.8$ min). Although, there was little difference between average proportion of time spent in each of the three types of off-task behavior across the sample (see Table 1), descriptively speaking, more children displayed instances of peer off-task behavior ($n = 116$), followed by off-task nonengaged ($n = 89$), and off-task other activity ($n = 68$; see Figures 2 through 4). On average, children spent 14% of class time in productive noninstruction, 9% of class time in unproductive noninstructional activities, and 57.01% of class time in teacher-managed instruction.

### Further Classification of Off-Task Behavior

Both total proportion of time spent off-task and the subtypes of this behavior related to individual and classroom variables. Zero-order correlations revealed that males (dummy coded = 0) were statistically significantly more likely to go off-task than females (dummy coded = 1; $r = -.16$, $p < .05$); although males were statistically significantly more likely to go off-task in another activity ($r = -.22$, $p < .01$; 3.77% vs. 1.01%) whereas females were statistically significantly more likely to go off-task with their peers ($r = .18$, $p < .05$). Males spent 2.03% of class time in this behavior, females spent 4.62%.

There was also a statistically significant positive correlation between off-task behavior and baseline literacy, where children who entered kindergarten with higher baseline literacy scores went off-task more ($r = .15$, $p < .05$). When investigating the subtypes of off-task behavior, however, this literacy and off-task behavior association is primarily driven by proportion of time spent off-task with a peer ($r = .28$ $p < .01$). We know this because other

Table 1
*Descriptive Statistics for Kindergarten and First Grade Student-Level Variables*

| | Kindergarten | | | | | First grade | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | $n$ | Minimum | Maximum | $M$ | $SD$ | $n$ | Minimum | Maximum | $M$ | $SD$ |
| Response inhibition | 167 | 0 | 56 | 29.26 | 18.86 | 120 | 0 | 59 | 43.47 | 11.62 |
| Attentional control | 162 | 414 | 478 | 456.90 | 11.46 | 119 | 436 | 491 | 470.64 | 9.14 |
| Working memory | 167 | 0 | 7 | 1.97 | 1.63 | 120 | 0 | 6 | 3.34 | 1.24 |
| Letter-word identification | 167 | 283 | 507 | 384.45 | 31.07 | 121 | 369 | 525 | 445.47 | 33.14 |
| Passage comprehension | 164 | 358 | 480 | 419.74 | 20.43 | 121 | 413 | 494 | 460.29 | 18.53 |
| Applied problems | 165 | 333 | 462 | 429.10 | 19.56 | 121 | 419 | 502 | 456.33 | 15.20 |
| Off-task total time | 172 | 0% | 65.06% | 8.87% | 11.04% | | | | | |
| Off-task nonengaged | 172 | 0% | 32.14% | 2.72% | 5.11% | | | | | |
| Off-task other activity | 172 | 0% | 20.03% | 2.09% | 3.81% | | | | | |
| Off-task peer | 172 | 0% | 25.89% | 2.70% | 4.05% | | | | | |
| Teacher-managed instruction | 172 | 19.09% | 95.24% | 57.01% | 26.33% | | | | | |
| Child-managed instruction | 172 | 4.28% | 60.06% | 31.11% | 15.18% | | | | | |
| Time in productive noninstruction | 172 | 2.78% | 27.66% | 14% | 7.03% | | | | | |
| Time in unproductive noninstruction | 172 | .74% | 16% | 8.73% | 5.89% | | | | | |

Table 2
*Pearson Bivariate Correlations of Off-Task, EF, Academic Achievement and Classroom Variables*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Off-task total | — | | | | | | | | | | | | | | | | | | | | |
| 2. Off-task nonengaged | .91* | — | | | | | | | | | | | | | | | | | | | |
| 3. Off-task other activity | .79* | .59* | — | | | | | | | | | | | | | | | | | | |
| 4. Off-task peer | .84* | .68* | .48* | — | | | | | | | | | | | | | | | | | |
| 5. K AC (EF) | .03 | −.12 | −.10 | .11 | — | | | | | | | | | | | | | | | | |
| 6. K WM (EF) | .05 | −.15* | −.06 | .22* | .35* | — | | | | | | | | | | | | | | | |
| 7. K RI (EF) | −.11 | −.08 | −.34* | .17* | .26* | .34* | — | | | | | | | | | | | | | | |
| 8. K WJ PC | .17* | −.05 | −.03 | .17* | .24* | .36* | .22* | — | | | | | | | | | | | | | |
| 9. K WJ LWID | .15* | .02 | .11 | .28* | .40* | .49* | .30* | .64* | — | | | | | | | | | | | | |
| 10. K WJ AP | .03 | −.05 | −.07 | .13 | .34* | .47* | .48* | .32* | .55* | — | | | | | | | | | | | |
| 11. First AC (EF) | .07 | −.02 | −.03 | .18 | .47* | .22* | .28* | .26* | .32* | .39* | — | | | | | | | | | | |
| 12. First WM (EF) | −.00 | −.03 | −.20* | .18 | .22* | .44* | .24* | .17 | .26* | .43* | .33* | — | | | | | | | | | |
| 13. First RI (EF) | −.03 | −.00 | −.07 | .16* | .27* | .33* | .42* | .18* | .28* | .39* | .36* | .30* | — | | | | | | | | |
| 14. First WJ PC | .03 | −.17* | −.11 | .04 | .28* | .47* | .30* | .47* | .59* | .46* | .29* | .37* | .26* | — | | | | | | | |
| 15. First WJ LWID | .06 | −.14 | −.09 | .06 | .29* | .46* | .31* | .50* | .68* | .48* | .30* | .32* | .25* | .88* | — | | | | | | |
| 16. First WJ AP | −.09 | −.12 | −.16* | .07 | .34* | .51* | .38* | .37* | .47* | .54* | .37* | .40* | .25* | .60* | .56* | — | | | | | |
| 17. Productive noninstruction | −.06 | −.05 | .00 | −.10 | .00 | −.09 | −.10 | −.05 | −.08 | −.05 | −.15* | −.18* | −.15* | −.13 | −.16* | −.08 | — | | | | |
| 18. Unproductive noninstruction | .10 | .09 | .05 | .11 | −.02 | −.07 | −.03 | −.06 | −.07 | .02 | −.11 | −.18* | .00 | −.17* | −.21* | −.12 | .34* | — | | | |
| 19. Teacher-managed instruction | −.25* | −.20* | −.20* | −.26* | −.02 | −.12 | −.01 | −.03 | −.17* | −.00 | .00 | −.01 | .00 | −.07 | −.11 | −.11 | −.28* | .06 | — | | |
| 20. Gender[a] | −.16* | −.11 | −.22* | .18* | .08 | .03 | .25* | −.06 | −.06 | .12 | .05 | .03 | .09 | −.10 | −.16 | −.14 | .04 | −.05 | −.09 | — | |
| 21. Age | .12 | .08 | .09 | .14 | .26* | .20* | .20* | .27* | .39* | .25* | .10 | .06 | .10 | .06 | .05 | .11 | .08 | .12 | .05 | −.02 | — |

*Note*. K = Kindergarten; AC = attentional control; EF = executive function; WM = working memory; RI = response inhibition; WJ PC = Woodcock-Johnson Passage Comprehension subtest; WJ LWID = Woodcock Johnson Letter-Word Identification subtest; WJ AP = Woodcock-Johnson Applied Problems subtest.
[a] male = 0, female = 1.
* $p < .05$.

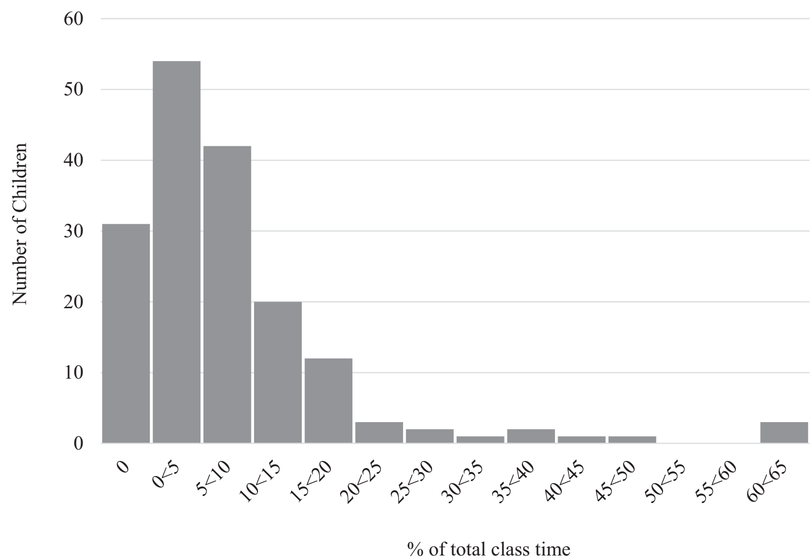


*Figure 1.* Distribution of percent of total class time children spent engaging in total off-task behavior.

correlations reveal that the more time a child spent off-task with a peer, the higher one's kindergarten and first grade achievement and EF scores more generally. However, the more time children spent off-task nonengaged and in another activity revealed mixed associations with EF and academic achievement.

All off-task behavior variables were not statistically significantly related to proportion of time spent in productive or unproductive noninstructional activities. However, the higher proportion of time spent off-task, the statistically significantly lower proportion of time spent in teacher-managed instruction. It is important to note, however, that all of these relations are zero-order correlations, so they do not control for potentially confounding variables. The coding of these behaviors attained moderate measurement reliability (ICC [2, 2] = .71), although only two coders were included in this analysis and further work should be done to establish reliability of coding subtypes of off-task behavior.

## Kindergarten Off-Task Behavior and Kindergarten and First Grade Executive Function

We investigated how the EF components in kindergarten—controlling for classroom and child variables—predicted each off-task variable in kindergarten (see Table 3). All models were Bonferroni corrected. The higher a child's response inhibition, the statistically significantly less proportion of time they spent off-task in another activity (ß = $-.40$, *SE* = .09, $p < .001$). The higher one's working memory the statistically significantly lower the proportion of time spent off-task nonengaged (ß = $-.18$, *SE* =

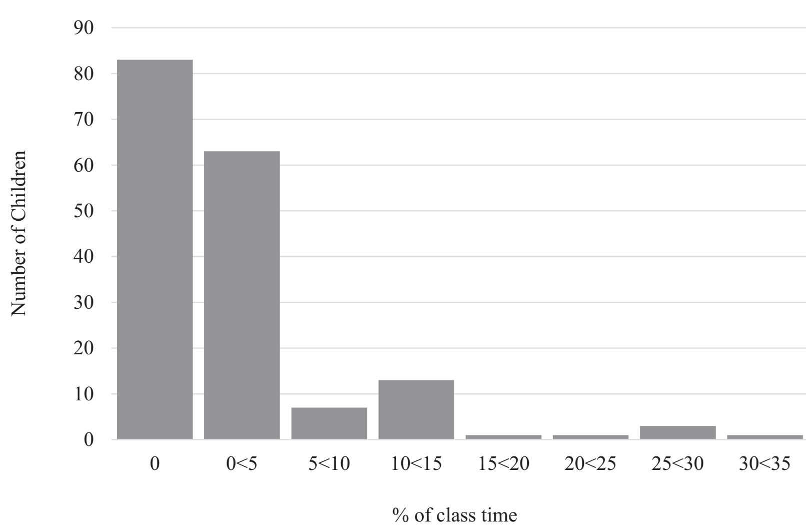


*Figure 2.* Distribution of percent of total class time children spent off-task nonengaged (specific type of off-task behavior).

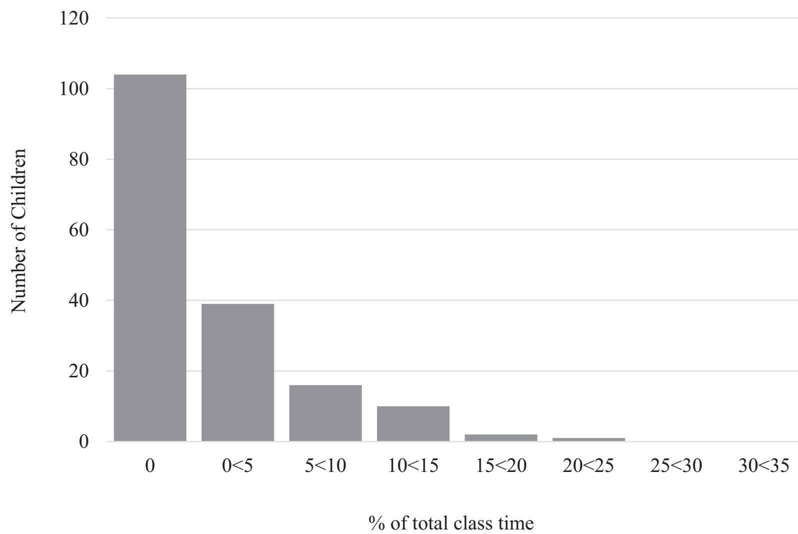


*Figure 3.* Distribution of percent of total class time children spent off-task other-activity (specific type of off-task behavior).

.06, $p < .01$). For predictions to first grade EF, results revealed that no off-task behaviors statistically significantly predicted gains in first grade executive functioning (see Table 4).

## Kindergarten Off-Task Behavior and First Grade Academic Achievement

After Bonferroni corrections, results (presented in Table 5) showed that total proportion of time off-task in kindergarten predicted statistically significantly fewer gains in letter-word identification from kindergarten to first grade (ß $= -.20$, $SE = .05$, $p < .01$). The higher proportion of time spent off-task nonengaged predicted statistically significantly fewer gains in reading comprehension in first grade (ß $= -.23$, $SE = .07$, $p < .01$). No off-task behavior subtypes statistically significantly predicted gains in first grade math achievement. To see whether kindergarten off-task behavior explained more variance above and beyond baseline executive functioning in prediction to first grade academic achievement, regressions were rerun controlling for kindergarten EF (see Appendix A). After Bonferroni corrections, higher proportion of time spent off-task (as a unitary construct) predicted statistically significantly fewer gains in first grade reading comprehension (ß $= -.17$, $SE = .05$, $p < .01$) and letter-word identification (ß $= -.22$, $SE = .04$, $p < .001$).

## Robustness Checks

Although we included classroom variables and clustered standard errors at the classroom level, we attempted to further tease apart child versus classroom influence on off-task behavior by also

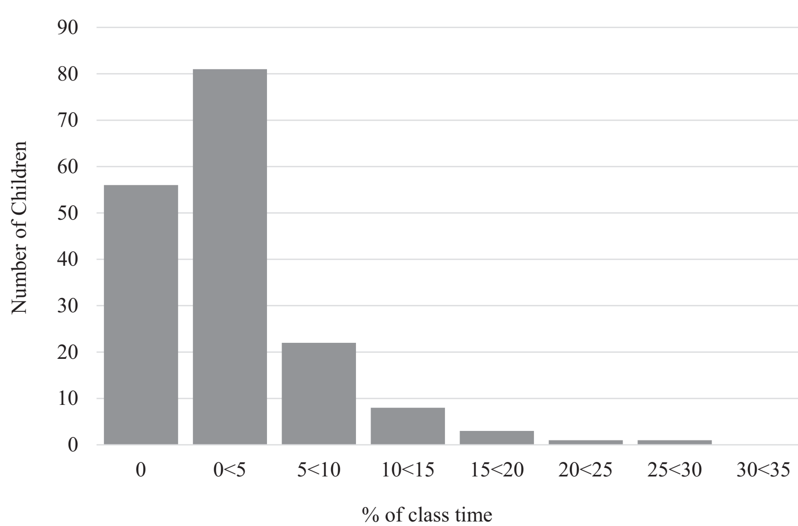


*Figure 4.* Distribution of percent of total class time children spent off-task with a peer (specific type of off-task behavior).

Table 3
*Fall Kindergarten EF Components Predicting Spring Kindergarten Off-Task Behavior*

| Variable | Total off-task | Off-task nonengaged | Off-task other activity | Off-task peer |
|---|---|---|---|---|
| Response inhibition | −.02 (.06) | .11 (.05) | **−.40 (.09)**[a] | **.18 (.06)** |
| Attention control | −.05 (.06) | **−.14 (.06)** | .01 (.09) | .003 (.06) |
| Working memory | .01 (.05) | **−.18 (.06)**[a] | .11 (.05) | .14 (.07)[†] |
| Productive noninstruction | −.08 (.15) | .06 (.15) | −.003 (.11) | −.13 (.13) |
| Unproductive noninstruction | .05 (.07) | .02 (.07) | −.002 (.07) | .12 (.08)[†] |
| Teacher managed | −.49 (.25)[†] | −.41 (.25) | −.35 (.18)[†] | **−.49 (.21)** |
| Gender | −.21 (.10)[†] | **−.31 (.11)** | −.11 (.14) | .12 (.11)[†] |
| Age | .15 (.16) | .20 (.20) | .20 (.16) | .04 (.18) |
| School/FRPL status (Reference is 1: 2% FRPL) | | | | |
| School 2: 61% | −.10 (.26) | −.10 (.26) | −.12 (.21) | −.02 (.27) |
| School 3: 69% | .10 (.33) | .04 (.31) | −.12 (.27) | .30 (.28) |
| School 4: 72% | 1.08 (.68) | .85 (.67) | .87 (.48)[†] | .97 (.58) |

*Note*. Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. EF = executive function; FRPL = free/reduced price lunch.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

mean-centering off-task behavior at the classroom level. We reran all regression analyses with classroom mean-centered off-task variables and outcomes and excluded classroom-level variables that when mean-centered would be 0 for each child (see Appendix B). Results remain stable for kindergarten EF predictions to kindergarten off-task behavior, although attentional control was a statistically significant predictor of off-task nonengaged behavior (ß = −.17, $SE$ = .06, $p < .01$; see Table B1 in Appendix B). For kindergarten off-task behavior predicting first Grade EF, results remain stable with no off-task variables statistically significantly predicting growth in any of the EF components in first grade (see Table B2 in Appendix B). For kindergarten off-task behavior predicting first grade academic achievement, no off-task behaviors predicted growth in academic achievement, which is different from our original models (see Table B3 in Appendix B).

Although clustering standard errors at the classroom level is one method of addressing the nesting of children within classrooms, we additionally conducted a robustness check using hierarchical linear modeling and nesting children within classrooms and schools. We reran models predicting the types of off-task behavior to gains in first grade EF and academic achievement. Models are presented in Appendix C. Results remain stable with the classroom mean-centered results, except that off-task nonengaged was a statistically significant negative predictor of reading comprehension in first grade—this finding aligns with our original (not classroom mean-centered) regression results (see Table C1 in Appendix C).

## Discussion

The present study examined how quantity and type of off-task behavior in Kindergarten related to direct assessments of EF and further predicted EF and academic achievement in first grade. Our study yielded four important findings. First, children's EF components upon entering kindergarten differentially predicted the type of off-task behavior they engaged in—and this remained

Table 4
*Kindergarten Off-Task Behavior Predicting Growth in EF from Kindergarten to First Grade*

| Variable | Attention control | | Working memory | | Response inhibition | |
|---|---|---|---|---|---|---|
| Baseline EF | **.38 (.10)**[a] | **.34 (.11)**[a] | **.52 (.11)**[a] | **.47 (.14)**[a] | **.40 (.10)**[a] | **.44 (.12)**[a] |
| Off-task total | .07 (.09) | | .03 (.12) | | .04 (.07) | |
| Nonengaged | | −.09 (.13) | | −.01 (.09) | | **−.23 (.08)** |
| Other activity | | −.09 (.10) | | −.23 (.09)[†] | | .13 (.10) |
| Peer | | .17 (.11)[†] | | .26 (.18) | | .07 (.11) |
| Productive noninstruction | −.17 (.09)[†] | −.15 (.09) | −.17 (.11) | −.14 (.12) | −.13 (.10) | −.13 (.10) |
| Unproductive noninstruction | −.04 (.10) | −.06 (.10) | −.10 (.13) | −.14 (.13) | .05 (.10) | .06 (.10) |
| Teacher managed | .04 (.11) | .07 (.12) | .08 (.14) | .10 (.14) | −.05 (.11) | −.04 (.11) |
| Gender | .23 (.20) | .16 (.19) | .04 (.25) | −.04 (.24) | .004 (.17) | −.02 (.18) |
| Age | .12 (.27) | .10 (.27) | .27 (.32) | .25 (.33) | .13 (.27) | .09 (.27) |
| School/FRPL status (Ref is 1: 2% FRPL) | | | | | | |
| School 2: 61% | .08 (.19) | .11 (.18) | −.06 (.27) | −.06 (.28) | −.22 (.23) | −.19 (.23) |
| School 3: 69% | −.17 (.23) | −.19 (.22) | −.50 (.27)[†] | −.55 (.27)[†] | .05 (.24) | .09 (.23) |
| School 4: 72% | −.14 (.32) | −.09 (.32) | −.23 (.37) | −.17 (.39) | −.12 (.28) | −.12 (.28) |

*Note*. Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. EF = executive function; FRPL = free/reduced price lunch.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

Table 5
*Kindergarten Off-Task Behavior Predicting Growth in Academic Achievement From Kindergarten to First Grade*

| Variable | Passage comprehension | | Letter-word identification | | Applied problems | |
|---|---|---|---|---|---|---|
| Baseline math/literacy | **.36 (.08)**[a] | **.31 (.08)**[a] | **.73 (.09)**[a] | **.71 (.09)**[a] | **.53 (.08)**[a] | **.49 (.09)**[a] |
| Off-task total | **−.15 (.07)** | | **−.20 (.05)**[a] | | **−.22 (.10)** | |
| Nonengaged | | **−.23 (.05)**[a] | | −.14 (.09)[†] | | −.13 (.12) |
| Other activity | | −.12 (.09) | | −.15 (.07)[†] | | −.15 (.08)[†] |
| Peer | | .17 (.10)[†] | | .02 (.10) | | .15 (.10) |
| Productive noninstruction | −.09 (.11) | −.06 (.11) | .15 (.08) | .14 (.08) | −.07 (.10) | −.04 (.10) |
| Unproductive noninstruction | −.13 (.11) | −.15 (.08)[†] | −.16 (.08)[†] | **−.17 (.08)** | −.02 (.10) | −.05 (.10) |
| Teacher managed | .02 (.10) | .05 (.10) | .12 (.09) | .15 (.09)[†] | −.14 (.09) | −.14 (.09)[†] |
| Gender | −.10 (.16) | −.17 (.16) | −.24 (.14)[†] | −.28 (.14)[†] | **−.41 (.15)** | **−.45 (.14)**[a] |
| Age | .12 (.24) | .11 (.24) | **−.43 (.19)** | −.42 (.20)[†] | −.49 (.24) | .07 (.23) |
| School/FRPL status (Ref is 1: 2% FRPL) | | | | | | |
| School 2: 61% | −.39 (.24) | −.37 (.24) | −.15 (.22) | −.15 (.22) | −.50 (.24)[†] | −.49 (.24)[†] |
| School 3: 69% | **−.96 (.23)** | **−.98 (.24)** | −.42 (.20)[†] | **−.15 (.22)** | **−.71 (.24)**[a] | **−.76 (.24)**[a] |
| School 4: 72% | −.61 (.29)[†] | −.59 (.29)[†] | −.17 (.23) | −.15 (.24) | **−.90 (.29)**[a] | **−.88 (.27)**[a] |

*Note*. Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. FRPL = free/reduced price lunch.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

robust when variables were mean-centered at the classroom level. This finding is important in illuminating the manifestation of EF in the classroom, and how future work might capitalize on observations of behavior that are most reflective of poor EF skills. Second, none of the kindergarten off-task behaviors significantly predicted growth in EF from kindergarten to first grade. Third, off-task behavior as a unitary construct predicted fewer gains in literacy achievement when controlling for classroom covariates—although when children's behavior was adjusted for the off-task behavior of their peers (behavior was classroom mean-centered), this significance disappeared. Fourth, proportion of time spent off-task nonengaged predicted significantly fewer gains in reading comprehension. This result remained significant in the hierarchical linear model but was not significant in the classroom mean-centered model.

## Off-Task Behavior in Kindergarten

Children in this study, on average, spent less time off-task than children in prior studies that used time-sampling "snapshot" observational procedures (Fischer, Thiessen, Godwin, Kloos, & Dickerson, 2013; Godwin et al., 2016; Karweit & Slavin, 1981), as opposed to quantifying the total duration of time off-task (i.e., the current study; Connor et al., 2009). It remains unclear whether this difference is due to the difference in observational techniques, or to differences in characteristics of the samples. The present study did find higher rates of off-task behavior than past work measuring total duration of off-task behavior in the school day (McLean et al., 2016), although past work was conducted with first graders who are likely more accustomed to the behavioral expectations of school and have more practice exercising behavioral self-regulation in this setting, compared to kindergarteners.

Discrepancies in frequency of off-task behavior observed in the literature speak to the importance of considering factors such as grade level, teacher expectations, and other contextual factors that likely influence off-task behavior. For example, in the present sample, on average, the more time children spent in teacher-managed instruction the less time they spent off-task. This finding highlights the significance of considering trade-offs between child engagement/behavior, classroom management, and providing opportunities for autonomous learning. It is also important to note that we measured children's off-task behavior in a single school day, and not as an average duration across multiple days. Given the reported variability in off-task behavior throughout the school year (Godwin et al., 2016), future studies should aim to not only directly compare and contrast observation techniques (rating scale vs. total duration of time) but also assess whether the results of these methodologies remain stable throughout the school year.

## Types of Off-Task Behavior

A major focus of the present study was to investigate types of off-task behavior and determine how more fine-grained classification relates to children's EF versus measuring off-task behavior as a unitary construct. This study succeeded in creating a reliable observational measure of distinguishable categories of off-task behavior: going off-task and engaging in a distracting activity (off-task other-activity), going off-task and passively disengaging or wandering (off-task nonengaged), and going off-task by talking or interacting with a peer (off-task peer).

These behaviors, although correlated, differed across gender, and in their contribution to both EF and academic achievement outcomes. For example, in examining correlations, we found a significant gender difference in total proportion of time spent off-task; however, males were more off-task engaging in another activity, whereas females exhibited more off-task behavior engaging with their peers. These results replicated Godwin and colleagues' (2016) findings regarding off-task categorizations of peer off-task and off-task environmental distraction (a category similar to the current study's off-task other-activity behavior). We also found that off-task peer behavior was positively related to EF and achievement outcomes in kindergarten, which explains why off-task behavior as a unitary construct was not as highly correlated with baseline EF and achievement outcomes, seeing as this construct includes instances of peer off-task behavior. This suggests that not all off-task behaviors are necessarily "bad" or indicative of poor underlying EF capacities, and we should cautious in deciding whether to treat this behavior as a unitary construct. Although speculative, one potential explanation for why children who go off-task with their peers have higher baseline scores could be the classroom climate whereby peer off-task behavior is more "acceptable", and children—regardless of baseline EF—are more likely to go off-task with another person distracting them from their classroom assignment. Going off-task with a peer could also be tied to having more social competence, a skill that demonstrates positive associations with academic achievement in early childhood (Malecki & Elliot, 2002; Welsh, Parke, Widaman, & O'Neil, 2001). In contrast to proportion of time spent off-task with a peer, off-task nonengaged and off-task other activity were more highly and negatively associated with direct assessments of EF. We examine the implications of these relations in a later section.

## Off-Task Behavior and Executive Function

Another focus of the current study was determining whether off-task behavior in kindergarten reflected worse EF skills, and if further separating off-task behavior into different subtypes differentially predicted EF components. We found that in kindergarten, no EF components predicted off-task behavior as a unitary construct, yet there was evidence for differential predictions of EF components to specific off-task behaviors. For instance, a lower response inhibition score in the fall of kindergarten predicted a higher proportion of time spent off-task in another activity in the spring. This finding is plausible considering how the definition of going off-task with another activity (in the current study) is most reflective of motor impulsivity (compared to the other off-task behaviors). For example, when a child went off-task in another activity in our study, it often appeared as if they could not inhibit the desire to physically interact with other materials or games that seemed more appealing than the academic task at hand. The measure we used to capture response inhibition (HTKS; Ponitz et al., 2008) has high motoric demands, so this might be why off-task other-activity related more strongly to this EF component.

Lower working memory (and attentional control—in the classroom mean-centered models) predicted a higher proportion of time spent off-task nonengaged. Attentional control is unsurprisingly related to this passive disengagement and "wandering" behavior considering focused attention is more clearly lacking with this type of off-task behavior. However, it is not entirely clear why having poor working memory would also be associated with being nonengaged. One explanation derives from evidence that children with poor working memory have more difficulty beginning a task and sustaining attention during a task, which is notably different from children who engage in more hyperactive/impulsive off-task behaviors (Aronen, Vuontela, Steenari, Salmi, & Carlson, 2005; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008). This distinction is similar to the distinction we make between off-task nonengaged and off-task other activity behavior subtypes in the current study. Classroom measures of working memory have also been developed and align well with the current study's definition of off-task nonengaged. For example, the Working Memory Rating Scale (Alloway, Gathercole, & Kirkwood, 2008) is a teacher report that includes 20 descriptions of behaviors indicative of poor working memory that overlap off-task nonengaged behavior, such as lapses in attention. The Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000) also includes a working memory dimension that comprises behaviors such as initiating tasks, remembering how to do a task, and having materials organized. The BRIEF Working Memory subscale is similar to definitions of off-task nonengaged behaviors, such as students wandering around the room or taking longer to complete a task due to frequent instances of mind-wandering. This offers support for our findings regarding off-task nonengaged behaviors being most reflective of working memory and attentional control.

We found that kindergarten off-task behavior did not predict gains in EF from kindergarten to first grade. This is an important finding that illustrates that growth in EF may not be influenced by children's overt behavior in the classroom—and instead other classroom factors or individual child differences not captured in this study may be contributing to growth in this skill in early elementary school.

## Off-Task Behavior and Academic Achievement

In examining predictions of kindergarten off-task behavior to first grade academic achievement, we found that off-task behavior as a unitary construct predicted fewer gains in letter-word identification from kindergarten to first grade. This result remained stable even after controlling for baseline EF (see Appendix A). The kindergarten year is a crucial period for developing the foundational skills of reading, and a majority of literacy time in the classrooms from the current study was spent engaging in these more code-focused instructional activities. Thus, going off-task more generally may be taking away time from these activities which may lead to substantially fewer gains in letter-word identification from kindergarten to first grade. It is important to note, however, that off-task behavior as a unitary construct did not predict gains in literacy in the classroom mean-centered models. This suggests that the more off-task children were from their peers did not predict fewer gains in literacy achievement, only absolute percent of time off-task predicted gains in this domain.

The higher proportion of time spent off-task nonengaged predicted fewer gains in reading comprehension. Although this finding was also not robust to the classroom mean-centered models, it was significant in the hierarchical linear model. Considering how off-task nonengaged was more strongly related to working memory than with other EF components in the current study, it is unsurprising that this working memory-like behavior would predict reading comprehension, given prior work on the relation between working memory and reading comprehension, compared to print-related skills (Cain, Bryant, & Oakhill, 2004; McClelland et al., 2014; McVay & Kane, 2012). However, direct assessments of working memory are typically stronger predictors of math than literacy (Allan, Hume, Allan, Farrington, & Lonigan, 2014; Blair & Razza, 2007; Willoughby, Blair, Wirth, Greenberg, & The Family Life Project Investigators, 2012), so it is surprising that proportion of time spent off-task nonengaged predicted literacy but not math achievement. A potential explanation for why we see this lack of predictability for math might be due to the fact that prior unpublished work with this study's kindergarten sample found that there was far less time spent in math than literacy instruction. Considering previous studies have shown how math in kindergarten is often repetitive (Engel, Claessens, & Finch, 2013), growth in math achievement may be less dependent on classroom experiences more generally. Given we did not include off-task behavior by specific academic domain, however, future studies should explicitly assess patterns of off-task behavior within different types of instruction to acquire a more complete picture as to why we may be seeing these off-task behavior relation differences by academic domain.

## Classroom Effects

As discussed throughout this article, off-task behavior is a dynamic interplay between child and classroom factors—which is why we included classroom mean-centered variables as robustness checks. However, a limitation of these mean-centered models is that mean-centering child off-task variables is highly dependent on the off-task behavior of one's peers and how many children we coded per classroom. In the current study, we only coded an average of 7.2 children per classroom, which is not particularly high or representative of classroom with 20 to 25 total students. Mean-centering also minimizes variability between classrooms; a classroom where all children are highly off-task versus a classroom where children are rarely off-task will be treated similarly in the models (e.g., children will not differ from the mean), yet there might be systematic child-level differences in these classrooms that is diluted when attempting to account for classroom-level differences. In other words, it may be correct to assume that children who are uniformly off-task dictates something amiss with the classroom itself (e.g., poor classroom management), but it could also be the case that children in that classroom are not intrinsically different from one another (e.g., all have worse behavioral self-regulation; are prone to go off-task) regardless of sharing the same environment. Furthermore, with mean-centered models we cannot include the classroom-level covariates we collected because all children would have a score of zero for each variable (e.g., because each child would be at the classroom mean). But excluding these variables also removes important information about the classroom climate that could be influencing children's off-task behavior or capturing variance in our outcomes that child-level variables do not account for. However, in our regression models that do not classroom mean-center variables, we find little evidence for classroom-level variables soaking up any variance in our outcomes—with the exception that high proportion of time spent in teacher disorganization predicted marginally fewer gains in letter-word identification in first grade. This lack of variance accounted for by measured classroom variables should also raise caution to interpretations of our classroom mean-centered models; if variance in teacher organization and disorganization is not accounting for differences in outcomes, it is unclear which classroom attributes would be accounting for the differences we see in the mean-centered models since these models only capture the classroom as a whole, but do not generate specifics about the classroom that is soaking up variance in our models.

It is important to consider that our classroom-level covariates are mutually exclusive with children's off-task behavior (except for teacher-managed instruction); which suggests a potential dependency of one variable on the other that could limit the ability of one variable to truly and independently predict an outcome. Although, given the wide range of variables present in the ISI coding system (close to 40 different instructional and noninstructional variables), it is unlikely that a small amount of time in one variable will automatically denote more time in a different variable (e.g., the coding system is not bivariate). Researchers interested in the dynamic interplay between classroom context and child characteristics should compare how mean-centering child variables versus including classroom-level covariates contributes to student outcomes. The current study illuminated how results and subsequent interpretations can change depending on which analytic method is used, so future work should be cautious in not only choosing analytic methods, but also in conducting the appropriate sensitivity analyses.

When examining zero-order correlations of the classroom variables to student variables, we also found that few associations between classroom factors and off-task behavior emerged. None of the productive or unproductive noninstructional activities significantly related to off-task behavior. This is surprising considering how teacher planning and organizing (productive noninstruction) might help decrease off-task behavior, whereas proportion of time spent waiting for the teacher or in disruptions (unproductive noninstruction) might breed more off-task behavior. We did, however, find a consistent negative relation between proportion of time spent in teacher-managed instruction and off-task behavior. This makes sense theoretically, considering that children may be less likely to be distracted if a teacher is managing their attention versus if they are working on an instructional activity on their own. Past work, however, has focused on differences between teacher-managed versus child-managed productive and nonproductive noninstructional activities (Day et al., 2015), which the current study did not undertake. With a larger sample, future work should focus on combinations of management and activity (e.g., teacher-managed productive noninstruction) as they relate to instances of off-task behavior. The current study also did not include changes in proportion of time spent in these noninstructional activities over time, which past work has illustrated is more predictive of student outcomes than a single observational time point (Connor et al., 2010).

Although not statistically significant across our sensitivity models, higher proportion of time spent in unproductive noninstructional activities predicted fewer gains in first grade letter-word identification scores and marginally fewer gains in first grade reading comprehension scores. This aligns with prior findings for literacy achievement (McLean et al., 2016). It is intriguing, however, that classroom variables such as proportion of time spent listening to directions/planning from the teacher (productive noninstruction) and proportion of time spent waiting for the teacher to start a lesson or experiencing a classroom disruption (unproductive noninstruction) would only impact achievement gains in literacy but not math. This might again be due to less time spent in math instruction (compared with literacy) in kindergarten, so when instructional time is lost (higher proportion of time spent in noninstruction), literacy instruction may take a bigger hit than math. Future studies should continue to include noninstructional activity factors in the study of achievement gains, to help decipher why some achievement domains may be more influenced by classroom experiences than others.

Overall, it will be important in future work to examine the interactions between classroom and child on resulting off-task behavior and subsequent executive function and academic achievement gains For example, in Nesbitt et al. (2015) they found that children's disengagement from learning opportunities (i.e., unoccupied, disruptive, or in time-out)—a similar construct to the present study's operationalization of off-task behavior—partially mediated the relation between EF skills in the fall of pre-K and gains in both literacy and mathematics at the end of pre-K. The present findings are consistent with those results and further demonstrate that the nature of the off-task behavior matters in predicting different academic gains. Future work should continue to examine different methodologies that address the influence of classroom context on child behavior and subsequent outcomes.

## Limitations

First, although the ISI coding system is a comprehensive measure, children were only observed for a single school day in kindergarten. Past studies have shown how various classroom factors change across the school year and that change in proportion of time spent in classroom factors from fall to spring is more important than total proportion of time spent at any given time point (Connor et al., 2010; Godwin et al., 2016). This may help explain why we did not see any significant predictions of productive and unproductive noninstruction to important student outcomes in this study. It was also up to each teacher's discretion when and for how long they wanted to be videorecorded, which produces inconsistency in not only length of observations between classrooms, but what time of day (e.g., morning vs. afternoon) these observations took place. Overall, this limits the generalizability of our findings.

Second, there are a few issues with the study sample. The rate of off-task behavior in the present sample is not normally distributed, and there are only a handful of children displaying high rates of off-task behavior; a majority of children in the current sample are off-task less than 15% of the time. Thus, despite emerging relations with student outcomes, the present findings might only be relevant to a subsample of students. However, when we excluded children two standard deviations above the mean, all predictions remained stable ($n = 7$). Another limitation related to the sample is that there was a high rate of missing data for first grade academic achievement outcomes (30%), thus a large amount of data was imputed for our predictions to first grade outcomes.

Additionally, it is also important to consider how there might be unobserved variables that contribute to children's off-task behavior that the current study did not capture. For example, we did not include variables assessing the difficulty of the instructional task, or how much time was spent in games or play/free-time, which past studies have demonstrated relations to attention, engagement and off-task behavior (Prykanowski, Martinez, Reichow, Conroy, & Huang, 2018; Timmons et al., 2015). Lastly, off-task behavior might also be very different in kindergarten classrooms where the structure of instruction and behavior expectations are typically more relaxed than with later grades, thus future work should replicate this study with first and second graders to see whether this pattern of findings holds.

# Conclusion

The current study offers a comprehensive investigation of off-task behavior in the kindergarten classroom and resulting student outcomes. We found that off-task behavior in the classroom related to individual differences in EF, yet separate EF components differentially predicted subtypes of off-task behavior (and not off-task behavior as a unitary construct). Future work should focus on these subcategorizations, and how their contribution to EF and academic achievement change over time. Specifically, researchers studying EF in the classroom should pay particular attention to instances of off-task nonengaged, and what teachers can do to address this behavior. Overall, this work sheds light on the importance of measuring EF in context, and how improved understanding of the manifestation of EF in the classroom—and resulting behavioral self-regulation—can help us develop more appropriate interventions.

# References

Abikoff, H., Gittelman, R., & Klein, D. F. (1980). Classroom observation code for hyperactive children: A replication of validity. *Journal of Consulting and Clinical Psychology, 48,* 555–565. http://dx.doi.org/10.1037/0022-006X.48.5.555

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

Alloway, T. P., Gathercole, S. E., & Kirkwood, H. J. (2008). *Working Memory Rating Scale*. London, UK: Pearson Assessment.

Aronen, E. T., Vuontela, V., Steenari, M. R., Salmi, J., & Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance at school. *Neurobiology of Learning and Memory, 83,* 33–42. http://dx.doi.org/10.1016/j.nlm.2004.06.010

Atkins, M. S., & Pelham, W. E. (1991). School-based assessment of attention deficit-hyperactivity disorder. *Journal of Learning Disabilities, 24,* 197–204. http://dx.doi.org/10.1177/002221949102400403

Austin, J. L., & Soeda, J. M. (2008). Fixed-time teacher attention to decrease off-task behaviors of typically developing third graders. *Journal of Applied Behavior Analysis, 41,* 279–283. http://dx.doi.org/10.1901/jaba.2008.41-279

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. http://dx.doi.org/10.1111/j.1467-8624.2007.01019.x

Bohlmann, N. L., & Downer, J. T. (2016). Self-regulation and task engagement as predictors of emergency language and literacy skills. *Early Education and Development, 27,* 18–37. http://dx.doi.org/10.1080/10409289.2015.1046784

Bohn, C. M., Roehrig, A. D., & Pressley, M. (2004). The first few days of school in the classroom of two more effective and four less effective primary-grade teachers. *The Elementary School Journal, 104,* 269–287. http://dx.doi.org/10.1086/499753

Cain, K. E., Bryant, P. E., & Oakhill, J. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42. http://dx.doi.org/10.1037/0022-0663.96.1.31

Chafouleas, S. M., McDougal, J. L., Riley-Tillman, C. R., Panahon, C. J., & Hilt, A. M. (2005). What do daily behavior report cards (DBRCs) measure? An initial comparison of DBRCs with direct observation for off-task behavior. *Psychology in the Schools, 42,* 669–676. http://dx.doi.org/10.1002/pits.20102

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., . . . Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38,* 85–99. http://dx.doi.org/10.3102/0013189X09332373

Connor, C. M., Ponitz, C. C., Phillips, B. M., Travis, Q. M., Glasney, S., & Morrison, F. J. (2010). First graders' literacy and self-regulation gains: The effect of individualizing student instruction. *Journal of School Psychology, 48,* 433–455. http://dx.doi.org/10.1016/j.jsp.2010.06.003

Day, S. L., Connor, C. M., & McClelland, M. M. (2015). Children's behavioral regulation and literacy: The impact of the first grade classroom environment. *Journal of School Psychology, 53,* 409–428. http://dx.doi.org/10.1016/j.jsp.2015.07.004

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

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis, 35,* 157–178. http://dx.doi.org/10.3102/0162373712461850

Farran, D. C., & Son-Yarbrough, W. (2001). Title I funded preschools as a developmental context for children's play and verbal behaviors. *Early Childhood Research Quarterly, 16,* 245–262. http://dx.doi.org/10.1016/S0885-2006(01)00100-4

Fischer, A., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention: Evidence from a new paradigm. *Journal of Experimental Child Psychology, 114,* 275–294. http://dx.doi.org/10.1016/j.jecp.2012.07.006

Gaastra, G. F., Groen, Y., Tucha, L., & Tucha, O. (2016). The effects of classroom interventions on off-task and disruptive classroom behavior in children with symptoms of attention-deficit/hyperactivity disorder: A meta-analytic review. *PLoS ONE, 11*(2), e0148841. http://dx.doi.org/10.1371/journal.pone.0148841

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance on laboratory analogues of classroom activities. *Applied Cognitive Psychology, 22,* 1019–1037. http://dx.doi.org/10.1002/acp.1407

Gill, P., & Remedios, R. (2013). How should researchers in Education operationalize on-task behaviours? *Cambridge Journal of Education, 43,* 199–222. http://dx.doi.org/10.1080/0305764X.2013.767878

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior rating inventory of executive function*. Odessa, FL: Psychological Assessment Resources.

Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction, 44,* 128–143. http://dx.doi.org/10.1016/j.learninstruc.2016.04.003

Griggs, M. S., Mikami, A. Y., & Rimm-Kaufman, S. E. (2016). Classroom quality and student behavior trajectories in elementary school. *Psychology in the Schools, 53,* 690–704. http://dx.doi.org/10.1002/pits.21941

Junod, R. V., Dupaul, G. J., Jitendra, A. K., Volpe, R. J., & Cleary, K. S. (2006). Classroom observations of students with and without ADHD: Differences across types of engagement. *Journal of School Psychology, 44,* 87–104. http://dx.doi.org/10.1016/j.jsp.2005.12.004

Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal, 18,* 157–171. http://dx.doi.org/10.3102/00028312018002157

Kilian, B., Hofer, M., Fries, S., & Kuhnle, C. (2010). The conflict between on-task and off-task actions in the classroom and its consequences for motivation and achievement. *European Journal of Psychology of Education, 25,* 67–85. http://dx.doi.org/10.1007/s10212-009-0007-8

Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., & Wilson, S. J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: A comparative evaluation of the educational relevance of selected measures. *Journal of Educational Psychology, 109,* 1084–1102.

Magi, K., Mannamaa, M., & Kikas, E. (2016). Profiles of self-regulation in elementary grades: Relations to math and reading skills. *Learning and Individual Differences, 51,* 37–48.

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17,* 1–23. http://dx.doi.org/10.1521/scpq.17.1.1.19902

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . Howes, C. (2008). Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills. *Child Development, 79,* 732–749.

Mather, N., & Woodcock, R. W. (2001). *Examiner's manual for the Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

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. http://dx.doi.org/10.1016/j.ecresq.2006.09.003

McClelland, M. M., & Cameron, C. E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development, 2011,* 29–44. http://dx.doi.org/10.1002/cd.302

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. http://dx.doi.org/10.1037/0012-1649.43.4.947

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology, 5,* 599. http://dx.doi.org/10.3389/fpsyg.2014.00599

McCoy, D. C. (2019). Measuring young children's executive function and self-regulation in classrooms and other real-world settings. *Clinical Child and Family Psychology Review*. Advance online publication. http://dx.doi.org/10.1007/s10567-019-00285-1

McLean, L., Sparapani, N., Toste, J. R., & Connor, C. M. (2016). Classroom quality as a predictor of first graders' time in non-instructional activities and literacy achievement. *Journal of School Psychology, 56,* 45–58. http://dx.doi.org/10.1016/j.jsp.2016.03.004

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141,* 302–320. http://dx.doi.org/10.1037/a0025250

Nesbitt, K. T., Farran, D. C., & Fuhs, M. W. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology, 51,* 865–878. http://dx.doi.org/10.1037/dev0000021

Noldus Information Technology. (2013). The observer Video-Pro: Interactive multimedia tutorial (Version 13.0) [Computer software]. Leesburg, VA: Noldus Information Technology, Inc.

Ocumpaugh, J., Baker, R. S. J. d., & Rodrigo, M. M. T. (2012). *Baker–Rodrigo Observation Method Protocol (BROMP)*. New York, NY: EdLab.

Pianta, R. C., & Rimm-Kaufman, S. (2006). The social ecology of the transition to school: Classrooms, families, and children. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 490–507). Malden, MA: Blackwell Publishing. http://dx.doi.org/10.1002/9780470757703.ch24

Platzman, K. A., Stoy, M. R., Brown, R. T., Coles, C. D., Smith, I. E., & Falek, A. (1992). Review of observation methods in attention deficit hyperactivity disorder (ADHD): Implications for diagnosis. *School Psychology Quarterly, 7,* 155–177. http://dx.doi.org/10.1037/h0088258

Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23,* 141–158. http://dx.doi.org/10.1016/j.ecresq.2007.01.004

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., . . . Woo, D. (2001). A study of effective first grade literacy instruction. *Scientific Studies of Reading, 15,* 35–58. http://dx.doi.org/10.1207/S1532799XSSR0501_2

Prykanowski, D. A., Martinez, J. R., Reichow, B., Conroy, M. A., & Huang, K. (2018). Measurement of young children's engagement and problem behavior in early childhood settings. *Behavioral Disorders, 44,* 1–10. http://dx.doi.org/10.1177/0198742918779793

Rimm-Kaufman, S., Le Para, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Elementary School Journal, 105,* 377–394. http://dx.doi.org/10.1086/429948

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45,* 958–972.

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York, NY: Guilford Press.

Timmons, K., Pelletier, J., & Corter, C. (2015). Understanding children's self-regulation within different classroom contexts. *Early Child Development and Care, 186,* 249–267. http://dx.doi.org/10.1080/03004430.2015.1027699

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition (WISC-III).*, San Antonio, TX: The Psychological Corporation.

Welsh, M., Parke, R. D., Widaman, K., & O'Neil, R. (2001). Linkages between children's social and academic competence: A longitudinal analysis. *Journal of School Psychology, 39,* 463–482. http://dx.doi.org/10.1016/S0022-4405(01)00084-X

Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., & The Family Life Project Investigators. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment, 24,* 226–239.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery- Revised.* Chicago, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *The Woodcock-Johnson Tests of Achievement—Third Edition (WJ—III).* Itasca, IL: Riverside Publishing Company.

# Appendix A

## Off-Task Behavior Predicting Academic Achievement: Controlling for Executive Function

Table A1

*Controlling for Kindergarten EF: Kindergarten Off-Task Behavior Predicting First Grade Academic Achievement*

| Variable | Passage comprehension | | Letter-word identification | | Applied problems | |
|---|---|---|---|---|---|---|
| Baseline math/literacy | **.25 (.09)** | **.25 (.09)** | **.63 (.09)[a]** | **.63 (.09)[a]** | **.34 (.10)[a]** | **.34 (.09)[a]** |
| Off-task total | **−.17 (.05)[a]** | | **−.22 (.04)[a]** | | **−.12 (.08)** | |
| Nonengaged | | −.07 (.11) | | −.05 (.09) | | .03 (.14) |
| Other activity | | −.12 (.09) | | −.14 (.07)[†] | | −.17 (.09)[†] |
| Peer | | −.03 (.11) | | −.07 (.10) | | −.03 (.12) |
| K EF (RI) | .12 (.08) | .08 (.10) | .10 (.07) | .07 (.09) | .14 (.09) | .08 (.10) |
| K EF (AC) | .06 (.08) | .06 (.08) | −.002 (.07) | .003 (.07) | .11 (.07) | .12 (.07) |
| K EF (WM) | **.28 (.08)[a]** | **.28 (.09)[a]** | .14 (.07)[†] | .15 (.07)[†] | **.21 (.08)** | **.24 (.09)** |
| Gender | −.18 (.15) | −.18 (.16) | **−.31 (.14)** | −.30 (.15)[†] | **−.44 (.13)[a]** | **−.42 (.13)[a]** |
| Age | −.14 (.23) | −.13 (.23) | **−.54 (.20)** | **−.53 (.21)** | −.14 (.20) | −.13 (.20) |
| School/FRPL status (Reference is 1: 2% FRPL) | | | | | | |
| School 2: 61% | −.28 (.21) | −.28 (.21) | −.15 (.22) | −.16 (.18) | **−.43 (.20)** | **−.43 (.19)** |
| School 3: 69% | **−.72 (.20)[a]** | **−.74 (.21)[a]** | −.34 (.18)[†] | −.35 (.18)[†] | **−.60 (.22)** | **−.63 (.22)** |
| School 4: 72% | −.28 (.21) | −.28 (.20) | −.03 (.19) | −.03 (.19) | **−.70 (.23)[a]** | **−.69 (.22)[a]** |

*Note.* Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. EF = executive function; FRPL = free/reduced price lunch.

[a] Values remain significant after Bonferroni correction.

[†] $p < .10$.

(*Appendices continue*)

## Appendix B

## Robustness Check: Models with Classroom Mean-Centered Variables

Table B1

*Classroom Mean-Centered Variables: Fall Kindergarten EF Components Predicting Spring Kindergarten Off-Task Behavior*

| Variable | Total off-task | Off-task nonengaged | Off-task other activity | Off-task peer |
|---|---|---|---|---|
| Response inhibition | −.06 (.06) | .10 (.05) | **−.44 (.11)**[a] | .15 (.08)[†] |
| Attentional control | −.04 (.02) | **−.17 (.06)**[a] | .03 (.09) | 008 (.05) |
| Working memory | .07 (.05) | **−.21 (.08)**[a] | .13 (.06) | .14 (.07)[†] |
| Gender | **−.24 (.12)** | **−.34 (.11)**[a] | **−.17 (.09)** | .16 (.11)[†] |
| Age | .18 (.18) | .22 (.18) | .25 (.16) | .09 (.08) |
| School/FRPL status (Reference is 1: 2% FRPL) | | | | |
| School 2: 61% | −.00 (.06) | −.00 (.06) | −.00 (.06) | −.00 (.07) |
| School 3: 69% | .00 (.07) | .00 (.07) | −.00 (.07) | .00 (.07) |
| School 4: 72% | .00 (.07) | .00 (.07) | .00 (.07) | .00 (.07) |

*Note*. Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. EF = executive function; FRPL = free/reduced price lunch.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

Table B2

*Classroom Mean-Centered Variables: Kindergarten Off-Task Behavior Predicting Growth in EF from Kindergarten to First Grade*

| Variable | K attention control | | K working memory | | K response inhibition | |
|---|---|---|---|---|---|---|
| Baseline EF | **.37 (.10)**[a] | **.32 (.11)** | **.54 (.12)**[a] | **.48 (.14)**[a] | **.42 (.10)**[a] | **.43 (.12)**[a] |
| Off-task total | −.008 (.09) | | −.03 (12) | | .06 (.07) | |
| Nonengaged | | −.16 (.14) | | −.04 (.18) | | −.11 (.10) |
| Other activity | | −.08 (.10) | | −.26 (.18) | | .11 (.10) |
| Peer | | .24 (.13)[†] | | .23 (.18) | | .12 (.11) |
| Gender | .20 (.21) | .12 (.20) | .07 (.24) | −.02 (.23) | .03 (.19) | .003 (.19) |
| Age | .33 (.31) | .32 (.31) | .29 (.36) | .26 (.37) | .21 (.38) | .19 (.38) |
| School/FRPL status (Reference is 1: 2% FRPL) | | | | | | |
| School 2: 61% | .00 (.12) | .00 (.13) | .00 (.17) | −.00 (.17) | −.00 (.13) | .00 (.13) |
| School 3: 69% | .00 (.13) | .00 (.13) | .00 (.17) | −.00 (.17) | −.00 (.13) | −.00 (.13) |
| School 4: 72% | −.00 (.16) | .00 (.16) | .00 (.22) | −.00 (.22) | −.00 (.16) | −.00 (.16) |

*Note*. Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. EF = executive function; FRPL = free/reduced price lunch.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

(*Appendices continue*)

Table B3
*Classroom Mean-Centered Variables: Kindergarten Off-Task Behavior Predicting Growth in Academic Achievement From Kindergarten to First Grade*

| Variable | First passage comprehension | | First letter-word identification | | First applied problems | |
|---|---|---|---|---|---|---|
| Baseline EF | **.34 (.09)[a]** | **.27 (.10)** | **.77 (.10)[a]** | **.73 (.11)[a]** | **.56 (.09)[a]** | **.51 (.10)[a]** |
| Off-task total | −.10 (.07) | | **−.14 (.05)** | | .01 (.08) | |
| Nonengaged | | **−.25 (.10)** | | −.13 (.10) | | −.05 (.13) |
| Other activity | | −.14 (.09) | | **−.16 (.07)** | | −.11 (.08) |
| Peer | | **.22 (.10)** | | .03 (.10) | | .17 (.10) |
| Gender | −.08 (.17) | −.17 (.17) | −.27 (.15)[†] | −.31 (.15)[†] | **−.46 (.15)** | **−.50 (.16)[b]** |
| Age | .19 (.27) | .17 (27) | −.28 (.24) | −.28 (.24) | .25 (.24) | .23 (.23) |
| School/FRPL status (Ref is 1: 2% FRPL) | | | | | | |
| School 2: 61% | .00 (.12) | −.00 (.12) | .00 (.10) | .00 (.10) | −.00 (.11) | .00 (.11) |
| School 3: 69% | .00 (.11) | −.00 (.11) | .00 (.10) | .00 (.10) | −.00 (.11) | .00 (.11) |
| School 4: 72% | .00 (.14) | −.00 (.14) | .00 (.13) | .00 (.13) | −.00 (.13) | −.00 (.13) |

*Note*. Coefficients are standardized. Standard errors are clustered at the classroom level. Values in boldface type are statistically significant. EF = executive function; FRPL = free/reduced price lunch.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

# Appendix C

## Robustness Check: Models Using Hierarchical Linear Modeling

Table C1
*Hierarchical Linear Model: Kindergarten Off-Task Behaviors Predicting First Grade EF and Academic Achievement*

| Variable | WM | RI | AC | LWID | AP | PC |
|---|---|---|---|---|---|---|
| Baseline score | **.49 (.13)[a]** | **.44 (.11)[a]** | **.34 (.10)[a]** | **.72 (.08)[a]** | **.50 (.08)[a]** | **.72 (.08)[a]** |
| Nonengaged | .03 (.17) | −.14 (.11) | −.09 (.12) | −.12 (.09) | −.11 (.10) | **−.22 (.08)[a]** |
| Other activity | **−.25 (.12)** | .12 (.11) | −.09 (.10) | **−.16 (.07)** | −.16 (.08)[†] | **−.16 (.07)** |
| Peer | .20 (.17) | .09 (.13) | **.24 (.12)** | .03 (.09) | .13 (.10) | .03 (.09) |
| Productive noninstruction | −.15 (.14) | −.13 (.10) | −.13 (.11) | −.04 (.09) | −.009 (.09) | −.04 (.09) |
| Unproductive noninstruction | −.07 (.12) | .04 (.10) | −.06 (.10) | −.04 (.07) | −.02 (.08) | −.05 (.07) |
| Teacher managed | .05 (.13) | −.01 (.10) | .04 (.10) | .01 (.08) | −.02 (.10) | .01 (.08) |
| Gender | .003 (.22) | −.01 (.18) | .13 (.18) | **−.43 (.19)** | **−.47 (.14)[a]** | **−.29 (.13)** |
| Age | .17 (.30) | .10 (.24) | .09 (.24) | **−.29 (.13)** | .05 (.20) | **−.43 (.19)** |
| School *SD* | .00 | .00 | .00 | .00 | .26 | .00 |
| Classroom *SD* | .00 | .00 | .00 | .02 | .003 | .02 |
| Residual *SD* | 1.13 | .91 | .90 | .65 | .73 | .65 |

*Note*. Betas are standardized. All assessments were collected in first grade. Values in boldface type are statistically significant. Baseline score = Baseline score of predicted outcome (e.g., WM); WM = working memory (executive function component); RI = response inhibition (executive function component); AC = attentional control (executive function component); LWID = letter-word identification; AP = applied problems; PC = passage comprehension.
[a] Values remain significant after Bonferroni correction.
[†] $p < .10$.

Received August 7, 2018
Revision received June 6, 2019
Accepted June 12, 2019 ■