Do Teacher Reports of Executive Functions Predict Reading Development? Evidence from a Nationally Representative Sample

Andrew Weaver

Published August 14th 2023

Do Teacher Reports of Executive Functions Predict Reading Development? Evidence from a Nationally Representative Sample

Abstract

This study explores whether teacher reports of executive functions predict change in reading performance (i.e., reading development) for elementary-aged students when controlling for direct assessments of executive functions and for teacher reports of students' literacy skills. Prior research has raised problems with the construct validity of teacher reports of executive functions but has yet to consider that these teacher reports might be related to teachers' perceptions of their students' literacy skills. The current study used Grades 3 through 5 data from nationally representative data (N = 6,945) of students collected between 2014 and 2016 to examine the contributions of teacher reports of executive functions to change in reading performance over the course of a year with autoregressive structural equation models. Measures of executive functions tapped attentional focusing (in Grades 3 and 4), working memory (in Grade 3), and inhibitory control (in Grade 4). When controlling for a direct assessment of the same facet of executive function as the teacher report, the teacher report of executive function predicted next year's reading. However, controlling for a teacher report of students' literacy skills reduced the effect of teacher reports of executive functions to nearly 0 across models while not reducing the effect of direct assessments of executive functions. This finding held across student race and home language subgroups in multigroup analyses. Based on these findings, teacher reports of executive functions do not capture information about executive functions that predicts of reading development beyond the teachers' perceptions of their students' literacy skills. Further research is needed to determine how teacher reports of EF could be designed to capture EFs as applied to reading.

Do Teacher Reports of Executive Functions Predict Reading Development? Evidence from a Nationally Representative Sample

Successful reading requires the coordination and strategic application of multiple skills (Scarborough, 2001). Recent models of reading highlight the necessity of domain-general cognitive processes in coordinating and applying reading skills in varied contexts (Butterfuss & Kendeou, 2018; Duke & Cartwright, 2019; Kim, 2017). Indeed, recent years have seen a growing interest in the role of executive functions (EFs), which are a set of higher-order cognitive processes that facilitate planning, problem solving, and the initiation and maintenance of goal-directed behavior (Pennington & Ozonoff, 1996). Research has found that EFs are associated with reading comprehension and basic word reading processes in English both within and across grade (Christopher et al., 2012; Cirino et al., 2019; Follmer, 2018; Kieffer et al., 2013; Kieffer & Christodoulou, 2020; Spencer et al., 2019; Zhang & Peng, 2023), with questions of both development and measurement that have yet to be addressed.

This growing body of research has led to instructional implications and policy decisions with direct consequences for reading teachers. Some large districts have recently adopted the required assessment of student EFs by teachers (Amin, 2021). For example, New York City began mandating the use of the Devereux Student Strengths Assessment (DESSA; LeBuffe et al., 2009), a social-emotional screener including indicators tapping EF related “goal-directed behavior” skills in 2021 following an $18 million contract. Instructional guidelines have been published with recommendations for how reading teachers can synthesize this research into their literacy teaching practice (e.g., Cartwright, 2023). All these policy and instructional prescriptions hinge on a solid research base about EFs and reading development. In this study, development is operationalized as the change in reading performance from year to year, specifically from the end of Grade 3 to the end of Grade 4, and from the end of Grade 4 to the end of Grade 5. Yet executive function measurement poses challenges for understanding how these important skills relate to reading instruction and development.

Reading researchers have most commonly measured EFs in elementary-aged children using direct assessment cognitive tasks (Follmer, 2018), but more recent attention has been given to reports that can be completed by a parent or teacher. Direct assessments are given to a child by a trained administrator (either a researcher or teacher) and attempt to capture the cognitive ability in a given EF skill. For example, a common working memory paradigm asks children to remember a string of letters or numbers, adding an extra digit or letter after each successful attempt until children fail to recall the string (e.g., Woodcock et al., 2001). By contrast, a parent or teacher report of EF might ask whether or to what extent a set of indicators match the behavior of a focal child. For example, a common teacher report of working memory asks teachers to rate how often a child loses track of what they are doing while working using a scale ranging from "Never" to "Often." (Gioia et al., 2000). Both measurement types are correlated with reading (Follmer, 2018), but two reviews of studies have found direct assessments of EFs and teacher reports of EFs do not consistently correlate with one another (Isquith et al., 2013; Toplak et al., 2013). Yet relatively little research has tested whether teacher reports of EFs make unique contributions to reading while controlling for direct assessments of EFs. "Unique" in this context means that teacher reports of EFs make a separate statistical contribution to reading when controlling for other predictors like direct assessments of EFs in the same model. If teacher reports of EFs make unique contributions to reading when controlling for direct assessment of EFs, then both could be useful as distinct and complementary measurements of EFs.

Yet some EF measurement researchers have argued that the low correlation between direct assessments of EFs and teacher reports of EFs may indicate that teacher reports of EFs measure other constructs than just EFs (e.g., McAuley et al., 2010). Such “task impurity” poses a serious validity issue for teacher reports of EFs, as these reports may not actually capture the construct (EFs) they are purported to measure. For reading teachers, reports of EF undoubtedly include teachers’ perceptions of their students’ literacy skills. Yet if these reports are not adding information about reading development beyond those perceptions of literacy skills, it is not clear what their utility is for predicting reading development. In turn, it is not clear what instructional implications for reading teachers should be derived from research using teacher reports of EFs to predict reading development. Additionally, research suggests that teachers may systematically underrate Black students, boys, and students identified as English Learners (Garcia et al., 2019). Thus, researchers need to test if teacher reports of EFs predict reading when controlling for a teacher report of students’ literacy skills.

The current study examines whether teacher reports of EFs predict change in reading achievement when controlling for direct assessments of EFs and teacher perceptions of literacy competency. The study utilizes a large, nationally representative dataset, the Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011 (ECLS-K:2011; Tourangeau et al., 2019), to examine these relations for a cohort of children entering Kindergarten in 2010, giving the study results generalizability and yielding statistical power to detect very small effects. This study uses data from Grades 3 through Grade 5 to test autoregressive structural equation models examining the contributions of teacher reports of EFs to reading development (i.e., predicting reading performance while controlling for reading performance measured the year prior) for elementary-aged children. Autoregressive models more rigorously test whether teacher reports of

EFs predict reading development as opposed to cross-sectional levels of reading. The predictors of interest are all measured in the spring of either Grade 3 or Grade 4 (depending on the model), with the reading outcome measured the following spring. The model first looks at the teacher report of EF as the only predictor of interest. Then, a direct assessment of the same facet of EF is introduced as a control to see whether both assessment types measure the same information about EFs. The final step additionally controls for teacher perceptions of literacy competency to test whether teacher reports of EF remain predictive and contribute additional variance beyond perceptions of literacy skills. The models are tested separately for three EF skills (attentional focusing, working memory, and inhibitory control) using one teacher rating and direct EF assessment for a given model. Doing so avoids potential issues with collinearity and provides the best opportunity to find evidence of the counter hypothesis that teacher rating scales of EFs do predict reading development alongside reports of literacy competency. Finally, multigroup analyses are used to see if there are differences in findings depending on student home language, gender, or racial/ethnic background.

Role of Executive Functions in Reading

Recent theories of reading comprehension have incorporated an explicit role for EFs (e.g., Duke & Cartwright, 2019), particularly in coordinating componential reading skills and maintaining reading activities in busy environments. EFs are often conceived of as three unified, yet distinct skills (Miyake et al., 2000): (1) attention shifting (or cognitive flexibility or set shifting) describes the ability to shift between tasks or rules; (2) working memory is the ability to hold and manipulate pieces of information in the mind (Baddeley & Hitch, 1974); and (3) inhibitory control describes the ability to withhold a dominant response (Miyake et al., 2000). In a review of research on reading comprehension and EFs, Butterfuss and Kendeou (2018)

highlight implicit roles of EFs present in several models of reading comprehension. The recent Deploying Reading in Varied Environments (DRIVE; Duke & Cartwright, 2019) model of reading utilizes a metaphor of driving to argue that the reading process involves an active reader coordinating many skills in different sociocultural environments to achieve a reading goal. Importantly, the DRIVE model posits that readers, like drivers, must manage many necessary processes using attention shifting (called cognitive flexibility), inhibitory control, and working memory. Thus, this model centers EFs as vitally important for managing many skills and behaviors when engaging in reading, particularly in the varied contexts children inhabit.
EFs might contribute to successful reading both in facilitating the componential reading processes and in regulating behavior during reading activities.

Empirical findings support the importance of EFs for reading skills, including English reading comprehension and word reading (Cutting et al., 2009; Christopher et al., 2012; Cirino et al., 2019; Follmer, 2018; Kieffer et al., 2013; Locascio et al., 2010; Sesma et al., 2009; Spencer et al., 2019). EFs may play a different role in reading for other languages and orthographies (Chung et al., 2018), but this is beyond the scope of the current study. For behavioral regulation, Duke and Cartwright (2019) highlight how inhibitory control might be necessary to ignore distracting stimuli such as ambient noise during reading. Kieffer, Vukovic, and Berry (2013) argued that classroom environments may need to be structured to maximize attention for students receiving reading interventions. Thus, while much of the empirical work on reading and EFs links EFs to specific components of reading, both research and theory suggest that EF processes ought to be important for behavioral regulation as well. The distinction between EFs as important for componential reading skills and EFs as important for behavioral regulation align with arguments about what direct assessments of EFs and teacher reports of EFs measure.

Although the importance of EF for reading is well documented, low correlations between direct assessments (i.e., assessments given directly to focal children) and EF reports (i.e., reports about a focal child completed by a parent or teacher) are often observed, even when measuring the same component of EF (Isquith et al., 2013; McAuley et al., 2010; Toplak et al., 2013). For example, one review of studies reported a median correlation of only $r = .19$ for direct assessments of EFs and EF reports (Toplak et al., 2013). Many researchers have interpreted these results to indicate that direct assessments of EF capture a purer form of EF in a laboratory setting, while parent or teacher reports of EFs capture more ecologically valid information about EF skills in everyday contexts (Anderson et al., 2002; Fuhs et al., 2015; Ten Eycke & Dewey, 2016). Indeed, some researchers argue that both forms of measurement have utility by providing two levels of assessment: specific EF components defined by direct assessments and the everyday manifestations of these components captured by parent or teacher reports (Isquith et al., 2013). In other words, a direct assessment might be conceived of as a measure of an EF skill in a quiet, uninterrupted testing environment, while a parent or teacher report taps how an EF skill is applied in daily contexts.

Despite the argument that each type of EF measures complementary aspects of EFs, studies using teacher reports of EFs to predict reading while controlling for direct assessments of EFs in the same model are rare. Studies on children ranging from preschool- to highschool-aged that have included both types in models find that parent or teacher reports of EFs make unique contributions to literacy skills when controlling for direct assessments of EFs (Dekker et al., 2017; Fuhs et al., 2015; Gerst et al., 2017; Ten Eycke & Dewey, 2016). While these previous studies cover a wide age range, only one (Fuhs et al., 2015) takes a developmental approach. In addition, all these studies do not examine reading specifically, but also include outcomes in other academic domains, such as mathematics. Finally, none of these studies have been conducted with nationally representative samples, limiting generalizability of their findings. Furthermore, these studies leave questions remaining about what else these teacher reports might be measuring.

Potential Issues with Using Teacher Reports of Executive Functions to Predict Reading

While prior work has highlighted the value of controlling for direct assessments of EFs when using teacher reports of EFs to predict reading, important critiques of EF rating scales remain unaddressed. Furthermore, this same work has highlighted the need to use a meaningful developmental approach. Research using a multitrait multimethod design (i.e., examining multiple EF skills using a mixture of direct assessments, teacher reports, and parent reports) identified high multitrait-monomethod correlations for teacher reports of EFs and direct EF assessments, which the authors interpreted as an indicator of potential test impurity (Dekker et al., 2017). The authors suggested that rating measures may be capturing wider information beyond the specific EFs purported to be measured. Indeed, a recent study testing cross-method concurrent and predictive validity of direct EF assessments and teacher reports of EFs found that direct EF assessments were more predictive (i.e., had larger path coefficients) than teacher reports of EFs for both direct assessments and teacher reports of academic performance (Soto et al., 2020). The authors concluded that direct assessments of EFs may have better validity evidence than teacher reports for predicting academic outcomes.

Furthermore, researchers have questioned whether the widely used Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000)—a popular EF rating scale used in many of the studies cited here and for one model in the current study—may measure general behavioral difficulty rather than specific EF skills (McAuley et al., 2010). Together, these findings suggest that EF skills captured by teacher reports are either highly related to teacher reports of academic skills, or that the two reports may be confounded with one another. If teacher reports of EFs and teacher reports of students' literacy skills do not uniquely predict reading development, then teacher reports of EFs need to be redesigned to better understand how EFs pertain to reading.

Finally, the relationship between student identity and teachers' reports of EFs poses another possible confound that has concerned researchers. In a study of students in Grades 3 through 5, Garcia and colleagues (2019) found that teachers systematically underrated the EFs of boys, Black students, and students identified as English learners. This study suggests that it is of critical interest to know the extent to which student background plays a role in teacher reports of EF. For example, teacher reports may explain unique variance in future reading for White students, but the predictiveness may be masked in whole sample estimates by the underrating of EF skills for minoritized students. Thus, research examining the predictiveness of teacher EF reports also needs to examine potential differences in findings based on student demographics.

Current Study

This study examines whether teacher reports of EFs predict reading development (i.e., change in reading over a year of schooling) beyond direct assessments of EFs and teacher perceptions of their student's literacy competency in the ECLS-K:2011. This study's contribution comes from the use of a nationally representative dataset to estimate coefficients and the examination of whether controlling for a teacher academic rating scale reduces the contribution of EF teacher rating scales to a trivial level. Using autoregressive structural equation models predicting change in reading achievement from Grade 3 to Grade 4, and from Grade 4 to Grade 5, the model first examined the contribution of teacher reports of EFs to reading development. In the next step the model controls for the established effect of direct assessments of EFs to see whether both types explain unique variance in reading development. Next, the model tested the contributions of teacher EF reports to reading development when controlling for teacher perceptions of literacy competency and direct assessments of EFs. Based on limitations of rating scales and specific concerns about teacher reports of EFs (e.g. Dekker et al., 2017; McAuley et al., 2010), it was hypothesized that contributions of teacher EF reports to next year's reading achievement would be reduced to a trivial level when controlling for teacher perceptions of literacy competency. Models are tested separately for rating scales tapping attentional focusing, working memory, and inhibitory control. Next, based on the idea that teachers' biases may lead to inaccurate EF reports of boys, multilingual students, and Black students (Garcia et al., 2019), differences were examined in these results for various subgroups using multigroup structural equation modeling. Here, it was hypothesized that teacher reports of EF might be more predictive of future reading for girls, monolingual students, and White students, based on a prior study which found evidence of these biases in teacher reports of EF (Garcia et al., 2019). This study examined the following research questions:

1. How are teacher reports and direct assessments of EFs related to one another?
2. Do teacher reports of EFs explain unique variance in Grades 3-4 and 4-5 reading development in autoregressive models:
    1. When controlling for direct assessments of EFs?
    2. When controlling for direct assessments of EFs and teachers' perceptions of students' literacy competency?
3. Do teacher reports of EFs explain more unique variance for White students, monolingual students, and girls in Grades 3-4 and 4-5 reading development when controlling for direct

assessments of EFs and teachers' perceptions of students' literacy competency in autoregressive models?

Method

Participants

The current study used data from the Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011 (ECLS-K:2011; Tourangeau et al., 2019), which was conducted by the National Center for Education Statistics (NCES) in the Institute of Education Sciences of the U.S. Department of Education between 2011 and 2016. The ECLS-K:2011 is a nationally representative dataset that followed a cohort of over 18,000 children from Kindergarten to Grade 5 and collected information from multiple sources including children, parents, teachers, and school administrators. As these data are publicly available and includes only de-identified data, this study is exempt from IRB review, though the NCES Ethics Review Board approved the original study and all participants provided written informed consent.

The current study utilized a sample weighted for non-responsiveness associated with child, parent, and reading teacher data from Grade 3 (collected in spring of 2014) through Grade 5 (collected in spring of 2016), resulting in a nationally representative sample of 6,945 students. The sample weights were created by NCES to account for the differential probability of survey response for different demographic groups, so that estimates are not biased due to systematic non-responsiveness (Tourangeau et al., 2019). This resulting weighted sample was balanced by gender (49% female). Furthermore, the weighted sample was ethnically and racially diverse (52% White, 24.8% Latine[1], 13% African American, 4.5% Asian, 1.2% American Indian/Alaskan Native, 0.4% Native Hawaiian/Pacific Islander, and 4.1% two or more races,

---

[1] Survey response category was listed as "Hispanic," but replaced here to reflect gender inclusivity (Zentella, 2017)

based on parent report). The median household income of the sample was $55,001-$60,000, while the most frequent reported household income (20%) was between $100,001 and $200,000. The families of 21.3% of the weighted sample reported household incomes below the federal poverty level. The median educational attainment for the sample's respondent caregiver was "some college," while the majority (85%) held at least a high school diploma.

Procedures

Data collection for the ECLS-K:2011 was conducted by trained and certified data collectors (Tourangeau et al., 2019). These data collectors successfully completed training including written exercises and an observation-based task, which required administering child assessments to an age-appropriate child. The assessment procedures altered slightly over the years of collection, but generally consisted of a data collector administering the battery to an individual child over the course of 60 minutes. Data collection occurred in the school, with direct assessments administered to sampled students individually by a trained data collector. Teachers completed questionnaires, including reports of executive functions and literacy competency, by themselves and gave completed questionnaires to data collectors during assessment visits. Some sampled students likely had the same teacher complete their questionnaire, resulting in non-independence of observations, but this information is not available in the public-use dataset. However, this statistical assumption violation is mitigated by using a robust estimator, as detailed below in the Analytic Plan. Data collection for assessment measures included in the present analyses occurred in the spring of 2014 for Grade 3 to the spring of 2016 for Grade 5.

Measures

Demographic information. Parents completed surveys including date of birth, gender, race and ethnicity, and primary language in the home during the fall and spring of Kindergarten.

If data were missing, parents were asked to complete information during subsequent waves of data collection. Using the reported date of birth and date of assessment, NCES calculated the age of children at the time of assessment in months. Race and ethnicity were reported in parent interviews and used to create dichotomous variables for the following categories: White, Latine (no race specified), Latine (race specified), African American, Asian, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, or two or more races. Primary language was indicated by parents and used to create a dichotomous variable to indicate whether families used a language other than English at home. Primary language was collected during the spring of Grade 3. For the present analyses the most recent year prior to the predicted spring was used. NCES also used parent report data to create a continuous measure of socioeconomic status. This composite was created using information on parents'/guardians' education and occupational prestige score, as well as household income.

Reading achievement. The reading assessment of the ECLS-K:2011 was designed using item response theory (IRT), which allows for scores to be put on a common vertical scale despite utilizing different test forms in different grades (Najarian et al., 2019). Thus, test forms in earlier grades included a higher proportion of questions tapping basic word reading skills, with subsequent years shifting to include proportionally more reading comprehension questions. By Grade 3 questions tapping basic word reading skills were no longer included. Comprehension questions required students to identify information specifically stated in texts; to make complex inferences from texts; and to judge the appropriateness and quality of texts. Children received a set of routing questions before receiving low, middle, or high difficulty second-stage forms. The current analyses utilized the theta score estimating students' latent ability through the IRT model. These scores were estimated using concurrent calibration and chain-linking to allow for longitudinal measurement within and across grades (Najarian et al., 2019). Reliability estimates for the IRT-based theta scores on the reading assessment are high (.86-.87 across Grades 3 to 5 for the overall sample) and are derived from the variance of repeated estimates of theta for each child compared with total sample variance (Najarian et al., 2019).

Executive functions – direct assessments. Direct assessments of EFs included measures tapping working memory in Grade 3, attention shifting in Grades 3 and 4, and a measure tapping inhibitory control in Grade 4.

Attention shifting (Grades 3 and 4). The Dimensional Change Card Sort (DCCS; Zelazo, 2006; Zelazo et al., 2013) was used to capture attention shifting (called "cognitive flexibility" in the ECLS-K:2011 manual; Tourangeau et al., 2019) in Grades 3 and 4. This task requires children to sort pictures of colored objects by either color or shape, with rules mixed across 30 trials. In Grades 3 and 4, children completed a developmentally appropriate computerized version of the DCCS, which requires children to sort cards using a keyboard and is based on the NIH Toolbox (Zelazo et al., 2013). This computerized version features sorting rules intermixed across the 30 trials, with one rule more common to build a response tendency. This version also uses a combined score based on both accuracy and reaction time. The publishers report a high test-retest reliability coefficient for the measure (.92; Zelazo et al., 2013).

Working memory (Grade 3 only). The Numbers Reversed task from the Woodcock-Johnson III Tests of Cognitive Ability (Woodcock et al., 2001) was used to measure working memory in Grade 3. This task requires children to repeat a sequence of numbers spoken aloud in reverse order. The task begins with up to 5 two-number sequences. If the child does not make three consecutive mistakes, they receive up to 5 three-number sequences. This process proceeds up to a maximum of eight numbers until the child makes three consecutive mistakes or completes all the number sequences. Although a variety of scores are available, these analyses used the *W* ability score, which is a type of standardized score that provides a common scale of equal intervals to represent both child's ability and task difficulty, allowing it to be considered a growth scale. The publishers of the Woodcock-Johnson III report a high split-half reliability coefficient for the measure (.87; McGrew & Woodcock, 2001)

Inhibitory control (Grade 4 only). The NIH Toolbox Flanker Inhibitory Control and Attention Task (Flanker; Zelazo et al., 2013) was used to measure inhibitory control in Grades 4-5 of the ECLS-K:2011. For this task children must identify the direction of a central arrow while inhibiting the direction of four flanking arrows on 20 test items. The score yielded reflects both speed and accuracy using an algorithm like the formula for the computerized version of the DCCS. As with the DCCS, the publishers report a high test-retest reliability coefficient for the measure (.92; Zelazo et al., 2013).

Executive functions – teacher reports. Teacher reports of EFs included measures tapping working memory in Grade 3 and a rating scale including attentional focusing (Grades 3 and 4) and inhibitory control (Grade 4 only).

Working memory (Grade 3). Four items from the Behavior Rating Inventory of Executive Function (BRIEF, Gioia et al., 2000) were completed by teachers in the spring of Grade 3 to assess working memory. Teachers indicated whether a given student "Never, Sometimes, or Often" demonstrated specific behaviors supposed to be related to working memory. For example, a specific indicator asks how often a student "Needs help from an adult to stay on task." Reliability was high among the four items (Cronbach's alpha = .91). Items were reverse coded for ease of interpretation.

Attentional focusing (Grades 3 and 4) and inhibitory control (Grade 4). An adapted version of the Temperament in Middle Childhood Questionnaire (TMCQ; Simonds & Rothbart, 2004) was completed by teachers to assess the attentional focusing and inhibitory control facets of EFs. The TMCQ also asks teachers to indicate how often children exhibit certain social skills and behaviors related to inhibitory control and attentional focusing. Teachers rate on a 5-point scale ranging from “almost always untrue” to “almost always true” with a middle option of “sometimes true, sometimes untrue.” Teachers were also given the option of “not applicable” if the statement did not apply to the child, which was coded as a missing response. The ECLS-K:2011 includes 6 of the 7 items from the original TMCQ Attentional Focusing subscale and 6 of the 8 items from the original TMCQ Inhibitory Control subscale. A sample indicator for the attentional focusing subscale asks whether it is true that the focal child “Is easily distracted when listening to a story.” A sample indicator for the inhibitory control subscale asks whether it is true that the focal child “Likes to plan carefully before doing something.” This adapted version also includes one item from the Inhibitory Control subscale of the Children’s Based Questionnaire short form (CBQ; Putnam & Rothbart, 2006), which is recommended for children aged 3-7 and was used in prior waves of the ECLS-K:2011. The TMCQ was designed as a parent-report, so item wording was changed on one item from the Attentional Focusing subscale and one item on the Inhibitory Control subscale to make the items appropriate to the school setting. Reliability for both subscales of the TMCQ was high (Cronbach alpha = .85-.96 across subscales and years). Items were reverse coded for ease of interpretation.

Literacy competency – teacher reports. Teachers were asked to rate students’ reading abilities, writing abilities, and oral language on a three-point scale based on curriculum standards for their grade level, with responses of “Below grade level,” “About on grade level," or “Above grade level" for each domain. Reliability was high among the three items (Cronbach's alpha = .87-.89 across the grades).

Analytic Plan

Structural equation modeling (SEM) in Mplus 8 (Muthén & Muthén, 2017) was used to investigate predictive relations of teacher reports of EFs with reading achievement while controlling for direct assessments of EFs, teacher perceptions of literacy, demographic covariates, and an autoregressor to examine change in reading. The demographic covariates included in each model are gender, age, race/ethnicity, and socioeconomic status. SEM allows for the estimation of latent factors for better construct coverage, as well as a model fitting and building process suitable for testing multiple hypothesized relations between variables. Each model used the maximum likelihood estimator with robust standard errors, which is robust to issues of non-normality and non-independence of observations (Muthén & Muthén, 2017). Each model also used full-information maximum likelihood to handle missing data. In the interest of transparency, please see Table A1 in the Appendix for a list of all variables used.

All models were autoregressive and predicted a given year's spring measure of reading achievement using measures collected in spring of the previous year to test effects of EF reports on change in reading. To maintain consistency across models, the analyses used data from Grades 3-4 for predictor variables, as teacher reports of literacy consisted of the same three items in these two grades. Grades 3 and 4 each offered two teacher reports of EFs with parallel direct assessments (i.e., a direct assessment of the same facet of EF as the teacher report). Attentional focusing was assessed with both direct and teacher reports in both years, working memory in Grade 3 only, and inhibitory control only in Grade 4. Thus, four final models were created assessing the predictive relations of teacher reports of EFs with reading development while controlling for direct assessments of EFs and teacher perceptions of literacy.

Each of the four models followed a three-step model building process. Each model included all variables of interest, with additional paths of interest added in each step. Each model also included demographic covariates and an autoregressor. In each step, the change in overall fit statistics was examined according to common benchmarks (e.g. Brown, 2006), including the Comparative Fit Index (CFI, considered good fit close to .95 or above; Bentler, 1990), the Tucker Lewis Index (TLI, considered good fit close to .95 or above; Bentler, 1990), the Root Mean Square Error of Approximation (RMSEA, considered good fit close to .06 or below; Hu & Bentler, 1999), and the Standardized Root Mean Square Residual (SRMR, considered good fit close to .08 or below; Hu & Bentler, 1999), as well as results from the Satorra-Bentler chi-square test to determine whether the addition of these paths improved model fit. To address the research question on the predictive utility of teacher reports of EFs and direct assessments of EFs for change in reading, the change in standardized path coefficients was examined for each step. The first step included a path from the teacher report of EF, as well as the autoregressor and demographic controls, to assess the contribution of these reports when not controlling for either direct assessments or teacher reports of literacy. Here, it was hypothesized that teacher reports of EF would contribute to future reading (i.e., the following year's reading measure). Next, a path was included from the parallel direct assessment of EF to future reading to examine whether teacher reports of EFs uniquely contribute to change in reading when controlling for the more widely known effect of direct assessments of EFs. Given prior research on the distinction between teacher reports and direct assessments (e.g., Toplak et al., 2013), it was hypothesized that both would make unique contributions to change in reading. Finally, a path was included from teacher reports of literacy to examine whether teacher reports of EF do contribute to reading development after accounting for a direct assessment of EF and teachers' perceptions of literacy competence. Here, it was hypothesized that the coefficient for teacher reports of EF would drop to a trivial level, while the coefficient for direct assessments of EF would remain predictive of reading development.

Finally, based on prior research about the bias in teacher reports of EF by student race, gender, and language status (Garcia et al., 2019), multiple group SEM was used to test whether coefficients differed for students of color relative to White students, girls relative to boys, and multilingual students relative to monolingual students. Here, it was hypothesized that teacher reports would be less predictive for marginalized students, and that teacher reports might remain predictive for White students, monolingual students, and boys. For these analyses, the groups for American Indian/Alaskan Native and Native Hawaiian/Pacific Islander were too small for models to run successfully. Thus, these groups were collapsed into one "Indigenous" category, allowing for examination of potential coefficient differences for these students rather than excluding them from multigroup analyses. For these analyses, CFAs were first conducted for each group separately by model. A lack of fit at this step indicated that a multigroup analysis would not be possible. Next, measurement invariance testing was conducted across subgroups. To do so, the difference in model fit was examined between a constrained and unconstrained version of the multigroup model using the Satorra-Bentler chi-square test. The constrained version fixed the factor loadings to be the same across subgroups, while the unconstrained allowed the intercepts and loadings to vary across subgroups. A significantly better fit for the unconstrained model indicated that the factors did not capture the same constructs for different subgroups. Then, the same model building process detailed above was conducted, but with path coefficients allowed to differ for each subgroup. Here, the differences in path coefficient magnitudes between subgroups were examined using the Wald test of parameter constraints. In accordance with the hypothesis, the comparison groups for these analyses were White students, monolingual students, and boys. These model results were also compared to the main findings.

Results

Sample Descriptives

Table 1 displays the means and standard deviations for the variables of interest. The data were examined for evidence of skewness and kurtosis in the distributions of the variables of interest. Most of the variables were within a reasonable range of skewness (between -1 and 1) and kurtosis (between -2 and 2) values. However, the direct assessments of executive functions all showed evidence of non-normality in the distributions. Thus, the maximum likelihood estimator with robust standard errors was used to mitigate this violation and to ensure that results are robust to issues of non-normality, as detailed above.

Confirmatory Factor Analysis (CFA)

Single-group CFA was used to create latent factors for teacher ratings of executive functions and teacher perceptions of literacy competency. The resulting fit was strong for most measurement models (CFI = .973 to .986, TLI = .957 to .981, RMSEA = .051 to .061, SRMR = .023 to .025), except for the Grade 4 inhibitory control model (CFI = .893, TLI = .859, RMSEA = .081, SRMR = .07). For this latter model, removing the first indicator of the inhibitory control factor ("The child can stop him/herself when s/he is told to stop;" standardized factor loading = .48) improved fit to an adequate level (CFI = .925, TLI = .896, RMSEA = .074, SRMR = .064). Tables A2 through A5 in the appendix display full models including factor loadings.

Correlations emerging from the CFA models among teacher ratings, direct assessments, and continuous covariates for each of the four CFA models are presented in Tables 2 and 3. Given the large sample size and relatedly strong statistical power, most of these relations were statistically significant. For three of the four models (Grade 3 attentional focusing, Grade 4 attentional focusing, and Grade 5 inhibitory control), the correlations between the direct assessment and teacher report of EF were modest (*r*s = .15 to .21) and aligned with Toplak and colleagues' (2013) median correlation of .19 for direct and teacher or parent reports of EF measures in a review of studies. For the Grade 3 working memory model, the correlation between the direct assessment and teacher report of working memory was moderate (r = .29). In all models, the teacher reports of EF and literacy competency were strongly associated (*r*s = .58 to .62), providing support for the hypothesis that teacher reports of EFs are capturing substantial variation overlapping with perceptions of literacy competency.

Both direct assessments and teacher reports of EF were moderately associated with reading. For three of the four models (Grade 3 attentional focusing, Grade 4 attentional focusing, and Grade 5 inhibitory control), the correlation with reading was slightly higher for the teacher reports (r = .38 to .39) than for the direct assessment (r = .29 to .32). For the Grade 3 working memory model, the correlation between the direct assessment and reading (r = .45) was marginally higher than the correlation between the teacher report and reading (r = .40). Teacher reports of literacy competency were strongly related to the direct assessment of reading (r = .68 for all models).

Single-Group Structural Equation Modeling

Table 4 displays the standardized path coefficients of all predictors and covariates, as well as the fit indices for these models, while Figures 1 through 3 display the coefficients for each predictor of interest over the three steps in the model building process. In the first step for all four models, as displayed in Figure 1 and the “Step 1” columns of Table 4, teacher reports of EF were predictive of change in reading (standardized path coefficients = 0.06 to 0.07, all *p*s < .001), albeit with small coefficients, when controlling for demographic covariates and an autoregressor.

In the second step, as displayed in Figure 2 and the “Step 2” columns of Table 4, controlling for the corresponding direct assessment of EF only trivially reduced the standardized path coefficient for teacher reports of EF (standardized path coefficient change = -0.002 to -0.011). This trivial change supports the hypothesis that direct assessments and teacher reports of EF capture different constructs. The path coefficient from the direct EF assessment to reading development was significant for all models (all *p*s < .05) but differed in magnitude and in relation to the path from teacher reports of EF. For the Grade 3 predicting Grade 4 working memory model, this coefficient (0.09, $p < .001$) was larger than the coefficient for the teacher report of EF (0.06, $p < .001$). For the other three models, the coefficients for the direct assessment of EF (.03 to .04, all *p*s < .05) were smaller than the coefficients for the teacher reports of EFs (.06 to .07, all *p*s < .001).

In the final step, as displayed in Figure 3 and the “Step 3” columns of Table 4, controlling for teacher perceptions of literacy competency reduced the coefficient on teacher reports of EF (standardized path coefficient change = -0.077 to -0.081) to a statistically nonsignificant and practically trivial level (standardized path coefficients = -0.02 to -0.01, all *p*s > .05) for all four models. Here, the direct assessments of EF path coefficients only trivially changed in magnitude (standardized path coefficient change = -0.007 to -0.024), though the path coefficient was no longer significant for the Grade 4 predicting Grade 5 attentional focusing model (standardized path coefficient = 0.02, $p = .096$). Teacher reports of literacy competency were meaningfully predictive of reading development (standardized path coefficient = 0.23 to 0.27, $p < .001$), taking into account the use of demographic controls and an autoregressor. Together, these results indicate that teacher reports of EF may be capturing teachers' perceptions of academic skills of students rather than specific information about EF skills that are relevant to reading.

Multigroup Analysis

Like the main analyses, multigroup analyses by student language and race revealed that teacher reports of EF did not predict reading development when controlling for literacy perceptions across groups. Preliminary CFAs and measurement invariance testing indicated that multigroup SEM analyses were only appropriate for race and language in the Grade 4 predicting Grade 5 attentional focusing model. For student gender, the unconstrained model fit significantly better (i.e. the EF rating indicators loaded differently on the factors for different subgroups) based on results from the Satorra-Bentler chi-square test (all *p*s < .001), indicating multigroup analyses would not be appropriate. For student home language and race (both *p*s = .17), the only non-significant difference between unconstrained and constrained models were for the Grade 4 predicting Grade 5 attentional focusing model. Thus, student home language and race multigroup analyses were conducted for the Grade 4 predicting Grade 5 model.

For home language, results from the Grade 4 predicting Grade 5 attentional focusing structural model indicated teacher reports of EF do not uniquely predict reading development when controlling for teacher perceptions of literacy competency for either monolingual or multilingual students. Table A6 in the appendix displays the coefficients and fit indices for each of the steps in the model building process. At each step, there were no significant differences (all *p*s > .05) in the structural path coefficients between monolingual and multilingual student groups. As with the main findings, including a path for teacher reports of literacy dropped the coefficient for teacher reports of EF to a marginal level for both monolingual (-0.003, $p > .05$) and multilingual (-0.03, $p > .05$) students.

For student race, results from the Grade 4 predicting Grade 5 attentional focusing structural model indicated teacher reports of EF do not uniquely predict reading development when controlling for teacher perceptions of literacy competency for students identified as White, Black, Latine, Indigenous, or Multiracial. Table A7 in the appendix displays the coefficients and fit indices for each of the steps in the model building process. For students identified as Asian or Multiracial, teacher reports of EF were never significantly predictive of reading development (all *p*s > .05). In the second step, the direct EF assessment was only significantly predictive for students identified as White (0.04, $p < .001$) and Latine (0.05, $p = .003$). In the third step, teacher reports of EF were predictive for students identified as White, but the sign on the coefficient flipped to indicate these reports were negatively predictive (-0.03, $p = .033$), though practically trivial. The only significant difference ($p = .0428$) in coefficient size was between teacher reports of literacy for students identified as White (0.27, $p < .001$) and Asian (0.13, $p = .02$).

Discussion

Across four models, teacher reports of EFs consistently uniquely predicted reading development when controlling for a direct assessment of EF, but not when controlling for teachers' ratings of students' literacy skills. This study provides evidence that teacher reports of EF do not add information about EFs beyond teachers' perceptions of their students' literacy skills when predicting reading development. This finding has implications for researchers using teacher ratings to evaluate the relationship between EFs and literacy skills, for teachers receiving instructional advice about supporting students EF and reading development, and for policymakers attempting to screen students for difficulties in EF. This study suggests that more research is needed to better understand what exactly teacher ratings of EFs do measure, especially in the larger context where teacher ratings of EFs are being mandated by large districts (e.g., Amin, 2021).

Questioning Validity of EF Reports in Predicting Reading Development

This paper contributes to the literature evaluating the validity of teacher reports of EFs by testing their association with direct assessments of EFs in a large, nationally representative dataset of elementary-aged children. Like past work (Biederman et al., 2008; Conklin et al., 2008; Soto et al., 2020; Toplak et al., 2013), the results showed modest associations between EF ratings and direct assessments across the four models (rs = .16 to .20). Past research with similar findings has argued that the two EF measure types may capture different cognitive functioning constructs (Toplak et al., 2013). Indeed, the mono-method correlations observed for the EF rating and literacy ratings were higher (r = .59 to .63) across the models than the modest cross-method correlations between the EF ratings and direct assessments. This finding provides further evidence that a parallel EF rating and direct assessment likely do not capture the same underlying construct, as these EF ratings are more strongly related to a teacher's rating of their student's literacy skills (a related, but theoretically distinct construct) than to a corresponding direct assessment purporting to measure the same facet of EF.

Researchers with similar findings has argued that both teacher ratings and direct assessments of EF may be capturing distinct, but complementary aspects of overall EFs (e.g., Miranda et al., 2015). In this argument, teacher ratings are thought to capture applied EFs in ecologically valid situations, while direct assessments capture a purer level of EF functioning.

The current results show that controlling for a direct assessment of EF only trivially reduces the coefficient for teacher ratings of EF on reading development. As with past research including elementary-aged children (Gerst et al., 2017; Soto et al., 2020; Ten Eycke & Dewey, 2016), both the EF ratings and direct assessments were predictive of change in reading, with differences in magnitude across the models. Indeed, at this step the standardized coefficient was generally larger for the teacher EF rating than the parallel direct EF assessment. Such a finding may indicate that both assessment types predict reading development, with each contributing distinct information. Given the multitude of skills involved in reading, and the use of a general reading proficiency assessment in the ECLS-K: 2011, more research is needed to examine exactly what aspects of reading each EF assessment type best predicts. Direct assessments of EFs may be more related to cognitive reading subcomponents like word reading and reading fluency, while teacher ratings of EFs may be more related to classroom behaviors that are associated with reading, like maintaining focus on reading in a busy classroom. However, more research is needed to determine how EFs support these various reading skills.

Aspects of EF Reports that Need Further Research

The findings from this study raise questions about what teacher reports of EFs measure, and how these reports should be used to predict reading development. The SEM analyses presented in this study tested whether teacher ratings of executive functions predict change in reading, when controlling for a direct assessment of executive functions and for teacher perceptions of literacy skills. As the results shown in the “Step 3” columns of Table 4 indicate, the inclusion of a teacher rating scale of literacy skills reduced the effect for teacher EF ratings to a trivial level. This finding adds onto recent work from Soto and colleagues (2020), who found that direct assessments of EF are more predictive of both academic rating scales and parallel direct assessments of academic performance than teacher reports of EFs in a path model with both assessment types. The current study additionally shows that teacher reports of EFs do not predict reading development when controlling for teachers' perceptions of their students' literacy skills, while a direct assessment of EF does remain predictive.

One possible explanation for this finding could be that students' EF and literacy skills as measured by teacher reports are highly related, but unrelated to EF skills as measured by direct assessment. Indeed, reading development and EF skills measured by teacher reports may be one form of bootstrapping that occurs between reading and other cognitive skills (e.g., Stanovich, 1986). Such an explanation would suggest that teacher reports are tapping some sort of ecologically valid behavioral EF skills that are not captured by direct assessments. Given this possibility, further assessment design may better capture the EF skills that pertain to reading. Cartwright and colleagues (2010) have developed a direct assessment of attention shifting (termed "cognitive flexibility" in their work) that incorporates phonological and semantic aspects of print as a "reading specific" measure of EF. Such reading-specific assessment development may be needed for teacher reports of EFs to capture how EFs are applied to reading activities in the classroom.

Teacher ratings of EFs may also be reflective of teachers' views of students' literacy skills. However, the teacher reports of EFs used in the ECLS-K:2011 do not often ask teachers about their students' literacy skills. Only a few indicators in the teacher reports of EFs are overtly related to literacy skills. For example, one indicator from the attentional focusing teacher report asks whether a focal student "Is easily distracted when listening to a story." Such an indicator might capture teachers' perceptions about students' listening comprehension skills in addition to information about their EF skills. Yet indicators overtly related to literacy seldom appear in teacher reports of EFs, and more research is needed to determine whether teacher reports of EFs and of literacy skills are confounded. The notion of a "halo effect" (Forgas & Laham, 2017) from psychological measurement may shed light on this finding. A halo effect describes the tendency for perceptions in one domain to influence perceptions in another domain. In other words, a teachers' overall perceptions of a student may influence their report of both literacy skills and EFs. Researchers might employ cognitive interviewing to ask teachers to narrate their thoughts aloud while completing both a teacher report of students' EFs and literacy skills to examine whether these interviews show evidence of a halo effect taking place. As is evident from this discussion, there are many areas of EF assessment design and reading that need to be addressed.

Multigroup analyses were used to see if student identity explained why teacher EF reports did not uniquely predict reading development when controlling for teachers' ratings of their students' literacy skills. These analyses tested differences based on student gender, language status, and racial or ethnic identity, given prior evidence of systematic bias based on these identity categories (Garcia et al., 2019). Multigroup analyses examining coefficient differences for groups based on student home language, and racial or ethnicity identity, revealed no substantive differences for subgroups, as shown in the "Step 3" columns of Tables A6 and A7. Specifically, teacher EF reports were not predictive for any subgroups when controlling for teacher perceptions of literacy. Although prior research has indicated that there may be systematic bias in teacher EF ratings based on student demographics, but the present study shows no evidence of this bias.

Limitations

Although this study benefits from the use of a nationally representative sample, there are methodological limitations that future research should address. First, the ECLS-K: 2011 utilized different teacher reports and direct assessments of EF over the course of data collection and did not always offer parallel assessments. For example, although an inhibitory control rating scale was collected at every time point, the parallel direct assessment was only administered in Grades 4-5. The limitations in available parallel assessments shaped the decision to only examine Grades 3-5 in this study. Similarly, although the reading test was designed using item-response theory and different forms for different grades to create a vertically scaled measure with good psychometric properties, the resulting score represents a broad "reading proficiency" score, which lacks specificity. Students in earlier grades received more items tapping basic reading skills, while students in later years received more questions on reading comprehension. The teacher rating of students' literacy skills was also limited in the ECLS-K: 2011 to three items rating reading, writing, and oral language. Although this measure was significantly and meaningfully predictive of reading development in all models, it is a very coarse measure of a construct as broad and complex as "literacy." Thus, more research is needed that goes beyond the use of general constructs such as listening and writing. Doing so may allow researchers to discern what aspects of literacy the teacher reports of executive function might measure. Multigroup analyses were limited due to the better fit of unconstrained measurement models, which suggests that the EF rating constructs fit differently for different subgroups. Future research should explore whether EF rating scales are appropriate for different subgroups of students based on home language, racial or ethnic identity, and gender.

Finally, while the availability of the ECLS-K: 2011 made it possible to conduct this study, these data were collected from 2010 to 2016. As a result, these students have finished their primary education and some of the participating teachers may have left the profession. The changes since this time indicate that further research is needed to examine whether these findings hold in classrooms. Indeed, one transformation since these data were collected has been the increasing focus on EFs for teachers, meaning it is possible that teachers are more adept at identifying EF skills beyond academic performance. Future research should examine whether these findings hold with the current cohort of elementary-aged children.

Conclusion

Teacher reports of EFs predict reading development when controlling for a direct assessment of EF. However, teacher reports of EFs were not predictive of reading development when controlling for a teacher report of students' literacy skills. This finding remains when examining subgroups based on student home language and student race. Although EF ratings may have utility in making diagnostic decisions about hyperactivity disorders (e.g., Miranda et al., 2015), the current study suggests that teacher reports of both EFs and literacy skills are highly related. More research is needed to better understand what teacher reports of EFs measure, and how to design these reports for use in reading research. While a wealth of prior research supports the importance of EFs for reading development (Follmer, 2018), the current findings suggest more research is needed to understand how best to measure EF skills as they pertain to reading. The findings question the use of existing teacher reports of EF for reading research, and the use of these ratings by school districts. Better understanding the constructs captured by EF assessment types will be invaluable for understanding how these important skills relate to the behavioral and cognitive regulation necessary for successful literacy development.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B140037 to New York University. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education. The author wishes to thank Michael Kieffer for his thoughtful feedback on earlier versions of this manuscript, as well as the anonymous reviewers and managing editor who provided constructive comments.

References

Aboud, K., Bailey, S., Petrill, S., & Cutting, L. E. (2016). Comprehending text versus reading words in young readers with varying reading ability: distinct patterns of functional connectivity from common processing hubs. *Developmental Science*, *19*(4), 632–656. https://doi.org/10.1111/desc.12422.

Amin, R. (2021, November 11). Social-emotional screener sparks pushback from NYC teachers and families. *Chalkbeat New York*. https://ny.chalkbeat.org/2021/11/11/22777057/nyc-social-emotional-screener-teacher-parent-pushback

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology, 8*(4), 231–240. https://doi.org/10.1076/chin.8.4.231.13509

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (8th ed., pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, *40*(1), 23-38.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Biederman J, Petty CR, Fried R, Black S, Faneuil A, Doyle AE, ... & Faraone SV (2008). Discordance between psychometric testing and questionnaire-based definitions of executive function deficits in individuals with ADHD. *Journal of Attention Disorders, 12*(1), 92–102.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Publications.

Butterfuss, R., & Kendeou, P. (2018). The role of executive functions in reading comprehension. *Educational Psychology Review*, *30*(3), 801–826. https://doi.org/10.1007/s10648-017-9422-6

Cartwright, K. B. (2023). *Executive skills and reading comprehension: A guide for educators*. Guilford Publications.

Cartwright, K. B., Marshall, T. R., Dandy, K. L., & Isaac, M. C. (2010). The development of graphophonological-semantic cognitive flexibility and its contribution to reading comprehension in beginning readers. *Journal of Cognition and Development, 11*(1), 61–85. https://doi.org/10.1080/15248370903453584

Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. *Journal of Experimental Psychology*, *141*(3), 470–488. https://doi.org/10.1037/a0027375

Chung, K. K. H., Lam, C. B., & Cheung, K. C. (2018). Visuomotor integration and executive functioning are uniquely linked to Chinese word reading and writing in kindergarten children. *Reading and Writing*, *31*(1), 155–171. https://doi.org/10.1007/s11145-017-9779-4

Cirino, P. T., Miciak, J., Ahmed, Y., Barnes, M. A., Taylor, W. P., & Gerst, E. H. (2019). Executive function: Association with multiple reading skills. *Reading and Writing, 32*(7), 1819–1846. https://doi.org/10.1007/s11145-018-9923-9

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
Conklin HM, Salorio CF, & Slomine BS (2008). Working memory performance following paediatric traumatic brain injury. *Brain Injury, 22*(11), 847–857.
Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia*, *59*(1), 34–54. https://doi.org/10.1007/s11881-009-0022-0
Dekker, M. C., Ziermans, T. B., Spruijt, A. M., & Swaab, H. (2017). Cognitive, parent and teacher rating measures of executive functioning: Shared and unique influences on school achievement. *Frontiers in Psychology*, *8*, Article 48. https://doi.org/10.3389/fpsyg.2017.00048
Duke, N. K., & Cartwright, K. B. (2019). The DRIVE model of reading: Deploying reading in varied environments. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 118–135). Routledge.
Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist*, *53*(1), 42–60. https://doi.org/10.1080/00461520.2017.1309295
Forgas, J. P., & Laham, S. M. (2017). Halo effects. In *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory*, (2nd ed., pp. 276–290). Routledge/Taylor & Francis Group.
Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2015). Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. *Journal of Educational Psychology*, *107*(1), 207–221. https://doi.org/10.1037/a0037366
Garcia, E. B., Sulik, M. J., & Obradović, J. (2019). Teachers' perceptions of students' executive functions: Disparities by gender, ethnicity, and ELL status. *Journal of Educational Psychology*, *111*(5), 918–931. https://doi.org/10.1037/edu0000308
Gerst, E. H., Cirino, P. T., Fletcher, J. M., & Yoshida, H. (2017). Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. *Child Neuropsychology*, *23*(4), 381–407. https://doi.org/10.1080/09297049.2015.1120860
Gioia, G. A., Isquith, P. A., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Psychological Assessment Resources.
Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.
Isquith, P. K., Roth, R. M., & Gioia, G. (2013). Contribution of rating scales to the assessment of executive functions. *Applied Neuropsychology: Child*, *2*(2), 125–132. https://doi.org/10.1080/21622965.2013.748389
Kieffer, M. J., & Christodoulou, J. A. (2020). Automaticity and control: How do executive functions and reading fluency interact in predicting reading comprehension? *Reading Research Quarterly, 55*(1), 147-166. https://doi.org/10.1002/rrq.289
Kieffer, M. J., Vukovic, R. K., & Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly*, *48*(4), 333–348. https://doi.org/10.1002/rrq.54
Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component

skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, *21*(4), 310–333. https://doi.org/10.1080/10888438.2017.1291643

LeBuffe, P. A., Shapiro, V. B., & Robitaille, J. L. (2018). The Devereux Student Strengths Assessment (DESSA) comprehensive system: Screening, assessing, planning, and monitoring. *Journal of Applied Developmental Psychology*, *55*, 62-70.

Locascio, G., Mahone, E. M., Eason, S. H., & Cutting, L. E. (2010). Executive dysfunction among children with reading comprehension deficits. *Journal of Learning Disabilities*, *43*(5), 441–454. https://doi.org/10.1177/0022219409355476

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*(1), 51–87. https://doi.org/10.1111/j.1469-7610.1996.tb01380.x

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function? *Journal of the International Neuropsychological Society*, *16*(3), 495–505. https://doi.org/10.1017/S1355617710000093

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III: Technical Manual*. Riverside Publishing.

Miranda, A., Colomer, C., Mercader, J., Fernández, M. I., & Presentación, M. J. (2015). Performance-based tests versus behavioral ratings in the assessment of executive functioning in preschoolers: Associations with ADHD symptoms and reading achievement. *Frontiers in Psychology*, *6*, Article 545. https://doi.org/10.3389/fpsyg.2015.00545

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Muthén, L. K., & Muthén, B. O. (2017). Mplus: Statistical Analysis with Latent Variables: User’s Guide (Version 8). Authors.

Najarian, M., Tourangeau, K., Nord, C., Wallner-Allen, K., & Vaden-Kiernan, N. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), Third-Grade, Fourth-Grade, and Fifth-Grade Psychometric Report* (NCES 2020-123). U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics.

Obradović, J., Sulik, M. J., Finch, J. E., & Tirado-Strayer, N. (2018). Assessing students' executive functions in the classroom: Validating a scalable group-based procedure. *Journal of Applied Developmental Psychology*, *55*, 4-13.
Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*(1), 51–87. https://doi.org/10.1111/j.1469-7610.1996.tb01380.x
Putnam, S.P., & Rothbart, M.K. (2006). Development of short and very short forms of the Children’s Behavior Questionnaire. *Journal of Personality Assessment, 87*(1): 103-113.

RAND Reading Study Group (2002). Reading for understanding: Toward an R&D program in reading comprehension. RAND.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (1st ed., pp. 97–110). Guilford Press.

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology, 15*(3), 232–246. https://doi.org/10.1080/09297040802220029

Simonds, J., & Rothbart, M.K. (2004, October). *The Temperament in Middle Childhood Questionnaire (TMCQ): A computerized self-report measure of temperament for ages 7–10* [Conference presentation]. The Occasional Temperament Conference, Athens, GA, United States.

Soto, E. F., Kofler, M. J., Singh, L. J., Wells, E. L., Irwin, L. N., Groves, N. B., & Miller, C. E. (2020). Executive functioning rating scales: Ecologically valid or construct invalid? *Neuropsychology*, *34*(6), 605-619. https://doi.org/10.1037/neu0000681

Spencer, M., Richmond, M. C., & Cutting, L. E. (2019). Considering the role of executive function in reading comprehension: A structural equation modeling approach. *Scientific Studies of Reading*, *24*(3), 179–199. https://doi.org/10.1080/10888438.2019.1643868

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *22,* 360-407.

Swanson, H. L., & Siegel, L. (2011). Learning disabilities as a working memory deficit. *Experimental Psychology*, *49*(1), 5-28.

Ten Eycke, K. D., & Dewey, D. (2016). Parent-report and performance-based measures of executive function assess different constructs. *Child Neuropsychology, 22*(8), 889–906. https://doi.org/10.1080/09297049.2015.1065961

Titley, J. E., D'Amato, R. C., & Koehler-Hak, K. M. (2014). Utilizing teacher ratings of student literacy to identify at-risk students: An analysis of data from the Early Childhood Longitudinal Study. *Contemporary School Psychology*, *18*(1), 59–68. https://doi.org/10.1007/s40688-013-0005-3

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

[Dataset] Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. & Najarian, M. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten-Fifth Grade Data File and Electronic Codebook, Public* (NCES 2019-051). National Center for Education Statistics.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities.* Riverside Publishing.

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

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. In P.D. Zelazo and P.J. Bauer (Eds.), *National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB): Validation for Children Between 3 and 15 Years.* Wiley.

Zentella, A. C. (2017). "Limpia, fja y da esplendor": Challenging the symbolic violence of the

Royal Spanish Academy. *Chiricú Journal: Latina/o Literatures, Arts, and Cultures, 1*(2), 21–42.
Zhang, Z., & Peng, P. (2023). Longitudinal reciprocal relations among reading, executive function, and social-emotional skills: Maybe not for all. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000787

Table 1
Means and standard deviations of all variables of interest.

Construct	Indicator	Grade 3	Grade 4	Grade 5
Reading	*DA*: ECLS-K:2011 Reading achievement (theta score)	1.14 (0.30)	1.31 (0.30)	1.48 (0.35)
Working Memory	*DA*: Numbers reversed (Z-score)	0.00 (1.00)	-	-
	TR: Remember (1-3 scale)	2.38 (0.69)	-	-
	TR: Multi-step (1-3 scale)	2.46 (0.68)	-	-
	TR: Adult help (1-3 scale)	2.34 (0.73)	-	-
	TR: Forgets (1-3 scale)	2.56 (0.64)	-	-
Attentional Focusing	*DA*: Dimension change card sort (accuracy/reaction time score)	7.15 (1.38)	7.65 (0.97)	-
	TR: Doing activites (1-5 scale)	3.57 (1.23)	3.61 (1.23)	-
	TR: Looks around (1-5 scale)	3.30 (1.21)	3.34 (1.22)	-
	TR: Distracted trying (1-5 scale)	3.50 (1.25)	3.55 (1.24)	-
	TR: Distracts easily (1-5 scale)	3.61 (1.20)	3.65 (1.20)	-
	TR: Needs to be told (1-5 scale)	3.47 (1.27)	3.52 (1.27)	-
	TR: Has hard time (1-5 scale)	3.58 (1.25)	3.62 (1.25)	-
Inhibitory Control	*DA*: Flanker (accuracy/reaction time score)	-	7.99 (0.99)	-
	TR: Can stop self (1-5 scale)	-	3.83 (1.26)	-
	TR: Stops self doing things (1-5 scale)	-	3.59 (1.15)	-
	TR: Easy time waiting (1-5 scale)	-	3.82 (1.10)	-
	TR: Plan carefully (1-5 scale)	-	3.19 (1.17)	-
	TR: Good following (1-5 scale)	-	3.83 (1.03)	-
	TR: Hard time slow rules (1-5 scale)	-	4.16 (1.03)	-
	TR: Hard time waiting talk (1-5 scale)	-	3.76 (1.23)	-
Literacy Perceptions	*TR*: Reading (1-3 scale)	2.07 (0.74)	2.06 (0.75)	-
	TR: Writing (1-3 scale)	1.89 (0.67)	1.90 (0.69)	-
	TR: Oral language (1-3 scale)	2.13 (0.59)	2.13 (0.61)	-
Age	Age at time of assessment (months)	109.14 (4.48)	121.12 (4.50)	-
SES	Socioeconomic status composite (NCES created)	-0.12 (0.77)	-0.12 (0.77)	-

Table 2
Correlations among Grade 3 predicting Grade 4 measures.

	1	2	3	4	5	6	7	8	9
1. *DA*: G4 Reading	1								
2. *DA*: G3 Reading	.84***	1							
3. *TR*: G3 Working memory	.40***	.42***	1						
4. *DA*: G3 Working memory	.45***	.45***	.29***	1					
5. *TR*: G3 Attentional focusing	.39***	.40***	.77***	.29***	1				
6. *DA*: G3 Attentional focusing	.29***	.32***	.22***	.23***	.20***	1			
7. *TR*: G3 Literacy perceptions	.68***	.67***	.61***	.40***	.58***	.28***	1		
8. G3 Age	.02	.04*	.02	.02	.04**	.04**	.03*	1	
9. SES	.43***	.43***	.17***	.21***	.18***	.14***	.34***	.01	1

Note. *DA* = Direct assessment; *TR* = Teacher rating.
* $p < .05$; *** $p < .001$

Table 3
Correlations among Grade 4 predicting Grade 5 measures.

	1	2	3	4	5	6	7	8	9
1. *DA*: G5 Reading	1								
2. *DA*: G4 Reading	.85***	1							
3. *TR*: G4 Attentional focusing	.38***	.40***	1						
4. *DA*: G4 Attentional focusing	.32***	.32***	.21***	1					
5. *TR*: G4 Inhibitory control	.38***	.40***	.82***	.20***	1				
6. *DA*: G4 Inhibitory control	.31***	.31***	.17***	.46***	.15***	1			
7. *TR*: G4 Literacy perceptions	.67***	.68***	.61***	.28***	.62***	.27***	1		
8. G4 Age	.02	.02	-.01	.06***	.01	.04***	.03***	1	
9. SES	.42***	.43***	.18***	.16***	.20***	.17***	.35***	.01	1

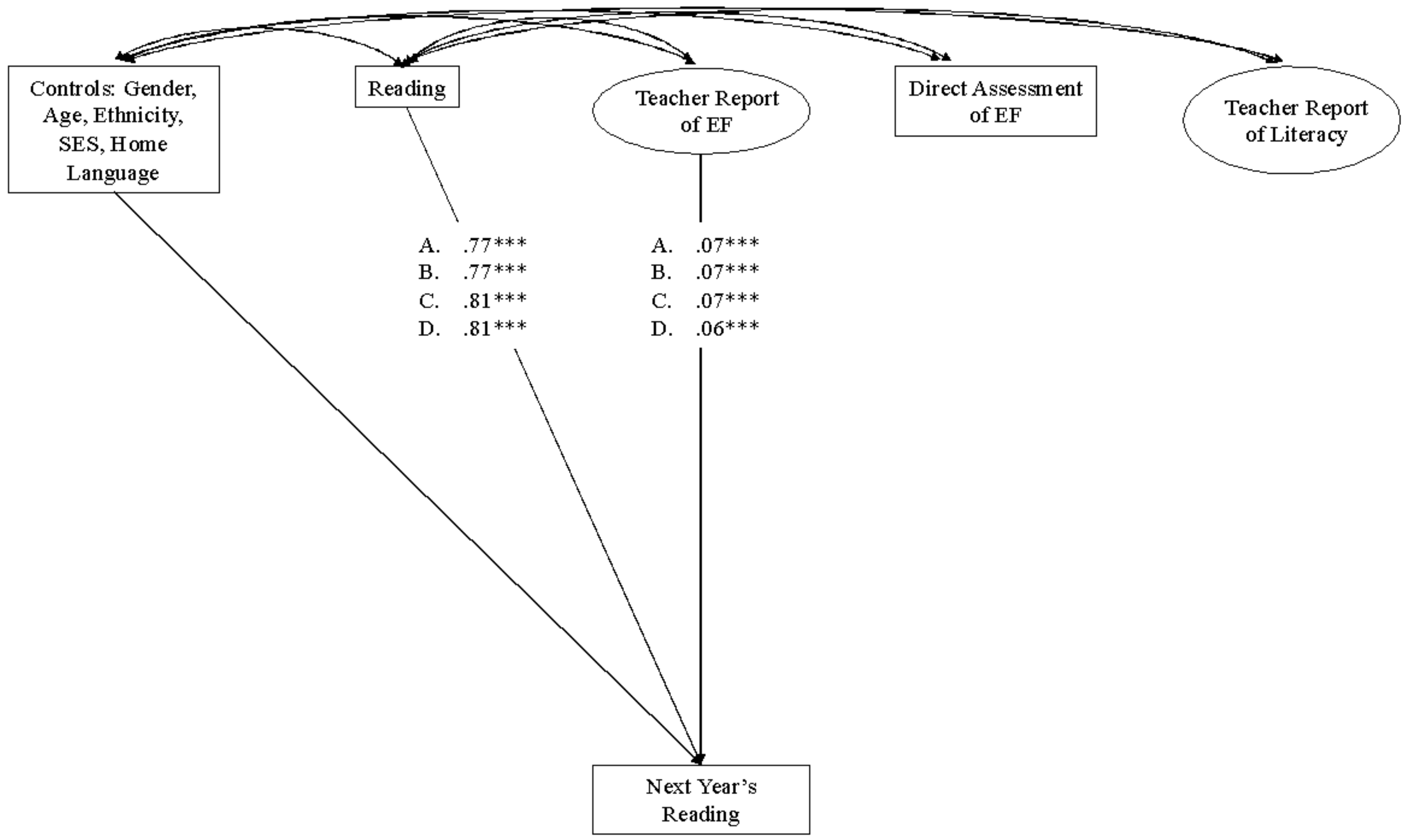
Note. *DA* = Direct assessment; *TR* = Teacher rating.
*** p < .001

Table 4
Standardized path coefficients and fit indices for main models.

Standardized path coefficients predicting next year's reading	Grade 3 working memory predicting Grade 4 reading			Grade 3 attentional focusing predicting Grade 4 reading			Grade 4 attentional focusing predicting Grade 5 reading			Grade 4 inhibitory control predicting Grade 5 reading		
	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
Predictors of interest												
TR: Executive function	0.07***	0.06***	-0.02	0.07***	0.07***	-0.02	0.07***	0.07***	-0.01	0.06***	0.06***	-0.02
DA: Executive function	-	0.09***	0.07***	-	0.04***	0.02*	-	0.03*	0.02	-	0.04***	0.03**
TR: Literacy competency	-	-	0.26***	-	-	0.27***	-	-	0.23***	-	-	0.23***
Controls												
Current year's reading	0.77***	0.74***	0.60***	0.77***	0.77***	0.61***	0.81***	0.80***	0.66***	0.81***	0.80***	0.66***
Female	-0.03**	-0.03**	-0.03***	-0.03***	-0.03**	-0.03***	-0.03***	-0.03***	-0.03***	-0.04***	-0.03***	-0.03**
Age	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*	-0.004	-0.01	-0.01	-0.004	-0.005	-0.01
Socioeconomic status	0.09***	0.09***	0.06***	0.09***	0.09***	0.06***	0.05***	0.05***	0.03***	0.05***	0.05***	0.03***
Black/African American	-0.04**	-0.03**	-0.03**	-0.04**	-0.03*	-0.03**	-0.03*	-0.02*	-0.02*	-0.02*	-0.02	-0.02*
Hispanic, race specified	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*
Hispanic, no race specified	-0.004	-0.003	-0.001	-0.003	-0.003	-0.001	-0.004	-0.003	-0.004	-0.001	-0.001	-0.004
Asian	0.01	0.01	0.01	0.01	0.01	0.01	0.005	0.005	0.01	0.005	0.004	0.005
Native Hawaiian/Pacific Islander	0.004	0.003	0.003	0.004	0.004	0.004	-0.01	-0.01	-0.01	-0.006	-0.01	-0.01
American Indian/Alaska Native	-0.002	-0.005	-0.01	-0.002	-0.003	-0.01	-0.002	-0.003	-0.003	0.000	-0.002	-0.003
Two or more races	0.005	0.005	0.002	0.005	0.005	0.002	0.01	0.01	0.01	0.01	0.01	0.01
Multilingual	0.001	-0.002	-0.001	-0.001	-0.001	0.002	-0.004	-0.004	-0.003	0.000	0.000	-0.001
Fit Indices												
CFI	0.891	0.903	0.975	0.931	0.933	0.977	0.938	0.94	0.982	0.887	0.887	0.947
TLI	0.821	0.851	0.960	0.900	0.908	0.968	0.91	0.917	0.974	0.837	0.844	0.926
RMSEA	0.07	0.066	0.034	0.057	0.056	0.033	0.054	0.053	0.029	0.056	0.055	0.038
SRMR	0.121	0.112	0.014	0.127	0.122	0.015	0.136	0.131	0.014	0.102	0.099	0.029
Likelihood ratio test		681.73***	3100.59***		251.97***	2798.75***		264.87***	2896.05***		142.114***	2594.586***
df		2	3		2	3		2	3		2	3

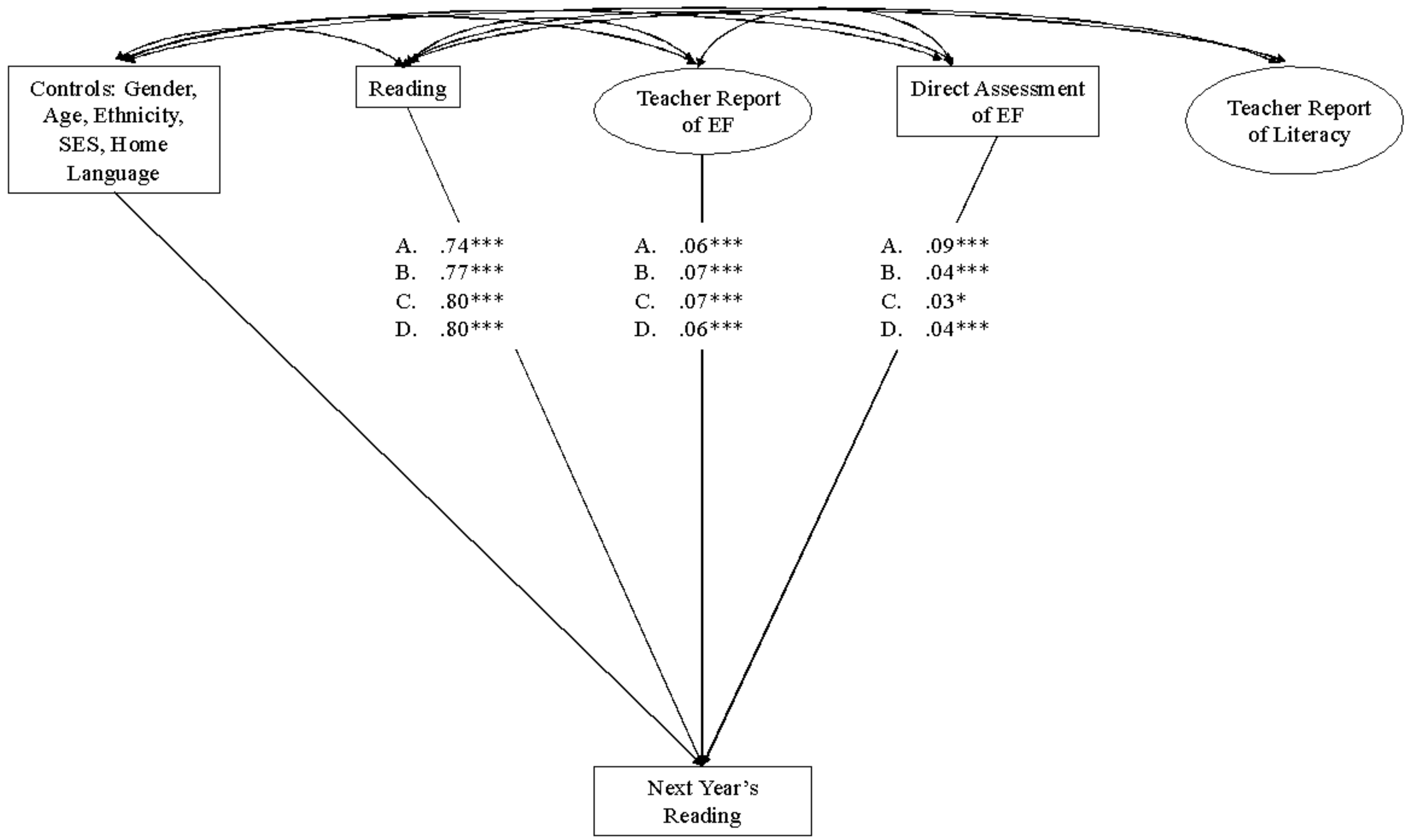
Note. *DA* = Direct assessment; *TR* = Teacher rating. Table excludes correlations among predictors.
* $p < .05$; ** $p < .005$; *** $p < .001$

Figure 1
Fitted structural equation model predicting next year's reading achievement with teacher report of EF for all four models, controlling for the autoregressor and demographic covariates.



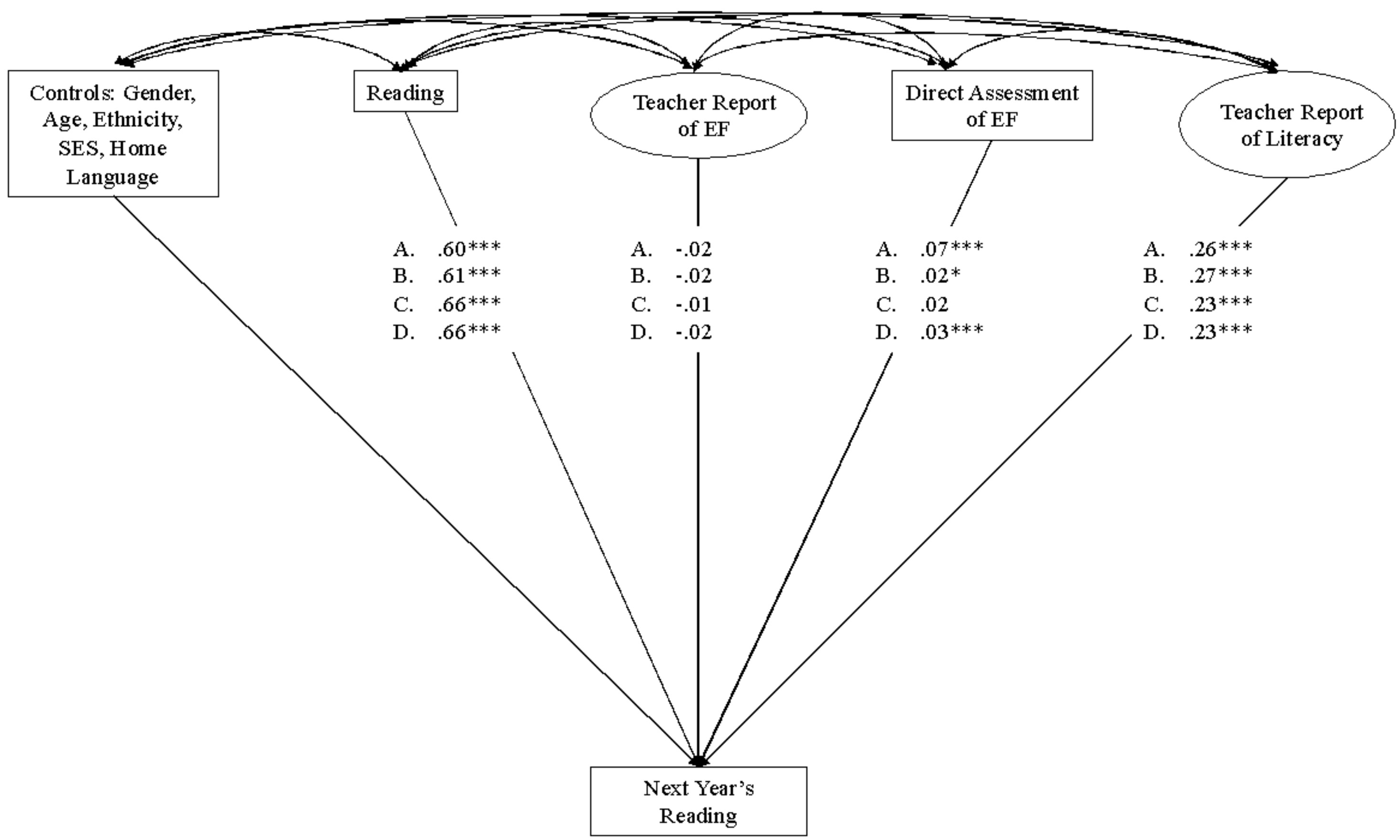
A. Grade 3 working memory predicting Grade 4 reading development
B. Grade 3 attentional focusing predicting Grade 4 reading development
C. Grade 4 attentional focusing predicting Grade 5 reading development
D. Grade 4 inhibitory control predicting Grade 5 reading development
*** $p < .001$

Figure 2
Fitted structural equation model predicting next year's reading achievement with teacher report of EF and direct assessment of EF for all four models, controlling for the autoregressor and demographic covariates.



A. Grade 3 working memory predicting Grade 4 reading development
B. Grade 3 attentional focusing predicting Grade 4 reading development
C. Grade 4 attentional focusing Grade 5 reading development
D. Grade 4 inhibitory control predicting Grade 5 reading development
* $p < .05$; *** $p < .001$

Figure 3
Fitted structural equation model predicting next year's reading achievement with teacher report of EF, direct assessment of EF, and teacher reports of literacy for all four models, controlling for the autoregressor and demographic covariates.



A. Grade 3 working memory predicting Grade 4 reading development
B. Grade 3 attentional focusing predicting Grade 4 reading development
C. Grade 4 attentional focusing Grade 5 reading development
D. Grade 4 inhibitory control predicting Grade 5 reading development
* *p* < .05; *** *p* < .001

Table A1
List of variable names from the ECLS-K:2011 (Tourangeau et al., 2019) dataset used in this study.

Construct	Indicator	Grade 3	Grade 4	Grade 5
Reading	*DA*: ECLS-K:2011 Reading achievement (theta score)	X7RTHETK5	X8RTHETK5	X9RTHETK5
Working Memory	*DA*: Numbers reversed (z-score)	X7NR_Z	-	-
	TR: Remember (1-3 scale)	T7REMBER	-	-
	TR: Multi-step (1-3 scale)	T7MLTSTP	-	-
	TR: Adult help (1-3 scale)	T7ADLHLP	-	-
	TR: Forgets (1-3 scale)	T7FORGTS	-	-
Attentional Focusing	*DA*: Dimension change card sort (accuracy/reaction time score)	X7DCCSSCR	X8DCCSSCR	-
	TR: Doing activities (1-5 scale)	T7BEZDAC	G8BEZDAC	-
	TR: Looks around (1-5 scale)	T7BLKARO	G8BLKARO	-
	TR: Distracted trying (1-5 scale)	T7BDSATN	G8BDSATN	-
	TR: Distracts easily (1-5 scale)	T7BEZDSL	G8BEZDSL	-
	TR: Needs to be told (1-5 scale)	T7BPYATN	G8BPYATN	-
	TR: Has hard time (1-5 scale)	T7BHTATN	G8BHTATN	-
Inhibitory Control	*DA*: Flanker (accuracy/reaction time score)	-	X8FLANKE R	-
	TR: Can stop self (1-5 scale)	-	G8BSPTLD	-
	TR: Stops self doing things (1-5 scale)	-	G8BSPQIK	-
	TR: Easy time waiting (1-5 scale)	-	G8BEZWAT	-
	TR: Plan carefully (1-5 scale)	-	G8BPLANS	-
	TR: Good following (1-5 scale)	-	G8BFLWIN	-
	TR: Hard time slow rules (1-5 scale)	-	G8BHTSLW	-
	TR: Hard time waiting talk (1-5 scale)	-	G8BHTTLK	-
Literacy Perceptions	*TR*: Reading (1-3 scale)	T7RTREAD	G8RTREAD	-
	TR: Writing (1-3 scale)	T7RTWRTE	G8RTWRTE	-
	TR: Oral language (1-3 scale)	T7RTOLAN	G8RTOLAN	-
Age	Age at time of assessment (months)	X7AGE	X8AGE	-
SES	Socioeconomic status composite (NCES created)	X9SESL_I	X9SESL_I	-
Race		X_RACETH_ R	X_RACETH_ R	
Gender		X_CHSEX_R	X_CHSEX_R	
Multilingual		X6LANGST	X6LANGST	
Sampling Weight		W9C79P_9T7 90	W9C79P_9T7 90	W9C79P_9T7 90
School ID (for Clustered Standard Errors)		S7_ID	S8_ID	S9_ID

Table A2
Measurement model for Grade 3 working memory predicting Grade 4 reading.

	Standardized factor loading
Grade 3 working memory	
Remember	0.840***
Multi-step	0.910***
Adult help	0.802***
Forgets	0.824***
Literacy competency	
Reading	0.861***
Writing	0.873***
Oral language	0.752***
Fit indices	
CFI	0.973
TLI	0.957
RMSEA	0.061
SRMR	0.024

*** $p < .001$

Table A3
Measurement model for Grade 3 attentional focusing Predicting Grade 4 reading.

	Standardized factor loading
Grade 3 attentional focusing	
Doing activities	0.867***
Looks around	0.810***
Distracted trying	0.931***
Distracts easily	0.837***
Needs to be told	0.900***
Has hard time	0.934***
Literacy competency	
Reading	0.858***
Writing	0.877***
Oral language	0.750***
Fit indices	
CFI	0.986
TLI	0.981
RMSEA	0.057
SRMR	0.025

*** $p < .001$

Table A4
Measurement model for Grade 4 attentional focusing predicting Grade 5 reading.

	Standardized factor loading
Grade 4 Attentional focusing	
Doing activities	0.885***
Looks around	0.825***
Distracted trying	0.935***
Distracts easily	0.840***
Needs to be told	0.899***
Has hard time	0.930***
Literacy competency	
Reading	0.870***
Writing	0.884***
Oral language	0.781***
Fit indices	
CFI	0.981
TLI	0.973
RMSEA	0.051
SRMR	0.023

*** $p < .001$

Table A5
Measurement model for Grade 4 inhibitory control predicting Grade 5 reading.

	Standardized factor loading
Grade 4 inhibitory control	
Can stop self	-
Stops self doing things	0.536***
Easy time waiting	0.662***
Plan carefully	0.759***
Good following	0.821***
Hard time slow rules	0.649***
Hard time waiting talk	0.615***
Literacy competency	
Reading	0.869***
Writing	0.885***
Oral language	0.781***
Fit indices	
CFI	0.925
TLI	0.896
RMSEA	0.074
SRMR	0.064

* $p < .05$; ** $p < .005$; *** $p < .001$

Table A6

Standardized path coefficients and fit indices for multigroup model examining home language with Grade 4 attentional focusing predicting Grade 5 reading.

Standardized path coefficients predicting Grade 5 reading	Step 1		Step 2		Step 3	
	Monolingual	Multilingual	Monolingual	Multilingual	Monolingual	Multilingual
Predictors of interest						
TR: Attentional focusing	0.08***	0.05*	0.08***	0.05*	-0.003	-0.03
DA: Attentional focusing	-	-	0.03*	0.02	0.02	0.01
TR: Literacy competency	-	-	-	-	0.23***	0.23***
Demographic controls						
Current year's reading	0.80***	0.81***	0.80***	0.80***	0.66***	0.68***
Female	-0.03**	-0.04*	-0.03**	-0.04*	-0.03***	-0.03*
Age	-0.01	0.02	-0.01	0.02	-0.02	0.01
Socioeconomic status	0.05***	0.04	0.05***	0.04	0.03***	0.03
Black/African American	-0.03*	0.03	-0.03*	0.03	-0.03*	0.03
Hispanic, race specified	-0.01	-0.05	-0.01	-0.05	-0.01	-0.03
Hispanic, no race specified	-0.001	-0.01	-0.001	-0.01	0.001	-0.01
Asian	-0.002	0.01	-0.003	0.01	-0.001	0.02
Indigenous	-0.004	0.0001	-0.005	0.02	-0.01	0.001
Two or more races	0.01	0.01	0.009	0.0001	0.01	0.01
Fit indices						
CFI	0.934		0.937		0.981	
TLI	0.911		0.918		0.975	
RMSEA	0.061		0.06		0.033	
SRMR	0.149		0.143		0.016	
Satorra-Bentler Chi-square test	-		128.03***		1738.79***	
df	-		4		6	

Note. *DA* = Direct assessment; *TR* = Teacher rating.
* p < .05; ** p < .005; *** p < .001

Table A7

Standardized path coefficients and fit indices for multigroup model examining race/ethnicity with Grade 4 attentional focusing predicting Grade 5 reading.

Standardized path coefficients predicting Grade 5 reading																		
	Step 1						Step 2						Step 3					
	White	Black	Latine	Asian	Indigenous	Multiracial	White	Black	Latine	Asian	Indigenous	Multiracial	White	Black	Latine	Asian	Indigenous	Multiracial
Predictors of interest																		
TR: Attentional focusing	0.06***	0.11**	0.09***	0.03	0.11*	0.003	0.06***	0.11**	0.08***	0.03	0.10*	0.006	-0.03*	0.05	-0.01	-0.01	0.04	-0.03
DA: Attentional focusing	-	-	-	-	-	-	0.04***	-0.04	0.05**	0.02	0.07	-0.01	0.03*	-0.04	0.04*	0.01	0.05	-0.01
TR: Literacy competency	-	-	-	-	-	-	-	-	-	-	-	-	0.27***	0.154*	0.23***	0.13*[a]	0.193	0.157*
Demographic controls																		
Current year's reading	0.81***	0.78***	0.81***	0.81***	0.87***	0.85***	0.80***	0.79***	0.80***	0.81***	0.87***	0.86***	0.63***	0.70***	0.66***	0.75***	0.72***	0.79***
Female	-0.04**	-0.04	-0.04*	-0.05	-0.05	-0.01	-0.04**	-0.04	-0.04*	-0.05	-0.04	-0.01	-0.04***	-0.04	-0.03*	-0.06	-0.05	-0.02
Age	-0.01	-0.01	0.01	-0.04	-0.07	0.05	-0.01	-0.01	0.01	-0.04	-0.08	0.05	-0.01	-0.01	0.003	-0.04	-0.05	0.06*
SES	0.04***	0.08**	0.06***	0.01	0.06	0.03	0.04***	0.08**	0.06**	0.01	0.07	0.03	0.02	0.07**	0.04*	0.01	0.07	-0.004
Multilingual	0.0001	0.06*	-0.02	0.01	0.01	0.02	-0.001	0.06*	-0.02	0.01	0.004	0.01	-0.001	0.06*	-0.01	0.02	0.005	0.005
Fit indices																		
CFI				0.93						0.933						0.978		
TLI				0.917						0.922						0.974		
RMSEA				0.07						0.069						0.04		
SRMR				0.184						0.178						0.024		
Satorra-Bentler Chi-square test				-						146.61***						1912.63***		
df				-						12						18		

Note. *DA* = Direct assessment; *TR* = Teacher rating.

* $p < .05$; ** $p < .005$; *** $p < .001$

[a] $p < .05$ for Wald test of parameter constraint compared to White students.