Developmental Psychology

Revisiting Age- and Schooling-Related Growth in School Readiness Skills: A Multimethod Validation Study

Jennifer K. Finders, G. John Geldhof, Jessica A. Dahlgren, Svea G. Olsen, and Megan M. McClelland

Online First Publication, June 6, 2022. http://dx.doi.org/10.1037/dev0001402

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

https://doi.org/10.1037/dev0001402

Revisiting Age- and Schooling-Related Growth in School Readiness Skills: A Multimethod Validation Study

Jennifer K. Finders[1], G. John Geldhof[2], Jessica A. Dahlgren[3], Svea G. Olsen[2], and Megan M. McClelland[2]
[1] Human Development and Family Studies, Purdue University
[2] Department of Human Development and Family Sciences, Oregon State University
[3] NPC Research, Portland, Oregon, United States

In the present study, we investigated the relative impact of age- versus schooling-related growth in school readiness skills using four modeling approaches that leverage natural variation in longitudinal data collected within the preschool year. Our goal was to demonstrate the applicability of different analytic techniques that do not rely on assumptions inherent in commonly applied methods (e.g., the school entrance cutoff method, regression discontinuity design) that selection into subsequent grades is based on birthdate alone and that the quality of experiences between grades are not responsible for differences in outcomes. Notably, these alternative methods also do not require data collected across multiple grades. Participants included 316 children (M_{age} = 54.77 months; 47.15% male) who mostly identified as White (64%) or Latinx (20%). A little over half of the sample attended Head Start preschools (54.75%). Four modeling techniques that leverage data collected at two timepoints in preschool were used to examine schooling effects on children's preliteracy, emergent math, and executive function (EF) skills. Results replicate evidence from previous research using traditional methods. Specifically, findings across all models demonstrate a schooling effect on preliteracy skills during the preschool year, above and beyond maturation, but not on emergent math or EF. We discuss the advantages and disadvantages of each analytical tool for researchers who are interested in answering questions about the effects of schooling with diverse data collection strategies, as well as broader implications for the integrity of educational and developmental science.

Keywords: schooling effects, schooling-related growth, school readiness, longitudinal data analysis, quasi-experimental designs

School readiness skills such as early literacy, math, and executive functions (EF; including attention, working memory, and inhibitory control), play a critical role in preparing children for formal schooling (Blair & Razza, 2007; Duncan et al., 2007; McClelland et al., 2014). These skills undergo tremendous development during early childhood (McClelland et al., 2015; Shonkoff & Phillips, 2000), but researchers have a tenuous understanding of the ways early childhood educational contexts scaffold the development of school readiness skills beyond maturational growth (Bailey et al., 2017). To address this issue, an emerging body of research has explored the unique contribution of school experiences on children's school readiness (i.e., "schooling effects"), relying primarily on the school entrance cutoff method and regression discontinuity designs (see Morrison et al., 2019, for a review). Results from such studies indicate an added benefit of schooling for preliteracy skills, but the evidence is less consistent with respect to emergent math and EF (Burrage et al., 2008; Kim & Morrison, 2018; Skibbe et al., 2011, 2013). Further, key assumptions made by a majority of past schooling effects research are that differences between the skills of children in higher versus lower grades directly reflect the amount of time spent in school, which is only the case when assignment to subsequent grades is based on birthdate alone (e.g., no other extraneous selection factors) and the quality of experiences across grades are not qualitatively distinct. In the present study, we discuss and apply four alternative statistical approaches that leverage longitudinal data and the natural variation of child ages within any single grade (i.e., preschool) to explore schooling-related growth in ways that explicitly overcome these shortcomings. These alternatives complement existing approaches to modeling schooling effects and expand researchers' ability to examine schooling effects using additional data collection strategies.

Jennifer K. Finders https://orcid.org/0000-0001-8368-4302

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100566 to Oregon State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This study was not preregistered. The datasets presented in this article are not readily available because Oregon State University's Institutional Review Board does not allow the sharing of the data from this study. The analysis code is available by request from the corresponding author.

Correspondence concerning this article should be addressed to Jennifer K. Finders, Human Development and Family Studies, Purdue University, 1202 West State Street, West Lafayette, IN 47907, United States. Email: jfinders@purdue.edu

Definition and Etiology of School Readiness Skills

School readiness refers to the skills and knowledge that contribute to children's learning and outcomes in school settings (Sabol

& Pianta, 2017). Of the many school readiness skills, preliteracy, emergent math, and EF are the most salient indicators of later school success (Claessens et al., 2009) and form the basis of most research on the topic. Significant variability in these skills exists before children enter kindergarten (Magnuson et al., 2004; NICHD Early Child Care Research Network, 2002). One explanation being that early learning environments are largely responsible for the development of such skills. For instance, experimental data from randomized controlled trials (RCTs) has provided strong support for the effects of preschool on early academic skills (e.g., Puma et al., 2010). However, disentangling age- and schooling-related growth poses several methodological challenges when considering data obtained from nonexperimental designs. When researchers compare children across grades (e.g., preschool vs. kindergarten), any observed differences reflect the combined effects of age *and* schooling. These collinear metrics of time make it difficult to estimate the unique effects of each. Adding to this complication, experiences and development interact such that the age at which children enter school may matter for how well they can take advantage of certain learning experiences (Li et al., 2013). In other words, maturation is necessary for learning, and early learning experiences matter for maturation (Bentin et al., 1991). Teasing apart these developmental processes requires the use of innovative methodological approaches that can estimate discrete school and age effects across outcomes. Although researchers have utilized quasi-experimental designs to examine schooling effects with nonexperimental data, many of these common methods have their own limitations, which we describe in the following sections.

Traditional Approaches to Examining Schooling Effects

School Entrance Cutoff Method

Extant schooling effects research has typically relied on study designs that match children on as many factors as possible (e.g., family demographics) while also finding groups who have spent appreciably different amounts of time in early educational contexts. Perhaps the most prominent of these designs is the school cutoff method. The school entrance cutoff method is a "natural experiment" that takes advantage of arbitrary cutoff dates which determine when children of almost identical ages can begin formal schooling (Morrison et al., 1995). For example, if the state cutoff date for kindergarten entry is August 1st, all children who reach age five before August 1st will begin kindergarten that year, while those born after that date must wait an additional year before enrolling. Using the school entrance cutoff method, researchers can constrain their sample to compare children born close to the cutoff date (e.g., within 3 months), but whose birthdates fall on either side of the cutoff; thereby, rendering different schooling experiences. This comparison allows researchers to examine how 1 year of schooling affects children's development relatively independent of age.

Studies utilizing the school entrance cutoff method comprise a growing body of literature demonstrating the role of educational experiences for enhancing school readiness skills. For instance, gains in early literacy skills are consistently attributed to the effects of schooling in the early elementary grades (Christian et al., 2000; Crone & Whitehurst, 1999; Cunningham & Carroll, 2011; Kim et al., 2021; Morrison et al., 1995). Moreover, schooling effects for literacy, and to a slightly lesser extent emergent math skills, have been documented into first grade (Bisanz et al., 1995; Christian et al., 2000; Morrison et al., 1997). The few studies that have examined schooling effects on EF have generated mixed findings, which seem to largely depend on the age of the children under examination and the nature of the EF task in question (Brod et al., 2017; Burrage et al., 2008; McCrea et al., 1999).

In general, the school entrance cutoff method provides a methodologically elegant means for controlling age to assess schooling effects and has been widely applied to examine schooling effects on diverse outcomes with data collected across school grades. However, it is not without limitations. This model restricts the usable sample to a fraction of available data and must be extended if a researcher wishes to also assess age-related development. Moreover, interpretation of the school entrance cutoff method rests on assumptions that the allocation of children to birthdates is random and grade level is determined by age only (Cahan & Cohen, 1989). Violation of these assumptions, particularly within the context of a small sample size, could result in selection bias.

Propensity Score Matching

To overcome limitations of the school entrance cutoff method, some researchers have applied propensity score matching to create comparable groups based on certain background characteristics (Morrison et al., 1995; Skibbe et al., 2011). For instance, one study used propensity score matching to equate children across schooling experiences and control for variables such as parent education and parent age that may account for selection into preschool (Skibbe et al., 2011). Results suggested that an extra year of preschool improved early literacy, but did not enhance EF skills (Skibbe et al., 2011). Propensity score matching has also been applied in studies examining the effects of full-day versus half-day preschool programs. For example, Leow and Wen (2017) demonstrated that differences in academic skills were not explained by the frequency at which children attended Head Start, even after equating children on demographic characteristics. However, a major limitation of propensity score matching when using nonexperimental data is that the effectiveness of the matching is heavily reliant on the availability and identification of observed covariates used to create equivalent groups (Newgard et al., 2004). In the context of testing schooling effects, omission of relevant unobserved covariates could lead to biased estimates for the statistical comparisons between children with more or less schooling experience.

Regression Discontinuity

Regression discontinuity is another popular approach to disentangling age and schooling effects that also takes advantage of the school entrance cutoff date (e.g., Gormley et al., 2008; Weiland & Yoshikawa, 2013). In this model, the effect of age is reflected in the slope of within-grade regressions by age, and the effect of schooling is reflected in the discontinuity between the two regression lines representing differences in grades (Cahan & Cohen, 1989). Children nearest to the school entrance cutoff are typically eliminated from consideration to create a cleaner comparison of children with varying levels of schooling. One study using this method found evidence for both age- and schooling-related growth on phonological awareness, with the schooling effect being four times larger than the age effect (Bentin et al., 1991). A regression discontinuity design in another study revealed schooling effects for various early literacy skills during the transition to school (Kim & Morrison, 2018). The regression discontinuity design has the advantage of allowing researchers to use a greater proportion of their sample and it can be flexibly applied in a variety of contexts, including longitudinal data analysis. However, this method assumes that model features, such as the covariates that estimate treatment effects, are appropriately specified. Not accounting for all variables that affect the assignment to preschool or kindergarten, such as differential attrition, can result in upwardly biased effects (Lipsey et al., 2015).

Alternative Methods for Modeling Schooling Effects

The pattern of results from previous work utilizing the school entrance cutoff method, propensity score matching, and regression discontinuity designs provides strong support for schooling effects on language and literacy, and more conflicting evidence for the effects of schooling on math and EF skills. Despite the robustness of findings across these three methods, even after adjusting for observed covariates, they may still contain bias due to the nature of the counterfactual condition (Zhai et al., 2014). In other words, researchers may reasonably conclude that there are effects of time in school on children's development, when in fact, a number of confounding factors may be responsible for these findings (e.g., the quality of educational experiences). For example, one study demonstrated that the home learning environment improves from when children are 36-months to 54-months as kindergarten entry approaches (Son & Morrison, 2010), suggesting extraneous factors that differ between cohorts may explain observed schooling effects. Indeed, nonrandom assignment poses a major threat to the internal validity of such techniques, thereby potentially misrepresenting the effect of schooling (Lipsey et al., 2015). Further, these models may require the reduction of sample size to children whose birthdates fall around the cutoff date, which can exacerbate issues of biased estimates. Limitations related to the nature of the counterfactual can be overcome by using theoretically informed models that take advantage of multiple time metrics in longitudinal data collected across the same school year (e.g., as discussed by Werner, 1957; see also, Lerner et al., 2009).

Developmental scientists and education researchers have demonstrated several plausible methods that are appropriate for obtaining independent estimates for the influence of age, time, and cohort (e.g., Baltes et al., 1970; Schaie, 1972). The most straightforward approach is to utilize multilevel regression modeling to investigate linear time effects at the within-child level independent of linear age effects at the between-child level (Curran & Bauer, 2011; Hoffman & Stawski, 2009). There is also the option to compare the intercept at school entry and the slope after school entry to determine whether there is a significant difference between these parameters that can be attributed to time in school (e.g., a form of discontinuity; Werner, 1957). Finally, one could compare a 6-month age difference at school entry against observed differences in child skills between the date of school entry and a date 6 months later by capitalizing on natural variation in children's birthdates. For instance, Goulet et al. (1974) discovered age and schooling- effects on vocabulary in preschool-aged children when examining intellectual functioning among children with similar ages and different amounts of schooling experience using longitudinal analysis.

Although the aforementioned methods present their own unique set of limitations, importantly, these strategies do not rely on assumptions that experiences are equivalent across grades, that covariates have been appropriately specified to estimate treatment effects in quasi-experimental designs, or that assignment to grades is based on birthdates alone. Therefore, they provide an additional means for separating the effects of age and time in school, and when examined together with the estimates from previous research, contribute to a more conclusive body of evidence on the effects of schooling. To our knowledge, this is the first systematic application of such alternative methods that leverage heterogeneity in children's ages within the same grade and dataset to examine schooling effects across a variety of outcomes. Such analysis is critical for demonstrating the validity of these less commonly applied alternatives for researchers who wish to overcome shortcomings of previous research and evaluate schooling effects with diverse data collection strategies.

Present Study

In the present study, we implement a series of complementary analyses for examining schooling effects to demonstrate alternative options to more traditional approaches that overcome issues related to nonrandom assignment into subsequent grades and the quality of schooling experiences in the comparison condition. We estimate four models that take advantage of variations in children's birthdates and the dates of assessment completion using two timepoints of data collected within a single preschool year. We start by demonstrating a basic regression approach in which age at time of assessment and time in school are specified as independent predictors of school readiness. Next, we test the equality of the time in school and age at school entry slopes in spline-like regression models. In our third approach, we randomly select one observation per participant in the fall or spring and examine age at time of assessment and time in school as predictors. Our last model uses coarsened exact matching to select a subset of fall and spring observations and perform pairwise comparisons between fall and spring outcomes. These four models overcome the inherent within-person collinearity that arises when chronological age and time in school are modeled simultaneously and allow researchers to compare outcomes when classroom experiences are otherwise similar within the same grade. We discuss the strengths and weaknesses of each methodological approach and argue that implementing such analyses offer a more holistic picture of the effects of schooling on key school readiness indicators than any single methodological tool on its own.

Method

Participants

Data came from 435 children who resided in the Pacific Northwest region of the United States and were participating in a study focused on developing a measure of EF skills. Children were recruited from Head Start and community-based preschools in the fall of 2011. To facilitate the matching-based analyses described below, we removed any cases that did not have complete data for the direct assessments (n = 10), who had extreme estimated ages at preschool entry (n = 12), whose classroom ID was missing (n = 1), and those who did not participate in both waves of the study or whose fall observation occurred more than 4 months after the beginning of the school year (n = 96).

The final analytic sample included 316 children who were between the ages of 48 and 61 months old at the beginning of the preschool year (M_{age} = 54.77 months, SD = 3.52 months). Children were nearly balanced on gender (47.15% male). Just over half of the participants were from families with low incomes, as indicated by their enrollment in Head Start preschool programs (54.75%). The sample was mostly White (64%) and Latinx (20%), which is representative of the broader region. The remainder of the sample identified as multiracial (11%), Asian/Pacific Islander (2%), Middle Eastern (<1%), African American (<1%), and other races/ethnicities (1%). Sixteen percent of the children were classified as English-language learners (ELLs) because they spoke Spanish as their primary language.

Chi-square tests did not indicate any significant differences between the original sample and the analytic sample on Head Start or ELL status, but the analytic sample was significantly less likely to be male compared with the original sample, $\chi^2(df = 1) = 9.67$, $p < .001$; odds ratio (OR) = .50. There were also no significant differences in child age, preliteracy, or emergent math at the start of preschool between the original and analytic sample. However, T tests revealed that the analytic sample performed significantly lower on EF in the fall of preschool (M_{diff} = 10.22, SE = 3.33) relative to the original sample, $t(403) = 3.07$, $p = .002$.

Procedure

Written consent was obtained from parents or primary caregivers before participation, and children gave verbal assent to participate in assessments. Parents filled out a background questionnaire to provide demographic information about their children during the preschool year (e.g., birthdate), and direct assessments of children's emergent math, preliteracy, and EF were collected by trained research assistants in the fall and spring of preschool. The fall and spring assessments were completed approximately 6 months apart. Assessments were administered in a designated space outside of the classroom (e.g., hallway) or a quiet corner inside the classroom. The assessments took about 15–20 minutes total to complete. Teachers indicated if children spoke Spanish as their primary language, in which case Spanish-speaking members of the research team administered assessments to these children in Spanish. All research activities were approved by the Institutional Review Board (IRB) at Oregon State University (Touch your Toes! Kindergarten Readiness Study, 4766). This study was not preregistered. The data sets presented in this article are not readily available because Oregon State University's IRB does not allow the sharing of the data from this study. The analysis code is available by request from the corresponding author.

Measures

School Readiness Outcomes

Emergent Math and Early Literacy. Children's emergent math and early literacy skills were assessed with the Applied Problems and Letter-Word Identification subtests of the Woodcock-Johnson (WJ) Battery III Tests of Achievement in English or the Batería III Woodcock-Muñoz Tests of Achievement in Spanish (Muñoz-Sandoval et al., 2005; Woodcock et al., 2001). These tests are norm referenced and require children to respond by either pointing to the correct answer or verbally expressing the answer until they incorrectly answer six questions in a row. The Applied Problems subtest assesses children's quantitative abilities related to counting objects, reading numbers, and basic addition and subtraction picture-problems. The Letter-Word Identification subtest assesses children's word-coding skills, including the ability to recognize and name letters and read words. We used W-scores in analyses, which are standardized and take into consideration the child's ability and the task difficulty. Both subtests have demonstrated high internal consistency and validity based on their correlations with other achievement tasks (Woodcock et al., 2001).

Executive Function Skills. Children's EF skills were directly assessed with the Head-Toes-Knees-Shoulders (HTKS) task, which measures the integration of working memory, inhibitory control, and attentional flexibility (McClelland et al., 2014). In this assessment, children are asked to do the opposite of what is instructed. For example, if the research assistant instructs them to touch their head, instead of following the command, children are directed to touch their toes (or vice versa). The same rules are then applied to knees and shoulders in the second section of the task, and then the instructions change so that children must remember new rules in the third section of the task (i.e., head goes with knees, and shoulders go with toes). Children receive 2 points for a correct response, 1 point for a self-correct, and 0 points for an incorrect response. Final scores on the HTKS are computed as the sum of children's performance across the 30 testing items and 17 practice items, with possible scores ranging from 0–94. The HTKS was translated into Spanish for administration with Spanish-speaking children. In previous research, the HTKS has demonstrated strong interrater reliability and predictive validity for children's academic outcomes in diverse groups of children (McClelland et al., 2014; Wanless, McClelland, Acock, et al., 2011).

Primary Independent Variables

Age. Age in months at the time of assessment was computed by subtracting children's birthdate from the date of each fall direct assessment and dividing by 30. The same procedure was implemented for age in months at the time of each spring direct assessment. Age at entry to preschool was computed by subtracting children's birthdate from the first day of school and dividing by 30. We estimated the first day of school as September 1st for all children in the sample, as this is a reasonable school start date in the location in which data were collected.

Time in School. The variable for schooling was computed by subtracting the first day of school (estimated as September 1st) from the date of each assessment and dividing by 30.

Covariates. Covariates were chosen because of their associations with children's school readiness (Castro et al., 2011; Cooper et al., 2011; Lee et al., 2014; Wanless, McClelland, Tominey, et al., 2011). These included child gender (1 = male, 0 = female), Head Start enrollment (1 = Head Start, 0 = no Head Start), and ELL status (1 = ELL, 0 = non-ELL).

Analytic Strategy

We fit a series of models to probe the robustness of prior results that had been found using traditional methods (i.e., school entrance cutoff, propensity score matching, and regression discontinuity), and to provide illustrative examples for researchers who wish to examine schooling effects using their own data. Specifically, we implemented the four following modeling approaches: (a) Naïve Regression Models for Simultaneous time and Age Effects, (b) Regression Models for time in School and Age at School Entry, (c) Regression Models for a Stratified Random Sample, and (d) Pairwise Comparisons Using an Age-Restricted Stratified Random Sample. Each model capitalizes on within-grade age heterogeneity and the fact that some older participants were the same age in the fall as younger participants were in the spring (see Figure 1). We ran analyses separately for each of the three outcome variables (preliteracy, emergent math, and EF). Regression models treated the HTKS measure as having a zero-inflated negative binomial distribution unless otherwise specified. Zero-inflated models allow researchers to predict expected count scores in combination with a model that predicts whether or not a case likely represents an excess zero (i.e., a score of zero above and beyond what is expected given a baseline distribution such as the negative binomial). Such models are especially suitable when modeling outcomes with participants who score at the floor, such as with the HTKS (see Figure 2).

Naïve Regression Models for Simultaneous Time and Age Effects

We first fit three-level regression models (observations nested in children, children nested in classrooms) in which age at time of assessment and time in school independently predicted the outcome measures. Statistical software like M*plus* automatically decomposes predictors and outcomes into independent within (and between) cluster components, leading to inherent collinearity at the within-person level for the two target predictors. That is, the number of days between fall and spring assessments is the same regardless of whether one treats this difference as a change in age or as a change in time in school.

One approach to resolving this issue would be to only consider variation of the time variables at the within-person level and without centering the data (e.g., by overriding the default in M*plus*). This approach would allow for a clean assessment of each predictor without collinearity but at the cost of assuming the predictor and outcome only covary at the within-person level. As an alternative, it is instead possible to fit the model using ordinary multilevel modeling software.[1] We fit these models in SAS using PROC MIXED. The models controlled for gender, Head Start status, and ELL status:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}Time\_In\_School_{ijk} + \beta_{2jk}Age\_at\_Assessment_{ijk} + e_{ijk}$$
$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k}Head\_Start_{jk} + \gamma_{02k}ELL_{jk} + \gamma_{03k}Gender_{jk} + u_{0jk}$$
$$\gamma_{00k} = \gamma_{000} + u_{00k}. \tag{1}$$

Significant schooling effects are indicated by a statistically significant coefficient for time in school.

We were only able to consider three-level models for emergent math and preliteracy skills, as common software packages do not support three-level zero-inflated negative binomial regression models. For the HTKS outcome, we instead fit a restricted two-level model in M*plus* that accounted for repeated measures nested in participants and additionally corrected standard errors for clustering at the classroom level. Time in school and age at assessment were only included at the within-level component of the model.

Regression Models for Time in School and Age at School Entry

Because the collinearity between time in school and age exists only at the within-person level, differences in the values are in reality only a function of age at school entry (i.e., a between-child variable). Thus, one approach to improving on the naïve model is to model only one of the two time metrics at the within-person level. This can be done in M*plus* with latent mean centering, although observed group mean centering is preferable when the group means are measured without error (Asparouhov & Muthén, 2019). We consider the exactly recorded dates used in the present study to be one example of such error-free measurement.

As such, we next fit a series of multilevel regression models that included both age at school entry and time in school as predictors of each outcome. The models were run in M*plus* Version 7.4 using TYPE = COMPLEX TWOLEVEL, where the multilevel component of the analysis treated repeated measures as nested in children and the standard errors were additionally corrected for clustering at the classroom level. Due to relatively low between-person variance in the time in school variable, we only included this predictor at the within-person level. The effect of time in school randomly varied across individuals. All models included gender, Head Start status, and ELL status as time-invariant covariates. The preliminary model for each outcome was:

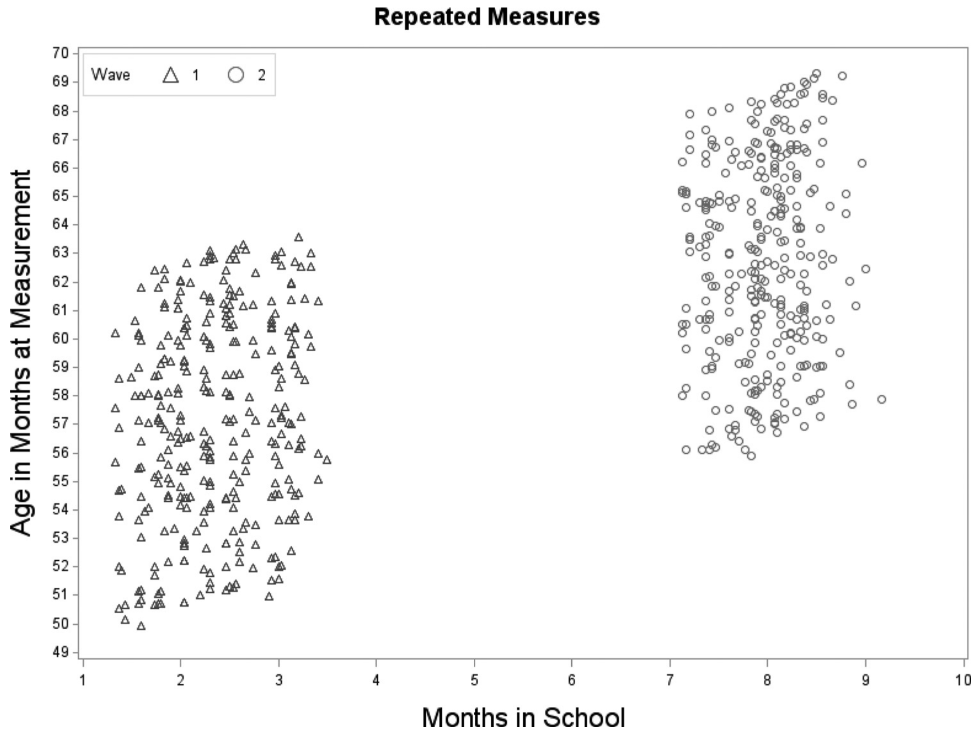
$$Y_{ij} = \beta_{0j} + \beta_{1j}Time\_in\_School_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}Head\_Start_j + \gamma_{02}ELL_j + \gamma_{03}Gender_j + \gamma_{04}Age\_at\_School\_Entry_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Head\_Start_j + \gamma_{12}ELL_j + \gamma_{13}Gender_j + \gamma_{14}Age\_at\_School\_Entry_j + u_{1j}. \tag{2}$$

A significant effect of time in school, after controlling for age at school entry, does not necessarily represent a significant schooling effect in this model, though. Age at school entry only provides a static between-person difference that cannot vary as a function of time in school. Any age-graded development that occurs after school entry will manifest in the model as an effect of the time in school variable.

To assess the relative effects of age versus time in school, we must assess whether the progression of time after school entry results in significantly greater gains in an outcome than would be expected given the progression of age alone. We propose doing this by testing the equality of the estimated effect of age (i.e., the effect of age at school entry, γ_{04} in the above model) and the conditional effects of time in school (i.e., $\gamma10$ if the effects of all covariates are zero and the appropriately calculated value of β_{1j} in all

[1] We note that the same goal could likely be accomplished in M*plus* using added model constraints.

Figure 1
Observed Ages and Time in School Values When the Head-Toes-Knees-Shoulders Task Was Administered



other cases). If a schooling effect existed, then we would expect to see an effect of time in school significantly greater than the effect of age at school entry. Figure 3 illustrates these possible associations.

Regression Models for a Stratified Random Sample

Although less powerful than the previous approach described, another method for eliminating redundancy between age and time in school is to remove the within-person component of the data altogether. That is, we can randomly select only one observation per child and fit a two-level model that accounts for nesting in classrooms:

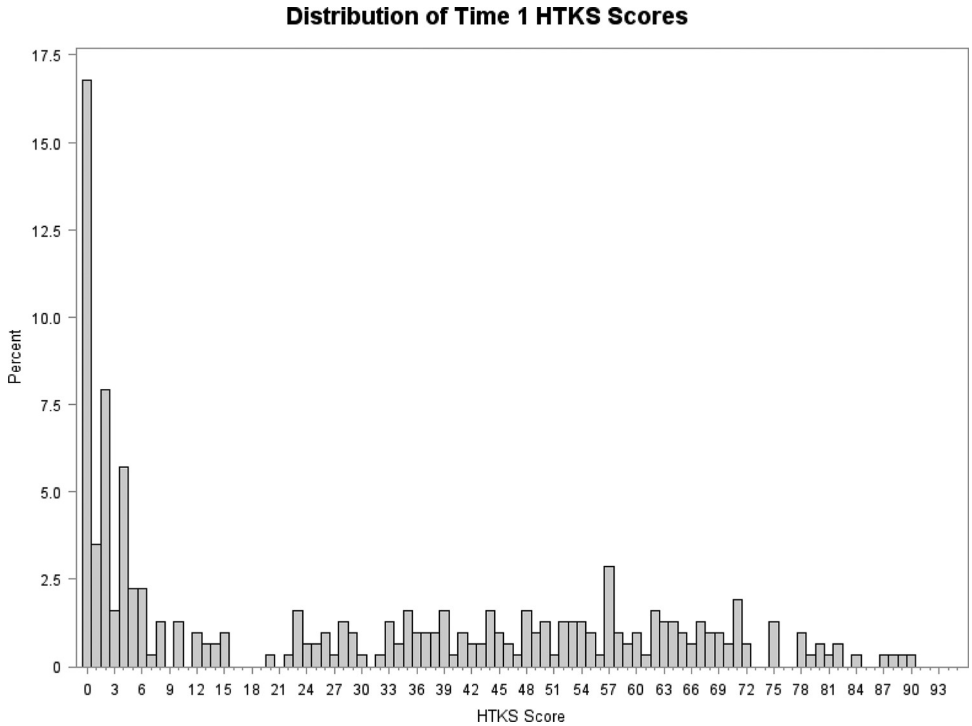
$$\begin{aligned}
Y_{jk} &= \beta_{0k} + \beta_{1k}Time_in_School_{jk} + \beta_{2k}Age_at_Assessment_{jk} \\
&\quad + \beta_{3k}Head_Start_{jk} + \beta_{4k}ELL_{jk} + \beta_{5k}Gender_{jk} + e_{jk} \\
\beta_{0k} &= \gamma_{00} + \gamma_{01}\overline{Time_in_School}_{.k} + \gamma_{02}\overline{Age_at_Assessment}_{.k} \\
&\quad + \gamma_{03}\overline{Head_Start}_{.k} + \gamma_{04}\overline{ELL}_{.k} + \gamma_{05}\overline{Gender}_{.k} + u_{0k} \\
\beta_{1k} &= \gamma_{10} + \gamma_{11}\overline{Time_in_School}_{.k} + \gamma_{12}\overline{Age_at_Assessment}_{.k} \\
&\quad + \gamma_{13}\overline{Head_Start}_{.k} + \gamma_{14}\overline{ELL}_{.k} + \gamma_{15}\overline{Gender}_{.k} + u_{1k} \\
\beta_{2k} &= \gamma_{20} + \gamma_{21}\overline{Time_in_School}_{.k} + \gamma_{22}\overline{Age_at_Assessment}_{.k} \\
&\quad + \gamma_{23}\overline{Head_Start}_{.k} + \gamma_{24}\overline{ELL}_{.k} + \gamma_{25}\overline{Gender}_{.k} + u_{2k}.
\end{aligned} \tag{3}$$

We took this approach for our third set of models, randomly selecting one observation (i.e., fall or spring) per child. The resulting dataset contained 167 observations in the fall and 158 observations in the spring. We ran the models in M*plus* Version 7.4 using TYPE = TWOLEVEL, where the multilevel component of the analysis treated observations as nested in classrooms. A significant effect for time in school would indicate a schooling effect. This model represents a modification of the naïve approach.

Pairwise Comparisons Using an Age-Restricted Stratified Random Sample

In addition to examining the joint influence of chronological age and school experience using linear or generalized linear models, we used coarsened exact matching to select a subset of fall and spring observations that were matched on age and then examined mean differences between the fall and spring assessments (e.g., Iacus et al., 2011). We used age at the administration of the HTKS when creating a matched sample, rounded to the nearest month. We then created samples of fall and spring observations that were matched on the rounded age variable as well as on Head Start status and ELL status. Part of the matching algorithm also ensured that only one observation per child would be included in the final matched sample and that the age range was restricted to between 56 and 64 months. Because the distribution of HTKS scores was nonnormal, we opted to examine schooling effects using Wilcoxon Rank Sum statistics. The Wilcoxon Rank-Sum test, which is equivalent to the Mann–Whitney U, compares the rank scores between two groups and can be thought of as a nonparametric analog to the independent

Figure 2
Distribution of Head-Toes-Knees-Shoulders (HTKS) Scores in the Fall of Preschool



samples *t* test.[2] Due to the limitations of nonparametric analyses, we ignored clustering within classrooms for these models.

Results

Descriptive statistics and correlations for the primary study variables are presented in Table 1. Children's fall and spring EF, preliteracy, and emergent math were strongly correlated with one another within and across time points (rs = .39 to .82, ps $< .001$). Age at entry to preschool was moderately correlated with fall EF ($r = .21$, $p < .001$), as well as emergent math in the fall ($r = .19$, $p < .001$) and spring ($r = .20$, $p < .001$), but it was not significantly correlated with preliteracy at either time point. Head Start status and ELL status were negatively correlated with children's fall and spring EF, preliteracy, and emergent math (rs = $-.21$ to $-.45$, ps $< .001$).

Naïve Regression Models for Simultaneous Time and Age Effects

Both age at time of assessment and time in school significantly predicted preliteracy skills, suggesting a significant schooling effect for preschool children's preliteracy skills beyond the effect of age. However, age at time of assessment, but not time in school, was a significant predictor of emergent math skills. A similar pattern of findings emerged for the count portion of the model for HTKS, but neither age nor time in school predicted excess zeros in the data. Thus, these results suggest a schooling effect for preliteracy but not for emergent math skills or executive function in preschool (see Table 2).

When examining the age and time in school variables, the raw data were correlated at approximately .63 ($p < .001$), with the caveat that this correlation necessarily conflates covariation at the between-person and between-classroom levels with exact collinearity at the within-person level.

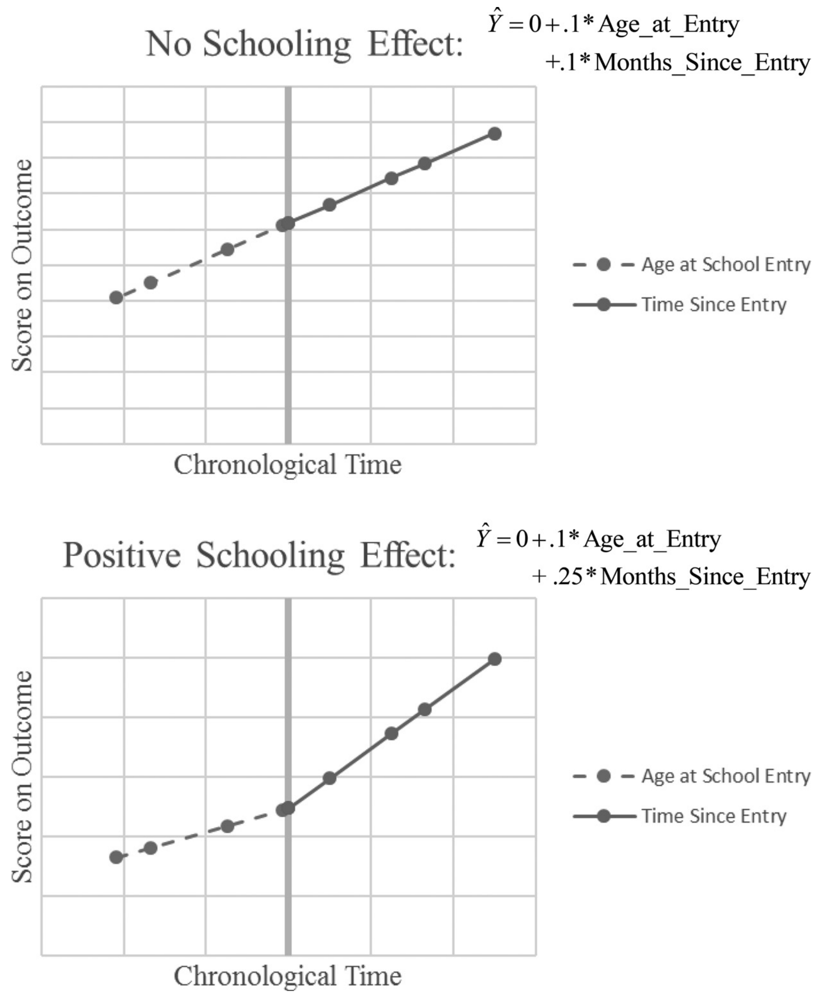
Regression Models for Time in School and Age at School Entry

In the following models, we model time in school at the within-person level only while modeling age at school entry at the between-person level only. Modeling each variable at a different level of analysis helps eliminate redundancy that might otherwise manifest in the data.

For preliteracy, preliminary models indicated that the slope of time in school did not significantly vary across individuals, scaled $\Delta\chi^2(df = 6) = 4.50$, $p = .61$. The multiple degrees of freedom represent a loss of slope variance and covariation between the slope and Level-2 variables, including the cross-level interaction between time in school and age at school entry. The random slope was dropped from the model as a means to improve parsimony. Time in school and age at school entry both significantly predicted preliteracy scores and a comparison of the slopes indicated a significant difference between the slopes (difference = -1.58 [SE = .42], $p < .001$), suggesting an added benefit of

[2] Note that we did not treat observations as paired because the matching did not occur on a case-by case basis but was instead stratified based on age and demographic variables.

Figure 3
Two Illustrative Examples of Schooling Effects From Spline Regression Models



Note. The dashed line represents the effect of age at school entry across individuals. The solid line represents the effect of time in school, treating the age at school entry as a constant (centered on the oldest individual at school entry). For the purpose of illustration, we assumed no interaction between Age at School Entry and Time in School.

schooling above and beyond age-based differences in preliteracy skills (see Table 3).

For math, preliminary models also indicated that the slope of time in school did not significantly vary across individuals, scaled $\Delta\chi^2(df = 6) = 3.26$, $p = .76$, and the random slope was dropped from the model. Time in school and age at school entry both significantly predicted emergent math scores, but a comparison of the slopes indicated no significant difference (difference = −.22 [*SE* = .27], $p = .41$), suggesting no added benefit of schooling above and beyond age-based differences in emergent math skills (see Table 3).

For EF, preliminary models indicated that random slopes for time predicting both the count and excess zero portions of the HTKS variable were not statistically significant, and neither were included in the reported HTKS model. The final regression model (see Table 3) indicated that both time in school and age at school entry predicted the count component of the HTKS, but these slopes were not significantly different from each other (difference = −.02 [*SE* = .02], $p = .21$). Time in school also predicted the zero-inflation component of HTKS, whereas the zero-inflation component was not significantly predicted by age at school entry. Despite this difference, the slopes themselves were not significantly different from each other (difference = .06 [*SE* = .05], $p = .28$). Therefore, the fitted model suggested no schooling effects for EF.

Regression Models for a Stratified Random Sample

The age and time in school variables remained strongly correlated in these models. Using the initial model for preliteracy as an example, age and time in school were correlated at .59 within classrooms and .88 between classrooms. The p value was $< .001$

Table 1
Descriptive Statistics and Correlations for Primary Study Variables

Variables	1	2	3	4	5	6	7	8	9	*M*	*SD*	Range
1. Fall HTKS	—									29.50	27.56	0−90
2. Spring HTKS	0.65***	—								43.19	28.92	0−93
3. Fall WJLW	0.39***	0.43***	—							334.80	26.92	264−477
4. Spring WJLW	0.35***	0.40***	0.82***	—						349.72	26.40	276−470
5. Fall WJAP	0.50***	0.59***	0.51***	0.55***	—					409.89	23.04	301−467
6. Spring WJAP	0.52***	0.62***	0.53***	0.55***	0.78***	—				420.02	21.33	332−481
7. Age at entry	0.21***	0.08	0.07	0.09	0.19***	0.20***	—			54.77	3.52	48.07−60.80
8. Gender (female)	−0.06	−0.06	0.00	−0.04	−0.07	−0.06	0.07	—		0.47	0.50	0−1
9. Head Start	−0.32***	−0.42***	−0.45***	−0.44***	−0.40***	−0.40***	0.12*	−0.03	—	0.55	0.50	0−1
10. ELL status	−0.27***	−0.34***	−0.21***	−0.24***	−0.41***	−0.44***	0.13*	0.10	0.40***	0.16	0.37	0−1

Note. HTKS = Head-Toes-Knees-Shoulders; WJLW = Woodcock Johnson Letter-Word Identification; WJAP = Woodcock Johnson Applied Problems; ELL = English-Language Learner.
* $p < .05$. *** $p < .001$.

at the within-person level but was not significant (i.e., $p = .96$) at the between-classroom level, indicating potential bias in the standard errors at the between-classroom level related to unresolved collinearity.

For preliteracy, initial models that included random slopes for age and time in school would not converge, nor would models that estimated the Level-2 variance of gender (model-estimated intraclass correlation coefficient [ICC] = .002; thus, indicating few differences in classroom gender compositions). Furthermore, between-classroom variation in preliteracy skills was not significantly correlated with classroom-level variation in age, time in school, or Head Start status, after controlling for classroom-level ELL prevalence, scaled $\Delta\chi^2(df = 3) = .13$, $p = .99$. Results from the final model indicated significant within-level effects for both age at time of assessment and time in school, suggesting a schooling effect for preliteracy skills (see Table 4).

For math, initial models that included random slopes for age and time in school would not converge, nor would models that estimated the Level-2 variance of gender (model-estimated ICC = .004; thus, indicating few differences in classroom gender compositions). Furthermore, between-classroom variation in emergent math skills was not significantly correlated with classroom-level variation in age, time in school, or Head Start status, after controlling for classroom-level ELL prevalence, scaled $\Delta\chi^2(df = 3) = .08$, $p = .99$. Results from the final model that included these constraints indicated a significant within-level effect for age at time of assessment but no added effect for time in school (see Table 4). Thus, the results did not support a schooling effect for emergent math skills.

Due to model complexity, random slopes for time in school and age at time of assessment were not considered in the model predicting EF skills. In addition, the models required that all between-cluster predictors only vary at the between-cluster level, so we manually decomposed time in school, age at assessment, gender, ELL status, and Head Start status into separate variables representing variation in cluster means and cluster mean-centered data. An initial model indicated that time in school, age at assessment, and gender did not significantly predict either component of HTKS at Level 2, and the Level-2 components of these predictors were dropped from the model (ΔBIC = −494.90; BIC = Bayesian information criterion). Results from the final model indicated a significant effect for age at time of assessment but no added effect for time in school for predicting the count component of HTKS at the within-classroom level, but neither measure of time was significant for predicting the probability of an excess zero (see Table 5). Thus, the results did not support a schooling effect for EF.

Pairwise Comparisons Using a Matched Sample

The coarsened exact match meant essentially no correlation between time and age at assessment (i.e., $r(250) = -.001$, $p = .98$) but a strong correlation between time and age at school entry (i.e.,

Table 2
Results From Naïve Regression Models Where Age at Time of Assessment and Time in School Simultaneously Predict Academic Scores

Predictor	Emergent math[a]	Preliteracy	HTKS count	HTKS zero inflation
Time in school (e_{1jk})	0.21 (0.31)	1.58 (0.39)***	0.02	−0.06
Age (β_{2jk})	1.77 (0.27)***	1.07 (0.36)**	0.05***	−0.06
Head Start (γ_{01k})	−12.82 (2.30)***	−23.15 (2.98)***	−0.56	0.86
ELL (γ_{02k})	−19.93 (2.91)***	−3.89 (3.84)	−0.60*	1.31
Male (γ_{03k})	−2.99 (1.90)	−2.14 (2.53)	−0.09	0.26
Level 1 residual (e_e)	109.38 (8.72)***	124.24 (9.90)***	0.48	NA
Level 2 variance (u_{0jk})	216.24 (23.74)***	419.02 (43.49)***	0.35	NA
Level 3 variance (u_{00k})	13.95 (11.86)	17.95 (25.27)	NA	NA

Note. ELL = English Language Learner; HTKS = Head-Toes-Knees-Shoulders. Parameter labels correspond to Equation 1; Level 1 residual for HTKS Count is a dispersion parameter. NA = not applicable.
[a] Standard errors in parentheses.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

Results From Spline Regression Models Where Age at School Entry and Time in School Simultaneously Predict Academic Scores and Executive Function

Predictor	Emergent math[a]	Preliteracy	HTKS count	HTKS zero inflation
Time in school (γ_{10})	1.97 (0.15)***	2.65 (0.18)***	0.08 (.01)***	−0.12 (.05)*
Age at school entry (γ_{04})	1.75 (0.21)***	1.07 (0.37)**	0.06 (.01)***	−0.06 (.04)
Male (γ_{03})	−2.79 (1.84)	−2.08 (2.42)	−0.09 (.08)	0.26 (.23)
ELL (γ_{02})	−20.56 (3.97)***	−4.84 (3.75)	−0.60 (.24)*	1.31 (.40)**
Head Start (γ_{01})	−13.23 (2.31)***	−23.40 (3.00)***	−0.56 (.23)*	0.86 (.35)*
Level 1 residual (e_{ij})	109.04 (10.04)***	123.84 (14.89)***	0.47 (.25)[b]	NA
Level 2 residual (u_{0j})	225.27 (33.41)***	427.72 (58.19)**	0.37 (51)	NA

Note. HTKS = Head-Toes-Knees-Shoulders; ELL = English Language Learner; Parameter labels correspond to Equation 2. NA = not applicable.
[a] Standard errors in parentheses. [b] Actually a dispersion factor.
* $p < .05$. ** $p < .01$. *** $p < .001$.

$r(250) = -.824$, $p < .001$). Thus, differences between time points represent the effect of schooling among age-matched peers.

Results from Wilcoxon Rank Sum tests comparing fall and spring assessments in matched samples replicated the regression-based findings, indicating a positive schooling effect for preliteracy skills but not early math skills or EF (see Table 6). After matching, the sample for these models included 126 observations per time point ($N = 252$). At each time point, the sample included 19 observations from ELL children enrolled in Head Start, 47 non-ELL children enrolled in Head Start, and 60 non-ELL children not enrolled in Head Start. Age at the time of the HTKS assessment was nearly identical across groups ($M = 59.46$, $SD = 1.97$ in the fall sample, $M = 59.45$, $SD = 1.95$ in the spring sample), reflecting the fact that we matched observations using this variable rounded to the nearest month. Estimated age at school entry was approximately separated by 6 months ($M = 57.11$, $SD = 1.94$ for the fall sample, $M = 51.47$, $SD = 1.95$ for the spring sample).

Discussion

The current study extends the body of literature investigating age- and schooling-related growth in school readiness skills by applying multiple analytic techniques meant to leverage natural variation in age within a single school grade. Using four alternative models, we were able to overcome several limitations of traditional methods, including nonrandom assignment to subsequent grades, the issue of quality in the counterfactual condition, and the restriction of data used in analyses. Notably, our results were consistent with findings from the school entrance cutoff method, propensity score matching, and regression discontinuity design, further validating these approaches for future application among researchers using diverse data collection strategies and increasing our confidence in the precision of the estimates. We first discuss substantive considerations within the broader context of the schooling effects literature and follow with a summary of the pros and cons of each method for researchers.

Aligning with previous research, all four modeling techniques revealed a significant schooling effect for preliteracy skills (e.g., Christian et al., 2000; Cunningham & Carroll, 2011; Kim & Morrison, 2018; Morrison et al., 1997, 1995). Given the robustness of findings across statistical methods, we interpret these effects as representing strong correlational evidence of the influence of preschool on preliteracy skills. We did not, however, observe a schooling effect for emergent math or EF. Prior work in these domains has been mixed, with a significant schooling effect for math and EF present in some studies (Bisanz et al., 1995; Brod et al., 2017; Christian et al., 2000; Loeb et al., 2007; Skibbe et al., 2013) but not in others (McCrea et al., 1999; Skibbe et al., 2011). Researchers hypothesize that the consistent schooling effect for early literacy skills, relative to the schooling effect for emergent math, may be a result of spending substantially more time in instructional activities related to reading in the preschool classroom compared with math (Morrison et al., 1997, 1995). Further, it may be the case that schooling effects on math are more apparent when assessing fine-grained math skills,

Table 4

Results From a Stratified Random Sample of Observations Where Age at Time of Assessment and Time in School Simultaneously Predict Academic Scores

	Preliteracy		Emergent math	
Predictor	Level 1[a]	Level 2	Level 1	Level 2
Time in school (γ_{10})	1.90 (0.56)**	NA	0.45 (0.40)	NA
Age (γ_{20})	0.89 (0.43)*	NA	1.75 (0.24)***	NA
Head Start (β_{3j}/γ_{03})	−21.76 (10.19)*	NA	−18.15 (4.54)***	NA
ELL (β_{4j}/γ_{04})	3.59 (3.57)	−66.87 (12.45)***	−13.70 (4.95)**	−52.77 (8.23)***
Male (β_{5j}/γ_{05})	−0.75 (2.60)	NA	−2.17 (2.06)	NA
Residual variance (e_{ij})	326.66 (52.06)***	5.44 (78.79)	302.56 (46.70)***	5.95 (16.14)

Note. ELL = English language learner; parameter labels correspond to Equation 3 specified as (Level 1/Level 2). NA = not applicable.
[a] Standard errors in parentheses.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Results From a Stratified Random Sample of Observations Where Age at Time of Assessment and Time in School Simultaneously Predict Executive Function

	Level 1[a]		Level 2	
Predictor	Count	Zero inflation	Count	Zero inflation
Time in school (γ_{10})	0.01 (.02)	0.04 (.11)	NA	NA
Age (γ_{20})	0.06 (.02)**	−0.09 (.05)	NA	NA
Head Start (β_{3j}/γ_{03})	−0.59 (.26)*	−0.40 (1.37)	−.25 (.13)	1.13 (.44)*
ELL (β_{4j}/γ_{04})	0.18 (.33)	2.13 (.65)**	−.72 (31)*	0.51 (.71)
Male (β_{5j}/γ_{05})	−0.05 (.08)	0.95 (.42)*	NA	NA
Residual variance (e_{ij})	NA	NA	.02 (.11)	NA
Dispersion	0.81 (.12)***	NA	NA	NA

Note. ELL = English language learner; parameter labels correspond to Equation 3 specified as (Level 1/Level 2); note that the dispersion parameter was not in that equation. NA = not applicable.
[a] Standard errors in parentheses.
* $p < .05$. ** $p < .01$. *** $p < .001$.

such as mental arithmetic, and may only be present in later grades as children develop more advanced cognitive abilities (Bisanz et al., 1995; Christian et al., 2000; Morrison et al., 1997). With regards to EF, one recent study documented a "catch up" effect (i.e., stronger schooling effect) among children who missed the cutoff to go to first grade and remained in kindergarten, suggesting that older children may be provided the context to practice EF when they attend a classroom with younger peers (e.g., Kim et al., 2021). Moreover, it is important to note that schooling effects on EF have mostly been documented with specific components of EF, such as working memory (e.g., Burrage et al., 2008; Finch, 2019). Our use of a global assessment may be one reason we did not find evidence for schooling effects on EF in this study. Regardless, the reasons behind the added benefit of schooling for preliteracy compared to the other domains of school readiness warrants further examination.

The current study also advances the field by introducing and discussing four theoretically informed alternative approaches to exploring age- and schooling-related growth in school readiness skills within a single school grade. Of the alternative models we implemented, the Naïve Regression Models for Simultaneous time and Age Effects is the most basic option for analyses. Although the conclusions drawn from this model in the present study are consistent with those of more advanced analyses, any application of this approach must come with the caveat that the progression of age and time in school are inherently collinear within individual participants. As illustrated in the Regression Models for a Stratified Random Sample approach, one can easily reduce redundancy in the data by sacrificing power and sample size and instead randomly selecting one observation per participant then modeling age and time in school as separate predictors. As our results showed, the stratified random sampling approach is not guaranteed to remove collinearity, however. A more rigorous approach, then, is the Regression Model for time in School and Age at School Entry, redefines age-graded development as reflecting each child's age at school entry and assesses equality of the time in school and age at entry slopes. Limitations to this model include the potential difficulty of capturing nonlinear development (e.g., quadratic growth as a function of age) and an assumption that age-graded differences before school entry reflect the same age-graded differences observed during the school year. The last approach, Pairwise Comparisons Using an Age-Restricted Stratified Random Sample, entails the exact matching of children who are similar ages at different waves of a study (e.g., fall vs. spring data collections; Goulet et al., 1974). This method closely mirrors the school cutoff approach but may allow for larger samples than the school entrance cutoff method. For instance, the school entrance cutoff method might select children born 3 months before a cutoff date to represent a later grade and children born 3 months after the cutoff date to represent an earlier grade, resulting in a 6-month age window. Matching as discussed above could produce a larger sample by pulling children from the full 6-month age window at both

Table 6

Descriptive and Inferential Statistics (Wilcoxon Ranked Sum) for the Matched Sample

Outcome	*M*	*SD*	Median	*W*	*p*
Math					
Fall	412.24	21.38	416.00		
Spring	415.78	19.14	419.00	15,290.50	.26
Preliteracy					
Fall	337.71	30.15	336.00		
Spring	348.41	24.81	349.00	14,073.50	.001
Executive function					
Fall	33.28	27.95	36.50		
Spring	39.49	29.23	45.50	14,940.00	.09

time points or could produce a more closely matched sample by pulling from the same 3-month window at both time points.

To summarize, we consider our study as replicating evidence of the schooling effect on preliteracy using several methodologically rigorous statistical designs that reduce the confounding effects of age and time within a single school year and overcome bias that may be present in traditional methods. This makes them potentially useful alternatives for researchers who use a variety of data collection protocols.

In addition to their strengths, each of the models also contain limitations, and researchers intending to apply one or more of these models to their own data should carefully weigh the pros and cons along with those of more common approaches when deciding on the most appropriate analytic strategy or strategies. Moreover, while we have discussed the advantages and disadvantages of each approach in detail, it is important to note that these are only a subset of the diverse modeling possibilities that can be adapted to meet the needs of individual data sets. Researchers should continue to explore novel approaches for disentangling the effects of age and schooling. Further, future work that investigates the processes underlying such effects, including the role of classroom instructional quality, is critical for understanding the mechanisms that lead to intended schooling-related growth. For instance, research suggests that it may be important to consider the amount, quality, and content of instruction together to capture the full complexity of classroom environments and their influence on achievement (Connor et al., 2014). Identifying these nuances will be a necessary step for informing interventions, programs, and practices within school-based education settings.

Implications

There are several broader implications of this work. Specifically, the alternative approaches described in this study expand the toolkit of quasi-experimental methods available to researchers who are interested in investigating policy and program effects (Gopalan et al., 2020). For instance, propensity score matching has become a popular technique for studying the effects of educational interventions, including full day kindergarten and Head Start (Leow & Wen, 2017). However, using the methods presented in this study may help researchers to overcome certain data limitations, such as those associated with estimating dosage across cohorts. Similarly, these approaches may complement the regression discontinuity design when used in the context of examining summer learning loss or summer programming, which may also be confounded with maturation (Finch, 2019; Zvoch & Stevens, 2011). Moreover, practitioners and administrators interested in evaluating kindergarten age cutoffs may leverage these methods to investigate the effects of varying eligibility policies for kindergarten entrance (Datar, 2006; Elder & Lubotsky, 2009). Finally, these alternative approaches offer additional opportunities for researchers to conduct replication analyses and implement robustness-checking practices, which may ultimately serve to enhance the integrity of developmental science (Duncan et al., 2014).

Limitations and Future Directions

Although our results support the robustness of previous findings using different analytic approaches, they can only provide correlational evidence of schooling effects. In the absence of RCTs, none of the approaches we highlighted deliver evidence of causality. An additional limitation of this study is the fact that a moderate proportion of the sample scored zero on the HTKS. This phenomenon has been documented in younger samples from families with low incomes (e.g., Schmitt et al., 2015), which is why we took into account recent recommendations to include the practice items in the total score (Fuhs et al., 2014; Gonzales et al., 2021). Despite these corrections, approximately 17% of the sample still demonstrated floor effects. When possible, we utilized modeling techniques that would allow us to examine these children independently from those with scores on the task (e.g., zero-inflated negative binomial regression). Yet, the task distribution may have prevented us from detecting schooling effects on EF skills that exist when measured in ways that capture more variability. Therefore, although we see it as a benefit that our measure of EF skills tapped into all three EF processes manifested in self-regulated behaviors, future work should explore schooling-related growth using a diverse set of standardized measures that are both unique to different EF processes and sensitive to developmental changes. Extending these designs to include other assessments of EF skills is a critical next step, as is replicating findings in larger and more diverse samples.

Relatedly, our analytic methods also faced some shortcomings that must be addressed. As an example, we were not able to fit a three-level model for the HTKS given that we determined a zero-inflated negative binomial distribution was warranted. Additionally, the methods we examined only relied on simultaneous estimation. Future research could build on these approaches, for example, by incorporating model-building approaches that test the added effect of each time metric after accounting for the other. Moreover, research should examine theoretically informed subgroup effects, such as whether schooling matters more for children based on their home learning activities (Coley et al., 2020).[3] Finally, aside from including a proxy for children from families with low incomes in our models (e.g., Head Start status), we were unable to account for variations in the quality of preschool experiences that children were exposed to, or the amount of time spent in academic instruction, which may influence schooling-related growth in early literacy skills (Fuller et al., 2017; McGinty et al., 2011).

Conclusion

In the present study, we revisited the question of whether schooling-related growth contributes to the development of school readiness skills above and beyond age-related maturation. We demonstrated how multiple alternative methods can be applied to studies examining the impacts of schooling within a single year to overcome potential biases present in traditional approaches. Our results suggest an added benefit of the preschool experience for fostering preliteracy skills but not emergent math or EF skills. These findings validate evidence from existing studies on schooling effects and contribute to a more comprehensive

[3] Based on a reviewer suggestion we tested subgroup differences for children who attended Head Start versus other programs and we did not find a different pattern of results.

understanding of the role of schooling in early skill development. Nevertheless, more research is needed to determine why schooling effects do not consistently emerge for the domains of EF and math and examine the mediating mechanisms through which preschool experience influences school readiness skills. Researchers are additionally encouraged to replicate these findings using larger and more diverse samples with differing preschool experiences and curricula.

References

Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling*, *26*(1), 119–142. https://doi.org/10.1080/10705511.2018.1511375

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Baltes, P. B., Baltes, M. M., & Reinert, G. (1970). The relationship between time of measurement and age in cognitive development of children: An application of cross-sectional sequences. *Human Development*, *13*(4), 258–268. https://doi.org/10.1159/000270894

Bentin, S., Hammer, R., & Cahan, S. (1991). The effects of aging and first grade schooling on the development of phonological awareness. *Psychological Science*, *2*(4), 271–275. https://doi.org/10.1111/j.1467-9280.1991.tb00148.x

Bisanz, J., Morrison, F. J., & Dunn, M. (1995). Effects of age and schooling on the acquisition of elementary quantitative skills. *Developmental Psychology*, *31*(2), 221–236. https://doi.org/10.1037/0012-1649.31.2.221

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*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does one year of schooling improve children's cognitive control and alter associated brain activation? *Psychological Science*, *28*(7), 967–978. https://doi.org/10.1177/0956797617699838

Burrage, M. S., Ponitz, C. C., McCready, E. A., Shah, P., Sims, B. C., Jewkes, A. M., & Morrison, F. J. (2008). Age- and schooling-related effects on executive functions in young children: A natural experiment. *Child Neuropsychology*, *14*(6), 510–524. https://doi.org/10.1080/09297040701756917

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, *60*(5), 1239–1249. https://doi.org/10.2307/1130797

Castro, D. C., Páez, M. M., Dickinson, D. K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child Development Perspectives*, *5*(1), 15–21. https://doi.org/10.1111/j.1750-8606.2010.00142.x

Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development*, *1*(4), 429–448. https://doi.org/10.1207/S15327647JCD0104_04

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, *28*(4), 415–427. https://doi.org/10.1016/j.econedurev.2008.09.003

Coley, R. L., Kruzik, C., & Votruba-Drzal, E. (2020). Do family investments explain growing socioeconomic disparities in children's reading, math, and science achievement during school versus summer months? *Journal of Educational Psychology*, *112*(6), 1183–1196. https://doi.org/10.1037/edu0000427

Connor, C. M., Spencer, M., Day, S. L., Giuliani, S., Ingebrand, S. W., McLean, L., & Morrison, F. J. (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*, *106*(3), 762–778. https://doi.org/10.1037/a0035921

Cooper, C. E., Osborne, C. A., Beck, A. N., & McLanahan, S. S. (2011). Partnership instability, school readiness, and gender disparities. *Sociology of Education*, *84*(3), 246–259. https://doi.org/10.1177/0038040711402361

Crone, D. A., & Whitehurst, G. J. (1999). Age and schooling effects on emergent literacy and early reading skills. *Journal of Educational Psychology*, *91*(4), 604–614. https://doi.org/10.1037/0022-0663.91.4.604

Cunningham, A., & Carroll, J. (2011). Age and schooling effects on early literacy and phoneme awareness. *Journal of Experimental Child Psychology*, *109*(2), 248–255. https://doi.org/10.1016/j.jecp.2010.12.005

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, *25*(1), 43–62. https://doi.org/10.1016/j.econedurev.2004.10.004

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, *50*(11), 2417–2425. https://doi.org/10.1037/a0037996

Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers. *The Journal of Human Resources*, *44*(3), 641–683. https://doi.org/10.3368/jhr.44.3.641

Finch, J. E. (2019). Do schools promote executive functions? Differential working memory growth across school-year and summer months. *AERA Open*, *5*(2), 2332858419848443. https://doi.org/10.1177/2332858419848443

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, *50*(6), 1698–1709. https://doi.org/10.1037/a0036633

Fuller, B., Bein, E., Bridges, M., Kim, Y., & Rabe-Hesketh, S. (2017). Do academic preschools yield stronger benefits? Cognitive emphasis, dosage, and early learning. *Journal of Applied Developmental Psychology*, *52*, 1–11. https://doi.org/10.1016/j.appdev.2017.05.001

Gonzales, C. R., Bowles, R., Geldhof, G. J., Cameron, C. E., Tracy, A., & McClelland, M. M. (2021). The Head-Toes-Knees-Shoulders Revised (HTKS-R): Development and psychometric properties of a revision to reduce floor effects. *Early Childhood Research Quarterly*, *56*, 320–332. https://doi.org/10.1016/j.ecresq.2021.03.008

Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, *44*(1), 218–243. https://doi.org/10.3102/0091732X20903302

Gormley, W. T., Jr., Phillips, D., & Gayer, T. (2008). The early years. Preschool programs can boost school readiness. *Science*, *320*(5884), 1723–1724. https://doi.org/10.1126/science.1156019

Goulet, L. R., Williams, K. G., & Hay, C. M. (1974). Longitudinal changes in intellectual functioning in preschool children: Schooling-and age-related effects. *Journal of Educational Psychology*, *66*(5), 657–662. https://doi.org/10.1037/h0037485

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis.

Research in Human Development, *6*(2–3), 97–120. https://doi.org/10.1080/15427600902911189

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345–361. https://doi.org/10.1198/jasa.2011.tm09599

Kim, M. H., & Morrison, F. J. (2018). Schooling effects on literacy skills during the transition to school. *AERA Open*, *4*(3), 1–15. https://doi.org/10.1177/2332858418798793

Kim, M. H., Ahmed, S. F., & Morrison, F. J. (2021). The effects of kindergarten and first grade schooling on executive function and academic skill development: Evidence from a school cutoff design. *Frontiers in Psychology*, *11*, 607973. https://doi.org/10.3389/fpsyg.2020.607973

Lee, R., Zhai, F., Brooks-Gunn, J., Han, W. J., & Waldfogel, J. (2014). Head start participation and school readiness: Evidence from the Early Childhood Longitudinal Study-Birth Cohort. *Developmental Psychology*, *50*(1), 202–215. https://doi.org/10.1037/a0032280

Leow, C., & Wen, X. (2017). Is full day better than half day? A propensity score analysis of the association between Head Start Program intensity and children's school performance in kindergarten. *Early Education and Development*, *28*(2), 224–239. https://doi.org/10.1080/10409289.2016.1208600

Lerner, R. M., Schwartz, S. J., & Phelps, E. (2009). Problematics of time and timing in the longitudinal study of human development: Theoretical and methodological issues. *Human Development*, *52*(1), 44–68. https://doi.org/10.1159/000189215

Li, W., Farkas, G., Duncan, G. J., Burchinal, M. R., & Vandell, D. L. (2013). Timing of high-quality child care and cognitive, language, and preacademic development. *Developmental Psychology*, *49*(8), 1440–1451. https://doi.org/10.1037/a0030613

Lipsey, M. W., Weiland, C., Yoshikawa, H., Wilson, S. J., & Hofer, K. G. (2015). The prekindergarten age-cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis*, *37*(3), 296–313. https://doi.org/10.3102/0162373714547266

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*(1), 52–66. https://doi.org/10.1016/j.econedurev.2005.11.005

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, *41*(1), 115–157. https://doi.org/10.3102/00028312041001115

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. https://doi.org/10.3389/fpsyg.2014.00599

McClelland, M. M., John Geldhof, G., Cameron, C. E., & Wanless, S. B. (2015). Development and self-regulation. In L. M. Lerner (Ed.), *Handbook of child psychology and developmental science* (pp. 1–43). Wiley. https://doi.org/10.1002/9781118963418.childpsy114

McCrea, S. M., Mueller, J. H., & Parrila, R. K. (1999). Quantitative analyses of schooling effects on executive function in young children. *Child Neuropsychology*, *5*(4), 242–250. https://doi.org/10.1076/0929-7049(199912)05:04;1-R;FT242

McGinty, A. S., Breit-Smith, A., Fan, X., Justice, L. M., & Kaderavek, J. N. (2011). Does intensity matter? Preschoolers' print knowledge development within a classroom-based intervention. *Early Childhood Research Quarterly*, *26*(3), 255–267. https://doi.org/10.1016/j.ecresq.2011.02.002

Morrison, F. J., Griffith, E. M., & Alberts, D. M. (1997). Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, *33*(2), 254–262. https://doi.org/10.1037/0012-1649.33.2.254

Morrison, F. J., Kim, M. H., Connor, C. M., & Grammer, J. K. (2019). The causal impact of schooling on children's development: Lessons for developmental science. *Current Directions in Psychological Science*, *28*(5), 441–449. https://doi.org/10.1177/0963721419855661

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology*, *31*(5), 789–799. https://doi.org/10.1037/0012-1649.31.5.789

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Pruebas de aprovechamiento*. Riverside Publishing.

Newgard, C. D., Hedges, J. R., Arthur, M., & Mullins, R. J. (2004). Advanced statistics: The propensity score—a method for estimating treatment effect in observational research. *Academic Emergency Medicine*, *11*(9), 953–961. https://doi.org/10.1111/j.1553-2712.2004.tb00788.x

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*(1), 133–164. https://doi.org/10.3102/00028312039001133

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., Jenkins, F., Fletcher, P., Quinn, L., Friedman, J., Ciarico, J., Rohacek, M., Adams, G., & Spier, E. (2010). *Head Start Impact Study: Final Report*. Administration for Children & Families.

Sabol, T. J., & Pianta, R. C. (2017). The state of young children in the United States: School readiness. In E. Votruba-Drzal & E. Dearing (Eds.), *Handbook of early childhood development programs, practices, and policies* (pp. 3–17). Wiley. https://doi.org/10.1002/9781118937334.ch1

Schaie, K. W. (1972). Limitations on the generalizability of growth curves of intelligence. *Human Development*, *15*(3), 141–152. https://doi.org/10.1159/000271238

Schmitt, S. A., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2015). Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly*, *30*, 20–31. https://doi.org/10.1016/j.ecresq.2014.08.001

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academy Press.

Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2011). Schooling effects on preschoolers' self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly*, *26*(1), 42–49. https://doi.org/10.1016/j.ecresq.2010.05.001

Skibbe, L. E., Hindman, A. H., Connor, C. M., Housey, M., & Morrison, F. J. (2013). Relative contributions of prekindergarten and kindergarten to children's literacy and mathematics skills. *Early Education and Development*, *24*(5), 687–703. https://doi.org/10.1080/10409289.2012.712888

Son, S. H., & Morrison, F. J. (2010). The nature and impact of changes in home learning environment on development of language and academic skills in preschool children. *Developmental Psychology*, *46*(5), 1103–1118. https://doi.org/10.1037/a0020065

Wanless, S. B., McClelland, M. M., Acock, A. C., Ponitz, C. C., Son, S.-H., Lan, X., Morrison, F. J., Chen, J.-L., Chen, F.-M., Lee, K., Sung, M., & Li, S. (2011). Measuring behavioral regulation in four cultures. *Psychological Assessment*, *23*(2), 364–378. https://doi.org/10.1037/a0021768

Wanless, S. B., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2011). The influence of demographic risk factors on children's behavioral regulation in prekindergarten and kindergarten. *Early Education and Development*, *22*(3), 461–488. https://doi.org/10.1080/10409289.2011.536132

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. https://doi.org/10.1111/cdev.12099

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development* (pp. 125–148). University of Minnesota Press.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Riverside Publishing Company.

Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2014). Head Start's impact is contingent on alternative type of care in comparison group. *Developmental Psychology*, *50*(12), 2572–2586. https://doi.org/10.1037/a0038205

Zvoch, K., & Stevens, J. J. (2011). Summer school and summer learning: An examination of the short-and longer term changes in student literacy. *Early Education and Development*, *22*(4), 649–675. https://doi.org/10.1080/10409289.2010.489891

Received April 24, 2021
Revision received March 1, 2022
Accepted April 20, 2022 ■