Published in *Educational Researcher*
Online First, October 19, 2023
https://doi.org/10.3102/0013189X231203646

Classroom Concentration of English learners and Their Reading Growth

Michael J. Kieffer ORCID: 0000-0001-8000-7832

Andrew W. Weaver ORCID: 0000-0002-2278-0517

New York University

Authors' bios: Michael J. Kieffer, Ed.D. is associate professor of literacy education at New York University, 239 Greene Street, New York, NY 10012; michael.kieffer@nyu.edu. His research focuses on language and literacy development in multilingual contexts. Andrew Weaver, M.Ed., is a doctoral candidate in Teaching and Learning at New York University, 239 Greene St, New York, NY 10003; aww304@nyu.edu. His research focuses on the role of social ties in the language and literacy development of multilingual students.

Acknowledgments: This study was supported, in part, by the Institute of Education Sciences, U.S. Department of Education through grant R305A200069 to New York University. Weaver's work on this study was supported, in part, 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 authors and do not represent views of the Institute or the U.S. Department of Education.

Running Head: EL Concentration

Classroom Concentration of English learners and Their Reading Growth

Abstract

In this brief, we use a national representative sample of ever-English learners (N = 783) to examine relations between English learner concentration within classrooms and reading growth between Kindergarten and Grade 5. Piecewise growth models were used to estimate relations for four developmental periods (K-1, Grade 1-2, Grade 2-3, and Grade 3-5). Results indicated non-significant, trivially sized relations for classroom English learner concentration across periods, with and without controls for school concentration and student characteristics. Results were robust across multiple specifications. Findings call into doubt the academic benefits of the common practice of grouping English learners together in particular classrooms.

Classroom Concentration of English learners and Their Reading Growth

Multilingual students classified as English learners (ELs) are often concentrated within particular classrooms and schools (Gándara & Orfield, 2012; National Academies of Sciences, Engineering, and Medicine [NASEM], 2017). This concentration reflects a variety of structural and geographic factors, but also results from intentional efforts to group students into particular classrooms based on their language needs (e.g., Estrada et al., 2020). The rationale for grouping ELs together within schools is that this will allow for more targeted and effective language instruction than would be possible in linguistically heterogeneous classrooms. This rationale results, in part, from an interpretation of policy mandates to provide special services to ELs, including those in the Every Student Succeeds Act and its precursors. Such mandates rarely specify isolated instruction for ELs, but they frequently lead educators to group ELs together to provide them mandated services (e.g., Estrada et al., 2020; Gándara & Orfield, 2012). This practice may also stem from educators' interpretation of guidelines advising targeted language instruction for ELs (e.g., Baker et al., 2014). If delivering such targeted services in concentrated classrooms is effective, it may be most consequential for reading, as English reading development (from early foundational skills to reading comprehension) involves linguistic processes that depend heavily on English language comprehension (August & Shanahan, 2006).

Despite the common practice of grouping ELs into specific classrooms, a wealth of theory and research criticize such efforts as harmful for both academic and social outcomes (Estrada et al., 2020; Gándara & Orfield, 2012). Theories of second language development and socialization (e.g., Ellis, 2005; Wong Fillmore, 1991) emphasize the importance of interaction with more proficient peers for language learning. Similarly, models of reading development (e.g., RAND Reading Study Group, 2002) acknowledge the sociocultural context—including classroom composition—as key to reading growth. Empirical work by Estrada and colleagues (2020) has demonstrated that ELs in elementary California classrooms with higher proportions of ELs had lower performance on state tests of English-language arts, math, and English language proficiency, controlling for various covariates. Such research and theory call into question the effectiveness of this pervasive policy.

Current Study

This study examines whether classroom EL concentration predicts ever-ELs' reading growth between Kindergarten and Grade 5 using a nationally representative sample of ever-ELs ($N = 783$) drawn from the Early Childhood Longitudinal Study-Kindergarten 2011 Cohort (ECLK:2011; Tourangeau et al., 2019). (Data are available at https://nces.ed.gov/ecls/kindergarten2011.asp). Ever-ELs were identified as students whose kindergarten teacher reported that they "participate in an instructional program designed to teach English language skills to children with limited English proficiency" (National Center for Educational Statistics, 2011, p. 13) Consistent with the Ever-EL or Total-EL framework (e.g., Hopkins et al., 2013), students remained in the longitudinal sample in later grades, even if they were no longer enrolled in EL services (see robustness checks regarding this decision in Supporting Online Material [SOM]). Sample descriptives are provided in SOM Table S1.

The outcome measure was an adaptive measure of overall reading achievement that emphasized basic reading skills in the early grades and reading comprehension in Grades 3-5 (Najarian et al., 2019; See SOM). Classroom EL concentration was reported by teachers. We controlled for administrator-reported school EL concentration as well as child and school characteristics. Child controls included socioeconomic status (ECLS-K:2011 composite), English oral proficiency, teacher-reported social-emotional skills (externalizing behaviors, internalizing behaviors, social skills), and executive functioning (cognitive flexibility, working memory, teacher-reported attentive behavior and inhibitory control). School controls including percent free-reduced lunch and percent students of color. All controls were measured in the same or a previous year to EL concentration.

We first describe school and classroom concentration of ever-ELs. Then, using latent growth modeling with clustered standard errors to account for nesting of children within schools, we examine whether classroom concentration predicts each slope for our best fitting piecewise model. We compare results in uncontrolled models and in models controlling for child and school covariates. Finally, we conduct a series of robustness checks to test alternative specifications. Based on prior theory and empirical work (e.g., NASEM, 2017; Estrada et al., 2020), we hypothesized that classroom EL concentration would negatively predict later growth in reading.

Results

Preliminary Analyses

The average ever-EL attends a school that enrolls 36%-40% ELs (depending on grade) and is enrolled in a classroom that enrolls 41%-60% ELs (SOM Table S2). By contrast, the average never-EL attends a school with much lower school EL concentrations (9%-11% ELs) and classroom EL concentrations (7%-15% ELs). Given the evidence of high EL concentration at both the school level and classroom level, we control for school EL concentration in our models to account for the contextual factors that may drive classroom EL concentration and to isolate (to some extent at least) the effects of intentional grouping.

Preliminary unconditional growth modeling indicated that a four-slope piecewise model provided the best fit to the data (see SOM). As shown in Figure 1, this model includes slopes from Fall, Kindergarten to Fall, Grade 1; from Fall, Grade 1 to Fall, Grade 2; from Fall, Grade 2 to Spring, Grade 3; and from Spring, Grade 3 to Spring, Grade 5. The average reading trajectory demonstrated a trend of decelerating growth over time, as found in prior research with ECLS-K and other datasets (e.g., Kieffer, 2011; Relyea & Amendum, 2020).

Classroom EL Composition Predicting Reading Growth

Classroom EL concentration (measured toward the beginning of each of the four time periods) did not significantly predict any of the four slopes for reading growth (all *p*s > .05). As shown in Table 1, these results were the same whether or not we controlled for school EL concentration along with student and school characteristics. The standardized path coefficients for classroom EL concentration were all trivial to small in magnitude (-0.082 to 0.084), and the standard errors were relatively small, indicating good precision for these null effects (see Figure 2).

Robustness Checks

Results were robust to the inclusion of various student and school controls, interactions with the class language program (i.e., English-as-a-second-language, bilingual programming or no services), and different specifications for the relations between slopes and intercept (see SOM). In addition, because our "ever-EL" specification combined both current and former ELs in a way that may dilute the link between EL concentration and English reading, we tested interactions to see whether the class EL concentration effects were different for current ELs in the most relevant grade for a given slope, compared to ever-ELs; these interactions were all nonsignificant, indicating the effects are not different for current ELs ($p > .05$; see SOM Table S7).

Discussion

In this study, we explored the relations between classroom EL concentration and reading growth in the elementary grades. Consistent with prior research (e.g., Estrada et al., 2020; NASEM, 2017), we found evidence of high concentration of ELs within particular classrooms, higher than we would expect based on school EL concentration alone. Based on arguments about the harms of segregation of various kinds (e.g., Gándara & Orfield, 2012) and about the importance of interaction with more proficient peers for language learning (e.g., Ellis, 2005; Wong Fillmore, 1991), we hypothesized that higher EL concentrations would be associated with slower reading growth. However, we found no evidence for such a negative effect. Rather, we found consistent null effects across four developmental periods. Given our large sample size, all of the effects had relatively small standard errors indicating “well-measured zeros” and providing good confidence that any effects we could not detect as statistically significant would be small or trivial in magnitude. Results were robust across a variety of alternative specifications. Despite the surprising lack of negative effects, the absence of positive effects raises questions about the common assumptions that underlie educators’ efforts to separate ELs into distinct classrooms.

There are multiple plausible explanations for our null effects, each of which raises questions for future research. In interpreting these effects, we draw on ecological theories of language learning (e.g., Kramsch, 2002; van Lier, 2002) that emphasize that all environments have particular affordances for language learning and related reading development. Although non-EL peers are one such affordance, others may be equally or more important. One possibility is that there are positive and negative effects of the various affordances that come with classroom EL concentration that cancel each other out. For instance, the positive affordance of more targeted language instruction in high-EL classrooms might be a benefit that is washed out by the lack of interaction with non-EL peers. Another possibility is that the quality of instruction (which we could not observe) is more predictive than the composition of the classroom. We found no interactions between EL concentration and the classroom language program (i.e., English-as-a-second-language, bilingual instruction, or no services), but these are relatively coarse categories. High-quality language instruction (or the lack thereof) may take place to a similar extent in classrooms with high and low EL concentration. Finally, it is possible that simply enrolling ELs with non-ELs in the same classrooms is necessary, but insufficient for reading growth, especially if ELs are not provided with many opportunities to interact with their non-EL peers during instruction, given the importance of such interaction for language learning (e.g., Ellis, 2005).

Well-intentioned educators separate ELs from non-ELs for instruction on the premise that this will facilitate more targeted language instruction and result in more rapid progress in reading and other areas. Our findings do not support this premise. Although we do not disagree with the research base on targeted language instruction for ELs (e.g., Baker et al., 2014), our findings suggest that educators should be cautious in assuming such instruction should be provided in isolated settings. Even if there may not be academic harms, concentrating ELs in particular classrooms may have negative social effects, restricting ELs' access to the linguistic, social, and cultural capital offered by their non-EL peers (e.g., Estrada et al., 2020; Gándara & Orfield, 2012).

This study has several limitations to note. First, although we conducted a variety of robustness checks, there was limited information in the ECLS-K:2011 dataset to investigate some hypotheses. For instance, we suspect that higher EL concentrations may be associated with more EL-specific resources (e.g., specialized EL curricula or bilingual paraprofessionals), but the ECLS-K:2011 dataset did not provide data to explore these questions. Second, although our data suggest greater classroom EL concentration than we would expect based on school EL concentration, our data do not allow us to fully isolate the extent to which this is due to intentional grouping as opposed to other factors. Our findings are robust to the inclusion of school EL concentration as a control, but further research on how educators make decisions to group ELs is warranted. Third, the available data on both EL status and classroom EL concentration were teacher-reported, so are prone to human error. Fourth, we focused on English reading because of its heavy English language demands, but further analyses in other subject areas are warranted. Fifth, future research should investigate other outcomes beyond academic performance, such as socioemotional skills and student self-concept, which may be more susceptible to EL concentration effects. Finally, future work using multiple-group approaches to measure academic growth (Kohli et al, 2015; Sullivan et al., 2016) to investigate the effects of EL concentration on students who are never ELs as compared to effects on ever-ELs is needed.

Despite these limitations, this study provides valuable new evidence on the relations between EL concentration and reading growth. They call into question the logic of educators' widespread practices of separating ELs from non-ELs for instruction.

References

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Routledge.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., ... & Newman-Gonchar, R. (2014). Teaching Academic Content and Literacy to English Learners in Elementary and Middle School. IES Practice Guide. NCEE 2014-4012. *What Works Clearinghouse*.

Carnock, J.T., and Ege, E. (2015). The Triple Segregation of Latino ELLs: What Can We Do? https://www.newamerica.org/education-policy/edcentral/latinos-segregation

Ellis, R. (2005). Principles of instructed language learning. System, 33, 209–224. doi.org/10.1016/j.system.2004.12.006

Estrada, P., Wang, H., & Farkas, T. (2020). Elementary English learner classroom composition and academic achievement: The role of classroom-level segregation, number of English proficiency levels, and opportunity to learn. *American Educational Research Journal, 57*(4), 1791–1836. doi.org/10.3102/0002831219887137

Gándara, P., & Orfield, G. (2012). Segregating Arizona's English learners: A return to the "Mexican Room"?. *Teachers College Record*, *114*(9), 1-27. doi.org/10.1177/016146811211400905

Hopkins, M., Thompson, K. D., Linquanti, R., Hakuta, K., & August, D. (2013). Fully accounting for English learner performance: A key issue in ESEA reauthorization. *Educational Researcher*, *42*(2), 101-108. doi.org/10.3102/0013189X12471426

Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to grade 8. *American Educational Research Journal*, *48*(5), 1187-1225. doi.org/10.3102/0002831211419490

Kohli, N. Sullivan, A. L., Sadeh, S., Zopluoglu, C. (2015). Longitudinal mathematics development of students with learning disabilities and students without disabilities: A comparison of linear, quadratic, and piecewise linear mixed effects models. *Journal of School Psychology, 53,* 105-120. doi.org/10.1016/j.jsp.2014.12.002

Kramsch, C. J. (Ed.). (2002). *Language acquisition and language socialization: Ecological perspectives*. Continuum.

Najarian, M., Tourangeau, K., Nord, C., Wallner-Allen, K., and 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. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubsearch.

National Academies of Sciences, Engineering, and Medicine. (2017). Promoting the Educational Success of Children and Youth Learning English: Promising Futures. Washington, DC: The National Academies Press.

National Center for Educational Statistics. (2011). Spring 2011 kindergarten teacher questionnaire (child level), Early Childhood Longitudinal Study: Kindergarten class of 2010-11. https://nces.ed.gov/ecls/pdf/kindergarten2011/ Spring_K_Classroom_Teacher_Child_Level.pdf

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

Relyea, J. E., & Amendum, S. J. (2020). English reading growth in Spanish-speaking bilingual students: Moderating effect of English proficiency on cross-linguistic influence. *Child Development*, *91*(4), 1150-1165. doi.org/10.1111/cdev.13288

Sulivan, A. L., Kohli, N., Farnsworth, E. M., Sadeh, S., & Jones, L. (2017). Longitudinal models of reading achievement of students with learning disabilities and without disabilities. *School Psychology Quarterly, 32*, 336-349. doi.org/10.1037/spq0000170

van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Springer.

Wong Fillmore, L. (1991). Second language learning in children: A model of language learning in social context. In Bialystok (Ed.), *Language processing in bilingual children* (pp. 49-69). Cambridge, UK: Cambridge University Press.

Table 1

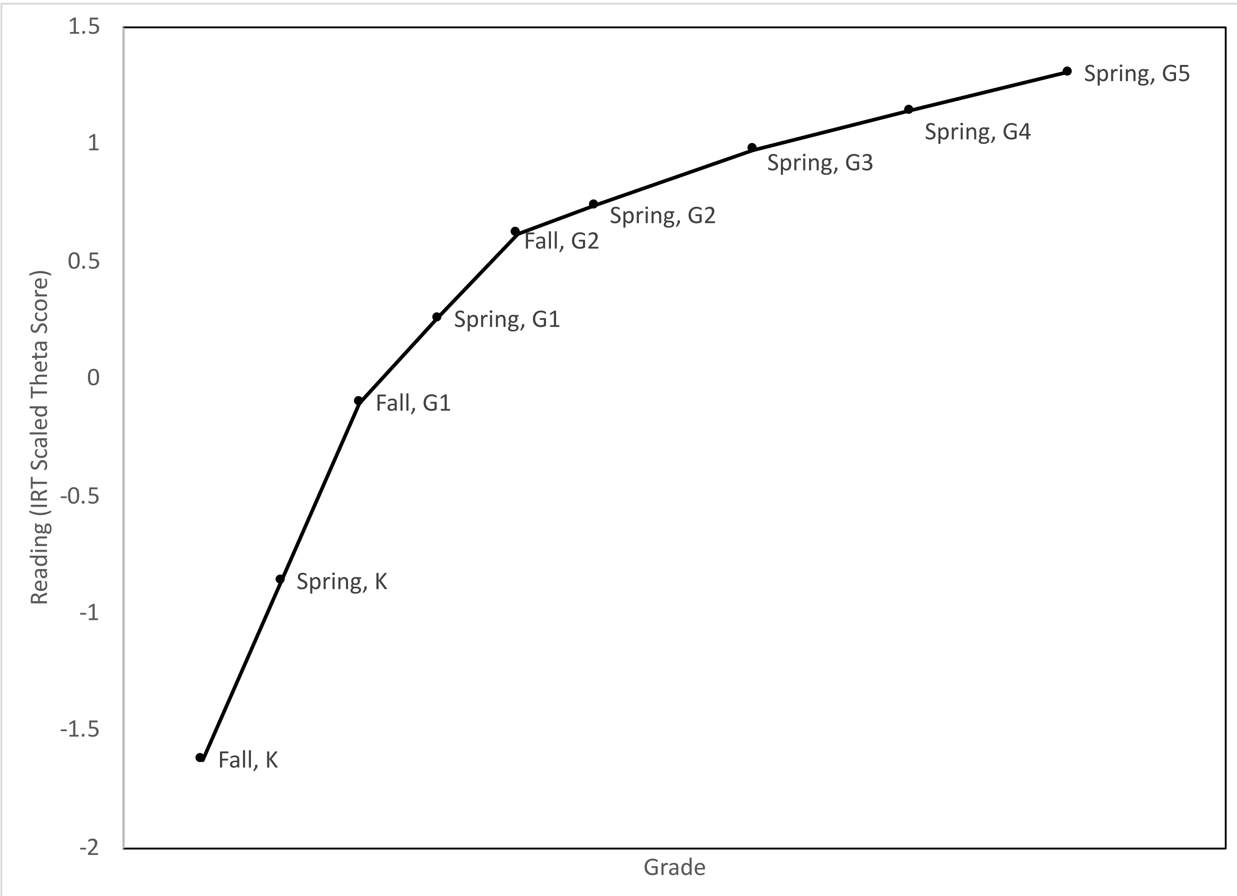
Standardized Path Coefficients and Standard Errors for Relations between Classroom EL Concentration and Reading Growth, with and without Controls (n = 783)

Outcome	Predictor	Without Controls	With Controls
Intercept (Fall, K status)	K Classroom EL Concentration	-0.082	-0.076
		(0.057)	(0.073)
Fall, K- Fall, G1 Slope	K Classroom EL Concentration	0.073	0.084
		(0.110)	(0.094)
Fall, G1 – Fall, G2 Slope	G1 Classroom EL Concentration	-0.023	-0.015
		(0.039)	(0.033)
Fall, G2 – Spring, G3 Slope	G2 Classroom EL Concentration	0.077	0.081
		(0.053)	(0.056)
Spring, G3 – Spring, G5 Slope	G4 Classroom EL Concentration	0.043	0.066
		(0.091)	(0.093)

Note. Models with controls include covariates for school EL concentration, child characteristics and school characteristics (see text above and SOM Table S4).

Figure 1

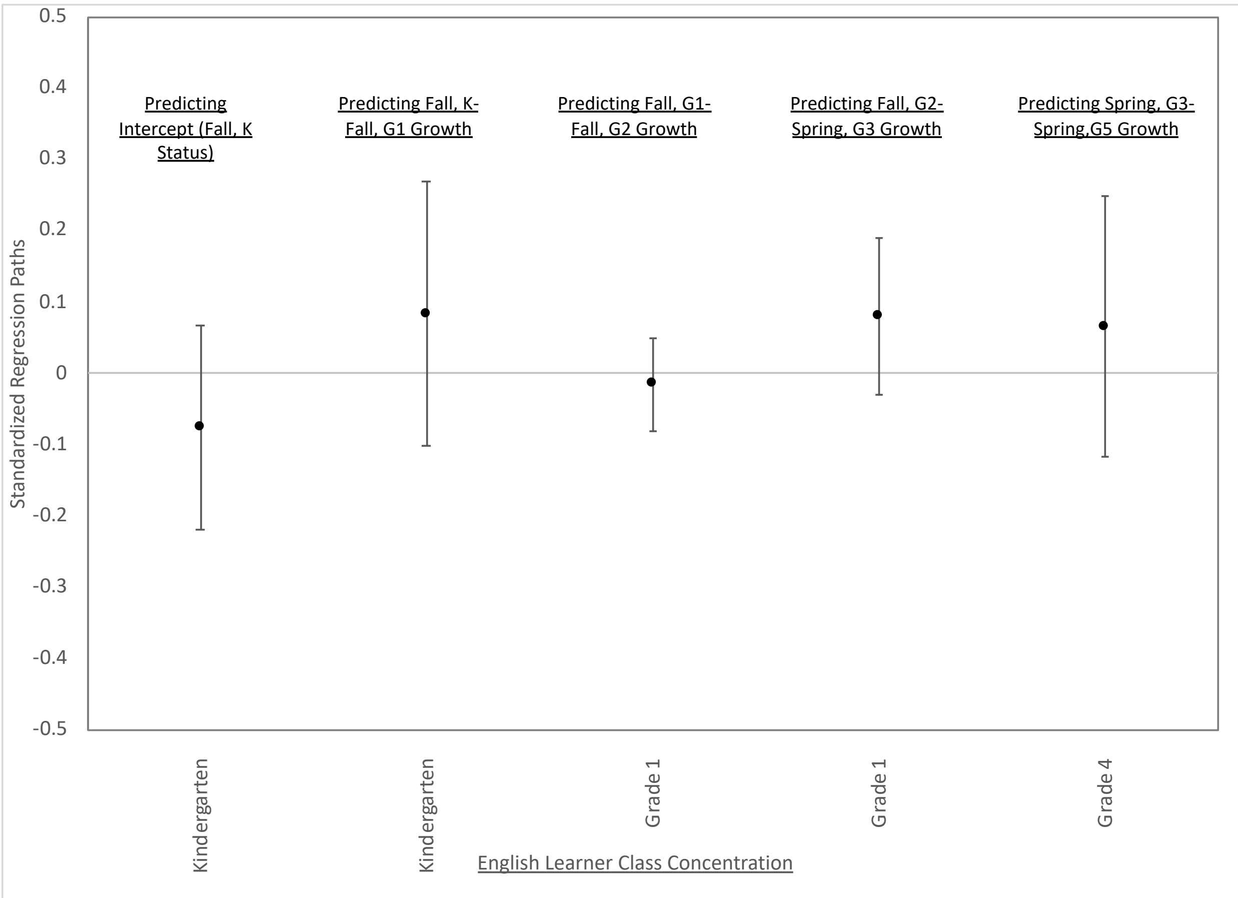
Fitted Trajectory for Reading Growth for All Ever-English Learners, Illustrating Four-slope Piecewise Growth Model (n = 783)



Note. K = Kindergarten; G = Grade.

Figure 2

Effect Size Estimates and Associated 95% Confidence Intervals for Relations between English Learner Classroom Concentration and Reading Growth (n = 783)



Note. G = Grade.

Classroom Concentration of English Learners and their Reading Growth

Supporting Online Material

Sample and Missing Data

The analytic sample includes 783 students who were ever-ELs (i.e., students whose kindergarten teacher reported that they "participate in an instructional program designed to teach English language skills to children with limited English proficiency") and who had non-missing longitudinal sampling weights for the seven main waves of data collection, including parent and child data (ECLS-K:2011 weight W9C29P_9B0). Given the latter criteria, the analytic sample is restricted to students with complete data for the seven main waves. We used the longitudinal sampling weights to account for attrition and preserve the nationally representativeness of ever-ELs of the sample. The sampling weights are appropriate for accounting for differential nonresponse patterns that can lead to bias in the estimates as well as differential probabilities of selection at each sampling stage, given the complex sampling design of ECLS-K:2011 (Tourangeau et al., 2019). In the fall of first grade and fall of second grade, only a stratified random subsample of children in one-third of the sample of primary sampling units participated. Given that this subsample was randomly selected within stratum, this missingness can be considered Missing at Random (MAR), and we took it into account using Full Information Maximum Likelihood (FIML). Monte Carlo studies (e.g., Enders & Bandalos, 2001) suggest that FIML is appropriate under the MAR assumption and superior to other approaches such as listwise and pairwise deletion. FIML was also used to account for missing data on EL concentration and covariates due to reasons other than attrition. Sample demographics are provided in Table S1.

Table S1

Sample Demographics (N = 783)

Race-ethnicity	Percentage
Latine	82.1%
Asian	9.9%
White	4.2%
Black	3.4%
Native American	0.3%
Multiracial-multiethnic	0.1%
Socioeconomic Status	Mean (SD)
Kindergarten Socioeconomic Status	-0.73 (0.57)
Grade 1 Socioeconomic Status	-0.76 (0.53)

Note: All racial-ethnic categories other than Latine are non-Latine. Socioeconomic Status is the ECLS-K composite of parental education, parental occupational prestige score, household income, with a full study sample mean of 0 and standard deviation of 1.

Reading Achievement Measure

The ECLS-K:2011 reading test can be considered a test of overall reading achievement that is appropriate for measuring growth over time, while also aiming to tap the most relevant skills in each time period. Drawing on the 2011 National Assessment of Educational Progress Reading Frameworks, current curriculum standards from five states, and the Common Core State Standards, the reading test includes items measuring knowledge and skills in three broad categories: basic reading skills, vocabulary, and reading comprehension (Najarian et al., 2019).

The weight of these three categories shifts across the time points: Basic reading skills constituted 50% of items in kindergarten, 40% and 20% in Grade 1 and 2, respectively, and were not assessed in Grades 3-5; reading comprehension items shifted from 35% of items in Kindergarten to 85% of items in Grade 5; and vocabulary items remained constant at between 15% and 20% of items across grades (Najarian et al., 2019). Item response theory methods were used to vertically scale the scores and account for the differences in forms across time.

Describing EL Concentration

As a preliminary analysis, we described the EL concentration experienced by the average ever-EL and the average never-EL. As shown in Table S2, the average ever-EL attends a school that enrolls 36%-40% ELs and is enrolled in a classroom that enrolls 41%-60% ELs. There is a modest trend toward lower concentrations in later grades, which is consistent with the idea that some ELs are becoming reclassified as former ELs (and more so than are new ELs arriving in the schools and classrooms). By contrast, the average never-EL attends a school with much lower school EL concentrations (9%-11% ELs) and classroom EL concentrations (7%-15% ELs). In addition, there is evidence of wide variation in school and classroom concentration, as indicated by the large standard deviations in Table S2. These are particularly large for ever-ELs' classroom concentration.

Table S2

School and Classroom English Learner Concentration, as Experienced by Ever-English Learners and Never-English Learners

	School Concentration Mean (*SD*)	Classroom Concentration Mean (*SD*)
	Ever-English learners (*n* = 783)	
Kindergarten	40.4% (25.8%)	56.0% (32.7%)
Grade 1	38.4% (22.0%)	59.8% (37.1%)
Grade 2	35.9% (22.4%)	53.1% (38.0%)
Grade 3	35.5% (22.0%)	53.7% (39.1%)
Grade 4	37.8% (21.9%)	48.3% (40.9%)
Grade 5	35.8% (22.1%)	41.4% (40.6%)
	Never-English Learners (*n* = 5930)	
Kindergarten	10.9% (18.3%)	8.1% (15.2%)
Grade 1	9.9% (15.1%)	8.3% (15.3%)
Grade 2	9.3% (14.2%)	7.4% (14.5%)
Grade 3	10.0% (15.2%)	14.2% (27.7%)
Grade 4	10.4% (14.7%)	14.9% (29.5%)
Grade 5	9.5% (22.1%)	14.9% (30.7%)

Preliminary Growth Analyses

Prior to addressing our research questions, we evaluated multiple specifications for growth in reading. In particular, we evaluated piecewise growth models that would allow us to segment growth into multiple meaningful developmental periods. This, in turn, allowed us to investigate the prediction of reading growth from EL concentration in more proximal periods (i.e., the same or prior grade) than would be possible with linear or polynomial growth models. As shown in Table S3, the four-slope piecewise model provided the best fit to the data. Figure 1 in the main text displays the unconditional version of this four-slope piecewise model for all ELs; as shown, this model includes slopes from Fall, Kindergarten to Fall, Grade 1; from Fall, Grade 1 to Fall, Grade 2; from Fall, Grade 2 to Spring, Grade 3; and from Spring, Grade 3 to Spring, Grade 5. This four-slope piecewise model fit better than a three-slope piecewise model (Satorra-Bentler scaled chi-square difference test = 57.20; $p < .0001$), a two-slope piecewise model (Satorra-Bentler scaled chi-square difference test = 310.67; $p < .0001$), and a linear model (Satorra-Bentler scaled chi-square difference test = 592.94; $p < .0001$). Quadratic and cubic specifications were also investigated, but encountered convergence problems; given our interest in segmenting growth into multiple periods, we did not pursue these specifications further.

It is worth noting that each slope covers both school time and summer time. This is an inevitable consequence of using the piecewise model (or any of the alternative specifications above), which require at least three time periods to establish each slope (with appropriate error terms; Singer & Willett, 2003). As a result, the slopes combine academic year growth with whatever growth or loss occurred during the summer periods. It is possible that this specification attenuated the effects of classroom EL concentration to some extent, which would not be expected to occur during the summer period. Future research with three waves of data per school year could explore this possibility.

Table S3

Comparing Goodness of Fit of Different Piecewise Specifications for Unconditional Growth Model

Specification	Slopes	Chi-square	AIC	BIC	RMSEA	CFI	TLI
Linear	Fall, K – Spring, G5	1723.231	7820.128	8263.125	0.232	0.158	0.242
Two-slope Piecewise	Fall,K – Spring, G2; Spring, G2 – Spring, G5	708.857	5091.972	5553.622	0.155	0.663	0.663
Three-slope Piecewise	Fall, K – Spring, G1; Spring, G1 – Spring, G2; Spring, G3 – Spring, G5	297.510	4242.371	4727.337	0.105	0.867	0.845
Four-slope Piecewise	Fall, K – Fall, G1; Fall, G1 – Fall, G2; Fall, G2 – Spring, G3; Spring, G3 –Spring, G5	169.696	3955.269	4468.214	0.086	0.928	0.896

Full Model

Table 1 in the main text provides the estimates for the relation between classroom EL concentration and reading growth for the models with and without controls. In Table S4, we provide all standardized estimates for those models. Note that not all variables are available at all rounds, particularly the fall rounds of testing, so earlier scores were used for controls in several cases (See superscripts and corresponding note at the bottom of the table). All variables were centered at their grand mean to facilitate convergence.

EL classroom concentration was specified as four time-invariant predictors. They are time-invariant in the technical sense that each predictor does not incorporate different values over time. However, across the four predictors, EL classroom concentration changes, so that its value is specific to the relevant period. With one exception, the predictors are each specified at the beginning of the particular time period for the relevant slope (see Table 1 in the main text) to predict future growth. For instance, Kindergarten EL concentration predicts the slope the runs from fall of Kindergarten to Fall of Grade 1, while Grade 1 EL concentration predicts the slope that runs from Fall of Grade 1 to Fall of Grade 2. One deviation from this pattern is that Grade 4 EL classroom concentration predicts the slope that starts in spring of Grade 3 and continues to spring of Grade 5; we made the decision that this would be the most appropriate time period for the predictor, because there alternative choice of spring of Grade 3 concentration might not reflect what students experience during fourth grade. Due to the nature of the piecewise growth model and how the degrees of freedom for modeling growth segments are constrained by the degrees of freedom of the number of testing occasions, we decided this was the best possible scenario.

All models included clustered standard errors to account for nesting of participating children within schools. It is worth noting that we could not account for nesting of children within classrooms, because classroom identifiers are not available in the public-use ECLS-K:2011 dataset. We recognized this limitation, but opted for the public-use dataset in the interest of ease of replicability. It is possible that failing to account for classroom-level variance may have biased our standard errors. However, this is unlikely to affect our conclusions, because accounting for classroom-level variance would have yielded wider standard errors, such that our null results would have remained null.

The final model presented includes socioeconomic status (ECLS-K:2011 composite), English proficiency, teacher-reported social-emotional skills (externalizing behaviors, internalizing behaviors, social skills), and executive functioning (cognitive flexibility, working memory, teacher-reported attentive behavior and inhibitory control) as child-level controls. These controls predicted reading achievement or growth in prior research using the ECLS-K: 1998 and ECLS-K: 2011 datasets (Morgan et al., 2008; Morgan et al., 2019; Tang & Dai, 2021). School controls including percent free-reduced lunch (which positively correlated with EL classroom concentration; r's =.30-.44) and percent students of color.

Table S4

Full Models with and without Controls, with Estimates Listed for All Controls (n = 783)

			Without Controls	With Controls
Fixed Effects (Standardized)	Initial (Fall, K) Status	Intercept	-2.747 (0.241)***	
		K Classroom EL Concentration	-0.082 (0.057)	-0.076 (0.073)
		K School EL Concentration		-0.011 (0.073)
		K Socioeconomic Status Composite		0.097 (0.059)
		K Externalizing Behaviors		0.133 (0.113)
		K Internalizing Behaviors		-0.031 (0.058)
		K Social Skills		0.039 (0.095)
		K English Proficiency Screener		-0.212 (0.068)**
		K Cognitive Flexibility		0.061 (0.053)
		K Working Memory		0.362 (0.060)***
		K Teacher-rated Attention		0.246 (0.089)**
		K Teacher-rated Inhibition		0.246 (0.089)**
		K School Percent on Free-reduced Lunch		0.003 (0.063)
		K School Percent Students of Color		0.079 (0.068)
	Fall, K- Fall, G1 Slope	Intercept	5.436 (2.025)**	
		K Classroom EL Concentration	0.073 (0.110)	0.084 (0.094)
		K School EL Concentration		0.019 (0.100)
		K Socio-economic status		0.081 (0.070)
		K Externalizing Behaviors		-0.113 (0.132)
		K Internalizing Behaviors		0.079 (0.080)
		K Social Skills		-0.093 (0.125)
		K English Proficiency Screener		0.229 (0.098)*
		K Cognitive Flexibility		-0.057 (0.070)
		K Working Memory		-0.338 (0.102)***
		K Teacher-rated Attention		-0.225 (0.117)

	K Teacher-rated Inhibition		0.173 (0.121)
	K School Percent on Free-reduced Lunch		-0.112 (0.083)
	K School Percent Students of Color		-0.121 (0.088)
Fall, G1 – Fall, G2 Slope	Intercept	2.415 (0.265)***	
	G1 Classroom EL Concentration	-0.023 (0.039)	-0.015 (0.033)
	G1 School EL Concentration		-0.066 (0.050)
	K Socio-economic Status		-0.143 (0.054)**
	Spring, K[a] Externalizing Behaviors		0.014 (0.044)
	Spring, K[a] Internalizing Behaviors		-0.097 (0.032)**
	Spring, K[a] Social Skills		-0.008 (0.052)
	Spring, K[a] English Proficiency Screener		0.051 (0.034)
	Spring, K[a] Cognitive Flexibility		0.076 (0.025)**
	Spring, K[a] Working Memory		0.028 (0.031)
	Spring, K[a] Teacher-rated Attention		0.109 (0.041)**
	Spring, K[a] Teacher-rated Inhibition		-0.070 (0.045)
	G1 School Percent Free-reduced Lunch		0.043 (0.042)
	G1 School Percent Students of Color		0.089 (0.055)
Fall, G2 – Spring, G3 Slope	Intercept	2.153 (0.242)***	
	G2 Classroom EL Concentration	0.077 (0.053)	0.081 (0.056)
	G2 School EL Concentration		0.016 (0.058)
	Spring, G1[b] Socioeconomic Status		-0.099 (0.058)
	Spring, G1[b] Externalizing Behaviors		-0.027 (0.069)
	Spring, G1[b] Internalizing Behaviors		0.033 (0.054)
	Spring, G1[b] Social Skills		-0.047 (0.074)
	Spring, G1[b] English Proficiency Screener		-0.066 (0.041)
	Spring, G1[b] Cognitive Flexibility		0.021 (0.037)
	Spring, G1[b] Working Memory		0.052 (0.039)
	Spring, G1[b] Teacher-rated Attention		0.185 (0.060)**

		Spring, G1[b] Teacher-rated Inhibition		-0.124 (0.085)
		G2 School Percent Free-reduced Lunch		0.041 (0.057)
		G2 School Percent Students of Color		-0.092 (0.068)
	Spring, G3 – Spring, G5 Slope	Intercept	3.589 (0.936)***	
		G4 Classroom EL Concentration	0.043 (0.091)	0.066 (0.093)
		G4 School EL Concentration		0.088 (0.098)
		Spring, G1[c] Socioeconomic Status		0.154 (0.078)
		Spring, G3 Externalizing Behaviors		-0.065 (0.107)
		Spring, G3 Internalizing Behaviors		0.131 (0.080)
		Spring, G3 Social Skills		0.112 (0.131)
		Spring, G3 Cognitive Flexibility[d]		0.011 (0.060)
		Spring, G3 Working Memory		-0.068 (0.082)
		Spring, G3 Teacher-rated Attention		0.333 (0.148)*
		Spring, G3 Teacher-rated Inhibition		-0.237 (0.158)
		G4 School Percent Free-reduced Lunch		0.094 (0.118)
		G4 School Percent Students of Color		-0.213 (0.102)*
Variance Components (unstandardized)	Initial Status		0.345 (0.054)***	0.193 (0.046)***
	Fall, K- Fall, G1 Slope		0.078 (0.058)	0.106 (0.067)
	Fall, G1 – Fall, G2 Slope		0.089 (0.023)***	0.093 (0.025)***
	Fall, G2 – Spring, G3 Slope		0.012 (0.003)***	0.013 (0.003)***
	Spring, G3 – Spring, G5 Slope		0.002 (0.001)	0.002 (0.001)*
	K-G1 with Intercept		-0.015 (0.046)	0.012 (0.041)
	G1-2 with Intercept		-0.150 (0.018)***	-0.108 (0.017)***
	G1-2 with K-G1		-0.016 (0.025)	-0.049 (0.033)
	G2-3 with Intercept		-0.040 (0.008)***	-0.027 (0.007***)
	G2-3 with K-G1		-0.021 (0.007)**	-0.032 (0.008)***
	G2-3 with G1-2		0.026 (0.005)***	0.029 (0.005)***
	G4-5 with Intercept		0.004 (0.003)	0.002 (0.003)

G4-5 with K-G1	0.000 (0.002)	-0.002 (0.002)
G4-5 with G1-2	-0.002 (0.002)	-0.001 (0.002)
G4-5 with G2-3	0.000 (0.001)	0.000 (0.001)

Note: [a]Controls that were not available in Fall, Grade 1, so Spring, K were used instead. [b]Controls that were not available in Fall, Grade 2, so Spring, Grade 1 were used instead. [c]Control that was not available after Spring, Grade 1 until Spring, Grade 5, so Spring, Grade 1 was used. [d]Measurement approach for cognitive flexibility for Grade 3 is different from those used in Kindergarten and Grade 1, so these estimates are not strictly comparable across time.

Robustness Checks

We conducted four robustness checks to determine if our results are sensitive to different specifications. First, we accounted for race/ethnicity by conducting a subgroup analysis with the Latine participants. Note: Latine is used here as a gender-inclusive alternative to Latino; it is also considered preferable to Latinx by some Spanish-speaking activists, because it is easily pronounceable in Spanish and more consistent with an existing Spanish suffix (Ochoa, 2022). Second, we tested whether results differed when using a specification for growth where slopes were regressed on the intercept. Third, because the longitudinal sample was defined by ever-EL status (thus including both current and former ELs in later grades), we explored interactions between EL concentration and contemporary current EL status in the grade in which concentration was measured. Fourth, we explored whether EL concentration interacted with the class language program (English-as-a-second-language, bilingual, or no services), because one reason for grouping ELs together is to provide native language instruction. Across all four of these robustness checks, the null effects of classroom EL concentration held, as described below.

Race/ethnicity

We initially attempted to control for race/ethnicity using a set of dummy variables, but this led to convergence problems. Instead, we conducted a subgroup analysis for the Latine participants, the largest racial-ethnic group of ELs in our data and nationally. As shown in Table S5, none of the path coefficients for classroom EL concentration were statistically significant, and their magnitudes were very similar to those in the final controlled model. Although we attempted parallel analyses for Asian students, the small sample size (*n* = 121) relative to the complexity of the model led to convergence problems.

Table S5
Model for Subsample of Latine Ever-English Learners, Compared to Final Controlled Model (n = 783)

		Final Controlled Model (*n* = 783)	Latine only (*n* = 613)
Initial (Fall, K) Status	K Classroom EL Concentration	-0.076 (0.073)	-0.131 (0.135)
Fall, K- Fall, G1 Slope	K Classroom EL Concentration	0.084 (0.094)	0.095 (0.131)
Fall, G1 – Fall, G2 Slope	G1 Classroom EL Concentration	-0.015 (0.033)	0.016 (0.029)
Fall, G2 – Spring, G3 Slope	G2 Classroom EL Concentration	0.081 (0.056)	0.013 (0.018)
Spring, G3 – Spring, G5 Slope	G4 Classroom EL Concentration	0.066 (0.093)	0.012 (0.013)

Note. Models include controls for school EL concentration, child characteristics and school characteristics (see Table S4).

Adding Regression Paths from Slopes to Intercept

Another possibility is that the growth model does not account appropriately for the relation between the slopes and the intercept. While these are allowed to covary in our final model, it is also possible to include one-way regression paths from each slope to the intercept, which could potentially control for these relations better. As shown in Table S6, this change did not lead to any statistically significant effects for classroom EL concentration, and the magnitudes were similar between this alternative specification and the final model.

Table S6

Model Incorporating a Regression Path for each Slope to the Intercept, in Place of Covariances, Compared to Final Controlled Model (n = 783)

		Final Controlled Model	Adding Paths from Slopes to Intercept
Initial (Fall, K) Status	K Classroom EL Concentration	-0.076 (0.073)	-0.071 (0.049)
Fall, K- Fall, G1 Slope	K Classroom EL Concentration	0.084 (0.094)	0.021 (0.050)
Fall, G1 – Fall, G2 Slope	G1 Classroom EL Concentration	-0.015 (0.033)	-0.031 (0.032)
Fall, G2 – Spring, G3 Slope	G2 Classroom EL Concentration	0.081 (0.056)	0.068 (0.055)
Spring, G3 – Spring, G5 Slope	G4 Classroom EL Concentration	0.066 (0.093)	0.047 (0.094)

Note. Models include controls for school EL concentration, child characteristics and school characteristics (see text above and SOM Table S4).

Interactions with Current EL Status

Next, because our sample consists of ever-EL students and EL status is a time-varying condition, we explored whether the results would differ for the subset of students who are current ELs in the particular grade in which EL concentration was measured. We did so using interaction terms between classroom EL concentration and contemporary current EL status in grades 1, 2, and 4 for the relevant slopes. As shown in Table S7, none of these interactions were statistically significant or practically meaningful in magnitude.

Table S7
Model with Interactions between Classroom EL Concentration and Contemporary Current EL status (n = 783)

		Model with Interactions with Current EL status
Initial (Fall, K) Status	K Classroom EL Concentration	-0.141 (0.121)
Fall, K- Fall, G1 Slope	K Classroom EL Concentration	0.110 (0.117)
Fall, G1 – Fall, G2 Slope	G1 Classroom EL Concentration	0.017 (0.052)
	G1 Current EL	-0.020 (0.022)
	G1 Classroom EL Concentration X Current EL	-0.027 (0.059)
Fall, G2 – Spring, G3 Slope	G2 Classroom EL Concentration	0.044 (0.034)
	G2 Current EL	-0.018 (0.012)
	G2 Classroom EL Concentration X Current EL	-0.014 (0.035)
Spring, G3 – Spring, G5 Slope	G4 Classroom EL Concentration	0.011 (0.021)
	G4 Current EL	-0.003 (0.008)
	G4 Classroom EL Concentration x Current EL	-0.002 (0.025)

Note. Models include controls for school EL concentration, child characteristics and school characteristics (see text above and SOM Table S4).

Interactions with Class Language Program

Because one rationale for grouping ELs together is to provide native language instruction, we hypothesized that the effects of EL concentration may be more positive when students are placed in bilingual classrooms. We investigated this by testing interactions between class language program and classroom EL concentration. Class language program was operationalized in three categories: ESL services, Bilingual programming (combining dual-language and transitional bilingual programming), and no language services. ESL services were the most common across most grades (K: 48%, G1: 46%, G2: 51%, G4: 34%), followed by bilingual programs (K: 31%, G1: 29%, G2: 23%, G4: 19%), and no language services (K: 21%, G1: 25%, G2: 26%, G4: 47%). As the largest category, ESL services served as the reference category while dummy variables for bilingual programs and no services were included and interacted with classroom EL concentration. As shown in Table S8, none of the interactions were statistically significant or practically meaningful. Two main effects were significant, but we hesitate to interpret these in the absence of significant interactions.

Table S8

Model with Interactions between Classroom EL Concentration and Class Language Program, with ESL Services as the Reference Category (n = 783)

		Model with Interactions with Class Language of Instruction
Intercept		
	K Classroom EL Concentration	-0.199 (0.101)
	K Bilingual	-0.019 (0.069)
	K No Services	-0.186 (0.079)*
	K Classroom EL Concentration X Bilingual	0.107 (0.088)
	K Classroom EL Concentration X No Services	-0.060 (0.086)
K1 Slope		
	K Classroom EL Concentration	0.314 (0.158)*
	K Bilingual	0.042 (0.094)
	K No Services	0.269 (0.112)*
	K Classroom EL Concentration X Bilingual	-0.191 (0.117)
	K Classroom EL Concentration X No Services	0.050 (0.112)
G12 Slope		
	G1 Classroom EL Concentration	-0.083 (0.052)
	G1 Bilingual	0.034 (0.054)
	G1 No Services	0.008 (0.036)
	G1 Classroom EL Concentration X Bilingual	0.015 (0.065)
	G1 Classroom EL Concentration X No Services	0.070 (0.044)
G23 Slope		
	G2 Classroom EL Concentration	0.035 (0.074)
	G2 Bilingual	-0.012 (0.055)
	G2 No Services	0.078 (0.051)
	G2 Classroom EL Concentration X Bilingual	0.073 (0.061)
	G2 Classroom EL Concentration X No Services	0.048 (0.060)

G35 Slope		
	G4 Classroom EL Concentration	-0.243 (0.192)
	G4 Bilingual	0.183 (0.116)
	G4 No Services	0.140 (0.105)
	G4 Classroom EL Concentration X Bilingual	0.227 (0.140)
	G4 Classroom EL Concentration X No Services	0.183 (0.159)

Note. Models include controls for school EL concentration, child characteristics and school characteristics (see text above and SOM Table S4).

SOM References

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. https://doi.org/10.1207/ S15328007SEM0803_5

Ochoa, Melissa K. (2022). Stop using 'Latinx' if you really want to be inclusive. *The Conversation. https://theconversation.com/stop-using-latinx-if-you-really-want-to-be-inclusive-189358*

Morgan, P. L., Farkas, G., Hillemeier, M. M., Pun, W. H., & Maczuga, S. (2019). Kindergarten children's executive functions predict their second-grade academic achievement and behavior. *Child development, 90*, 1802-1816.

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities, 41*, 417-436.

Najarian, M., Tourangeau, K., Nord, C., Wallner-Allen, K., and 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. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubsearch.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Tang, X., & Dai, T. (2021). How do classroom behaviors predict longitudinal reading achievement? A conditional autoregressive latent growth analysis. *Early Childhood Research Quarterly, 54*, 239-251.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. and Najarian, M. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten–Fifth Grade Data File and Electronic Codebook, Public Version* (NCES 2019-051). U.S. Department of Education. Washington, DC: National Center for Education Statistics.