**Dual Language Education and Academic Growth**

Angela Johnson

American Institutes for Research

anjohnson@air.org

**Structured Abstract**

Background or Context: Dual language education aims to foster the development of bilingualism, biliteracy, sociocultural competence, and academic skills in all school subjects. Early correlational research suggest that participation in dual language education is associated with higher achievement. Recent studies leveraged more comprehensive sets of baseline characteristics and found that dual language students improved their achievement status more than other students. A major limitation to these studies is their lack of ability to model within-student growth. Thus, we lack evidence on the relation between dual language education and growth trajectories within and across grades.

Purpose: This paper reports academic achievement and growth in grades 2 to 8 for Hispanic participants and nonparticipants of a Spanish-English dual language program. It extends the literature by providing novel evidence on seasonal patterns of learning for multilingual students across the elementary and middle grades.

Research Design: Applying a piecewise multilevel model to rich assessment data on a large sample of Hispanic students in a district in the Midwest, I compare the math and English reading growth rates of participants to nonparticipants of a dual language program. By separately specifying growth terms for each school year and summer, I test whether any differences in growth rates between DL and non-DL students expand, stay the same, or diminish across grade levels.

Conclusion: Dual language participants started 2nd grade with lower achievement than nonparticipants. In math, dual language participants grew more than nonparticipants during each year in grades 2 to 5 but lost more learning during summers. In English reading, dual language participants grew more during some school years and lost less during summers. These findings suggest summer learning opportunities are crucial for addressing achievement disparities.

**Biography**

Angela Johnson is a senior researcher at the American Institutes for Research. Her research interests include education policy, multilingual learners, and students with disabilities.

**Key Words**

Dual language education, English Learners, Education Policy

**Acknowledgements**

The author is grateful for support from the anonymous school district and valuable feedback from Nate Jensen and Megan Kuhfeld.

**Dual Language Education and Academic Growth**

Dual language education—an inclusive category for programs that provide academic content instruction in both English and a partner language—aims to foster the development of bilingualism, biliteracy, sociocultural competence, and academic skills in all school subjects. In the past few decades, a variety of program models have been implemented throughout the US and continue to expand. Dual language programs are currently implemented across 42 states and the District of Columbia; further, the programs are gaining prominence in policy contexts new to adopting dual language education, such as Delaware (Dual Language Schools, 2020).

Early waves of correlational research suggest that participation in dual language education is associated with higher achievement (Collier & Thomas, 2017). Recent studies leveraged more comprehensive sets of baseline characteristics to compare the achievement of students in dual language to other programs (e.g., Steele et al., 2017; Valentino & Reardon, 2015) and found that dual language students improved their achievement status more than other students (i.e., relative positions among peers). A major limitation to these studies is their lack of ability to model within-student growth (i.e., how much the students actually learned). Thus, we lack evidence on the relation between dual language education and growth trajectories within and across grades.

This paper provides descriptive evidence to address this critical gap in the research. Applying a piecewise multilevel model to rich assessment data on a large sample of Hispanic1 students (N=14,577) in a district in the Midwest, I compare the math and English reading growth rates of participants to nonparticipants2 of a dual language program. My research questions are:

1. How do Hispanic dual language participants’ movements in achievement status across grades compare to nonparticipants’?
2. How does Hispanic dual language participants’ achievement growth within and across grades compare to nonparticipants’?

This study makes two important contributions. First, it leverages vertically-scaled measures to report the first estimates of within- and across-year growth in the dual language education literature. I observe over 150,000 test scores each in math and English reading, measured in fall, winter, and spring of each academic year from 2nd to 8th grade, at up to 20 time points per student. These unique data allow me to estimate growth rates in each grade, as well as loss rates in summers after each grade, disaggregating seasonal patterns of learning unfound in dual language education research. Second, the study context is novel to the body of large-scale research studies on dual language education. Most extant studies addressed dual language programs in California, North Carolina, Oregon, Texas, and Utah. I contribute new evidence from a policy context that was largely unexplored in previous work. Unlike the recent experimental and quasi-experimental papers, this study is descriptive and does not aim to estimate the causal impact of dual language education. Instead, it extends the literature by providing novel evidence on seasonal patterns of learning for multilingual students across the elementary and middle grades.

**Positionality Statement**

Researchers are in a position of power and privilege from which we can contribute to disrupting systemic oppression and reframing research for the benefit of all students. Positionality represents the researcher’s world view in relation to the social and political context of the research and affects every phase of the research process. I am a first-generation immigrant and a former English Learner (EL)3. As a student who received and then exited EL services in US public schools, I attended both fully-segregated and English-only instructional environments. My research is also informed and my perspective shaped by a decade of training and practice as an English as a Second Language instructor in higher education and an English proficiency test developer. I acknowledge that my views are influenced by privileges associated with my US citizenship and non-disabled status.

**Dual Language Program Features**

The central feature of dual language programs is the provision of literacy and content instruction in both English and a partner language such as Spanish for promoting bilingualism, biliteracy, grade-level academic achievement, and sociocultural competencies (Howard et al., 2018). Two-way dual language programs enroll a mix of native users of English and native users of the partner language, with the idea that the two groups of students would model languages for each other. One-way programs use both English and the partner language to teach native users of the partner language.

Students who are developing English proficiency, to whom federal education policy refer as ELs, are an important target population for both two-way and one-way programs. Though the current study’s population of interest includes both non-ELs and ELs, I highlight below the theoretical framework linking dual language education to EL student achievement, as ELs tend to constitute a considerable portion of students in most dual language programs. In dual language programs, ELs receive home language support in content learning for an extended duration. Transfer theory and underlying proficiency theory explain that students who are developing English proficiency acquire literacy and academic concepts more easily in their home language than in English, and when they learn new concepts, the knowledge is added to their repertoire of skills and transferred to English (August & Shanahan, 2006; Cummins, 1979; Goldenberg, 2008). More recently, Phillips Galloway et al. (2020) found that core academic language skills in the home language and in English are positively related, suggesting that bilingual learners are able to leverage knowledge of shared communicative functions across languages and have greater opportunity for developing insights into how languages work. Ultimately, students in dual language programs gain literacy and academic skills in both English and their home language. However, initial exposure to English in dual language programs is reduced (e.g., typically 10% to 50% of instruction in English) relative to English-only or English-dominant instruction; thus, we might expect EL students in dual language programs to outperform their EL peers in other instructional services in English reading and academic content not immediately but in the long run (Valentino & Reardon, 2015).

In addition, dual language programs can elevate the social status of the partner language and boost ELs’ academic expectations and success (Garcia, 2002; Hult & Hornberger, 2016; Ruiz, 1984). Dual language programs can reframe students’ status from “English Learner” to “native users of the partner language,” undoing negative effects of the EL label on students’ outcomes (Umansky, 2016). In the case of two-way dual language programs, the placement of emergent and native English users in the same classroom can reduce linguistic and academic segregation. As a result, EL students in dual language programs can be exposed to higher-quality academic discourse in English than they would in EL-only classrooms. As the population of EL students in the US continues to grow, dual language programs offer a potentially effective approach to addressing their academic needs.

However, dual language program design and implementation, and consequently program quality, vary greatly by context (Howard et al., 2018). A range of factors crucial to program quality and student achievement differ across programs, such as the proportion of instruction allocated to each language (e.g., 90:10 or 50:50), the academic subjects taught in each language, the language proficiencies of its teachers, and curriculum materials (Valdés, 2017). Many dual language programs phase out partner language instruction after elementary school, with only one or two courses taught in the partner language starting in middle school. As a result, students (EL or non-EL) may develop interpersonal communication proficiency in the partner language but have limited opportunity to develop academic partner language skills or apply partner language proficiency in the learning of academic subjects.

**Dual Language and Academic Achievement**

A mature line of descriptive studies documented positive correlations between participation in dual language programs and academic achievement (see Bialystok, 2018 and Howard et al., 2018 for reviews). Generally, research has found that longer program duration (at least six years of partner language instruction) is associated with better downstream achievement outcomes for ELs and non-ELs. Across a large body of work, Collier and Thomas used multivariate regression to compare the achievement of dual language students with students in other programs, such as English as a Second Language and short-term home language education. In a recent review, they summarized this work by stating that they found dual language programs completely closed the achievement gap between ELs and English-proficient students, with effect sizes twice as large as those of other programs (Collier & Thomas, 2017). In a series of studies focused on EL and non-EL students in schools in California, Lindholm-Leary and colleagues reported that Latinx students in dual language programs performed better than Latinx students in mainstream classes (e.g., Lindholm-Leary & Block, 2010; Lindholm-Leary & Hernández, 2011). However, many of the early studies either insufficiently addressed selection bias or used test scores that violated the assumption of vertical scaling.

Three recent studies leveraged rigorous experimental and quasi-experimental designs to provide robust estimates on the relation between dual language program participation and academic achievement. Steele and colleagues (2017) conducted a randomized experiment in Portland, Oregon, using a lottery to assign kindergarten students to a dual language immersion program. They found positive intent-to-treat effects on reading achievement in 5th (.13 standard deviations (SDs)) and 8th grade (.22 SDs) but no effect on math achievement for ELs and non-ELs. Bibler (2021) used school-choice lottery data from Charlotte Mecklenburg School District and found that attending a dual language school increased math and reading achievement by 0.04 and 0.05 SD per year, respectively, and that the promising results applied to ELs and non-ELs alike. Valentino and Reardon (2015) applied multilevel modeling to achievement data from a district in California to examine the rate of change in standardized achievement measures across grade levels. Controlling for an extensive list of covariates, including parent program preference, they compared the spring California Standards Test (CST) scores of ELs who enrolled in dual language immersion to students in three other language programs. The study found that in terms of SD units of reading achievement scores, ELs in dual language immersion started the lowest among all programs in 2nd grade but experienced the most gains between 2nd and 7th grade.

**The Need to Examine Growth and Seasonal Learning**

Taken together, extant research on dual language programs provided evidence on students’ achievement status. However, the studies faced key limitations in modeling academic *growth* over time. In most study contexts, the test was not vertically scaled; as a result, scale scores could not be compared across grade levels. Valentino and Reardon (2015) addressed this by comparing students’ movement in achievement status relative to grade-level peers. But this approach does not allow us to examine within-student growth. When a student moves up in rank, we assume but cannot be sure of the direction or amount of academic growth. In other words, the rank approach simply presents relative order but lacks the information to describe magnitude or meaning between any two points of interest. When we see a student move up the ranks, her new position does not confirm learning. Maybe her peers grew, and she grew more; or perhaps she stayed the same while her peers lost learning; she might have even lost some learning but moved up the ranks because her peers lost more. As such, analysis of relative status does not capture within-student growth patterns contributing to the achievement gaps.

In addition, the extant research relies on scores from annual standardized testing. As a result, we lack information about within-year growth between fall and spring and about summer learning growth or loss. For example, Alvaer (2019) used vertically-scaled measures of Spanish reading and found that within a kindergarten cohort of Latinx EL students, dual language students grew at similar rates as students in short-term bilingual programs between kindergarten and 3rd grade. Since this study used annual assessment data, it is impossible to identify within-year patterns that contributed to this finding. Variations in patterns of learning by grade level and season are well-documented in recent research (Kuhfeld, 2019; von Hippel & Hamrock, 2019). Learning gains during the school year tend to be followed by stagnation or loss during the summer; this pattern applies across student groups.

Inferences on the effectiveness of teachers and school programs appear to be sensitive to the inclusion of summers (Atteberry & Mangan, 2020), which means that estimating a single effect size for each year masks important differences in seasonal learning rates. In exploring the efficacy of dual language education, it is important to examine within-year growth to track seasonal learning. While we may not expect the effects of instructional programs to extend beyond the school year or hold schools responsible for summer loss, educators and policymakers need to be informed if the benefits of school-year programs are diminished or undone during the summer. Given the efforts and resources invested in program design and implementation, districts and schools strive to help students not only maximize learning gains during the school year but also maintain or augment their knowledge and skills while school is out of session. A breakdown of school-year and summer growth will help decisionmakers identify ways in which students’ long-term learning growth can be sustained.

If dual language instruction is more effective at closing achievement gaps than other programs during the school year, we should see dual language students making more learning gains than other students between fall and spring. Similarly, we need to interrogate what happens to student learning while they are out of school during summer break. If dual language programs are also more effective than other programs at providing students with extended learning opportunities in the summer, through family engagement or other mechanisms, then we would see dual language students making gains or losing less learning than other students. However, it is also possible that dual language programs are more effective during the year when students are in class but are no more (or less) effective than other programs at supporting students while they are out of school. In this case, we would want to identify any gaps in summer learning rates so that schools can revise program designs to help students maintain or augment their school-year learning in the summer. This is not possible by only examining spring-to-spring achievement status, as previous research has done.

Seasonal learning patterns for students who use another language at home—and participation in dual language programs—merit investigation because of the students’ unique home contexts. In a recent qualitative study examining Latinx students’ mathematics learning, Stoehr et al. (2022) found that teacher-parent collaborations, especially in a bilingual context, can be beneficial to student learning. Language barriers, on the other hand, hinder family involvement (Vera et al., 2012). When schools provide academic materials only in English for students to practice at home, Spanish-speaking families may not be able to help the students using those English materials. Dual language programs have the potential to increase family involvement, which creates more out-of-school learning opportunities. For example, if families read to students in Spanish during summer vacation, students would continue to add to their repertoire of literacy skills, preparing them to acquire more advanced concepts in both Spanish and English in the fall. In this case, students would grow or at least maintain their skills during summers. However, if families do not pursue reading and other achievement-enhancing activities on their own regardless of language, then students might flatline or decline over the summer, and having Spanish instruction in school would not make a difference.

McCormick et al. (2021) found that during the summer prior to kindergarten, dual language learners (defined as students who use a language other than English) showed larger drops in language skills compared to monolingual learners. In a paper looking at kindergarten to fourth grade students, Johnson (2022) showed that compared to monolingual English users, students who are developing English proficiency grew more during the school year but also lost more learning over the summers. To my best knowledge, there is no study examining summer learning trajectories during the K-12 grades for students in dual language programs. Research is needed, first to identify which pattern holds and then explore how dual language program features, such as teacher-family interactions and summer learning support, shape students’ seasonal learning.

**Current Study**

This study addresses these important gaps in the research by estimating the association between participation in a dual language program and students’ achievement growth within and across years. I leverage rich longitudinal data from a diverse district in the Midwest with a long history of serving a large Spanish-speaking student population. A district-wide curriculum change in 2011-12 resulted in the expansion of bilingual/dual language education from K-2 to K-12, which has since served a large group of Hispanic students and been the primary support program for Spanish-speaking ELs. Controlling for student- and school-level covariates, I compare achievement and growth for Hispanic students who participated in the long-term dual language program to Hispanic students who did not participate.

**Study Context**

 This study is conducted in a research-practice partnership with a large school district in the Midwest. The district enrolls about 40,000 students in total. Slightly more than half the students are Hispanic; another quarter are White; and the rest of the students are members of other racial/ethnic groups. Just under two-thirds of the students come from low-income families. The district has a long history of serving students from diverse linguistic backgrounds, with current ELs comprising over 30% of its total student population. Though the district serves students with about 100 home languages, about 80% of the ELs use Spanish at home.

**Spanish-English Dual Language Instruction**

Spanish-English bilingual education was first implemented in the district in the 1970s and expanded over the years. For decades, Spanish instruction was available to students in kindergarten through 2nd grade in most elementary schools. Prior to the introduction of the dual language program, Spanish-speaking ELs in the district received language service through this transitional bilingual program, English as a second language, and/or sheltered instruction. Starting in the 2011-12 school year, Spanish-English two-way dual language instruction was made available across the district, such that any Spanish-speaking EL student enrolling in 2nd grade or below in 2011-12, as well as all subsequent cohorts, would be able to enroll in dual language from elementary to high school. Students who are native and fluent users of English are also eligible to participate. The remainder of this paper will refer to this Spanish-English dual language program as DL.

In 2011-12, all Spanish-speaking EL students in the district became eligible to participate in a DL by default, though students’ families have the right to opt out of participation. Since DL started, it has been the primary support service for Spanish-speaking ELs in the district. Other support services, such as transitional (K-2) bilingual, English as a Second Language, and sheltered instruction, remain available to ELs in the district. Spanish-speaking ELs who opt out of DL can choose to receive or decline these alternative services. More than 85% Spanish-speaking ever-ELs who enrolled in the district in or after 2011-12 enrolled in DL during at least one school year; almost all students who enroll stay in the program. ELs can stay in DL after gaining fluent English proficiency and reclassifying; in other words, exiting EL status does not lead to changes or removal of services. Very few students (about 1%) who enroll in DL, ever-EL or never-EL, exit the program.

**Program Features**

DL offers instruction in Spanish and English between kindergarten and 12th grade. EL and non-EL participants receive the same instruction. Kindergarten instruction is delivered 80% in Spanish and 20% in English (80:20), followed by 70:30 and 60:40 in 1st and 2nd grade, respectively, and 50:50 between 3rd and 6th grade. In 7th and 8th grade (middle school), students take language arts in both Spanish and English, social studies in Spanish, and math and science in English. From 9th to 12th grade, students take Spanish language arts and at least one additional core subject course in Spanish every year. In addition, all 9th grade DL students are required to take the Advanced Placement Spanish Language exam.

**Data and Sample**

 The data for this study come from the district’s administrative records between school years 2009-10 and 2019-20. Variables include students’ (a) demographic characteristics (e.g., biological sex, race/ethnicity, home language); (b) eligibility to receive EL, free or reduced-price lunch (FRPL), and Special Education (SPED)4 services and participation; (c) course enrollment, which identifies DL participation; and (d) academic assessment scores.

**Assessment Scores**

This study focuses on NWEA’s MAP Growth assessment. MAP Growth is a computer adaptive test, which provides a precise measurement even for students above or below grade level.5 The assessments are aligned to the state’s content standards. States and districts choose to administer MAP Growth to their students for various purposes, including monitoring academic achievement and growth. The tests are untimed, though they typically take approximately 40 to 60 minutes for students to complete. States and districts decide when and how many times to administer MAP Growth, depending on their assessment need and context. Testing typically takes place three times per academic year—in the fall, winter, and spring. Up to and through winter 2020, students took the tests in school during regular school hours on electronic devices provided by the school, with school staff (often their classroom teacher) proctoring the test sessions and providing necessary support with navigation of the test interface. The assessment is vertically scaled to allow for the estimation of gains across time. Test scores are reported on the Rasch unIT (RIT) scale, where RIT is a linear transformation of the logit scale units of the Rasch item response theory model.

In this district, students took the MAP Growth assessments three times a year, from the fall of 2nd grade to the winter of 8th grade between 2004-05 and winter 2020. I use data for cohorts of students (defined by the year of entering 2nd grade) who were in 2nd grade after the start of DL, from 2011-12 to winter 2020. Eighth-grade spring testing is optional in the district, therefore test scores in that term are excluded.

**Student and School Covariates**

I observe a comprehensive set of student demographic characteristics and school-level student composition variables. I include key student- and school-level covariates in my analyses (described in the Analysis section) because prior research (e.g., Valentino & Reardon, 2015) has shown them to be strongly correlated with student achievement and used them as statistical controls. For the purpose of multilevel growth modeling, student- and school-level covariates are treated as constant across time. Annual support service eligibility flags are used to generate ever-eligible indicators at the student level. Each student thus has an ever-EL, an ever-FRPL, and an ever-SPED variable that has the same value across all time points in the data. Since very few DL students leave the program, ever-participating is essentially equivalent to consistently participating; therefore, I use ever-DL as a student-level indicator for participation. I also include an indicator for refusing EL service at any time before the fall of 2nd grade. All ELs are entitled to receiving support service, but students’ families can choose to opt out of service. Though individual families’ rationale for refusing EL service are unknown, students whose families refused service may be systematically different from students whose families consistently accepted EL service. For example, some students may have had high English proficiencies close to but not quite meeting the cut scores for the state’s English proficiency test; their families may have refused EL service believing that their child’s English proficiency and other skills are sufficient to succeed in a general education setting without language support. Using student-level data, I calculate the percentages of students enrolled at each school who are eligible for EL and FRPL services. Students are nested within the first school attended in the district, and school-level characteristics are treated as constant across time.6

**Sample**

I retain only Hispanic students (including ELs and non-ELs) in the analytic sample since the Spanish-English DL was designed primarily to serve Spanish-speaking students and the vast majority of participants have been students whose reported race/ethnicity is Hispanic. However, I do not observe Spanish proficiency or complete English proficiency data for the Hispanic students in my sample, therefore, I am unable to distinguish language proficiency beyond EL status. In the descriptive analyses outlined in the next section, I compare the academic growth of DL participants (students in cohorts after start of the program, or “post-program cohorts”, who enrolled) to nonparticipants (students in post-program cohorts who could have enrolled but did not). It is important to note that nonparticipants in post-DL cohorts shared time-specific schooling experiences with participants but may have differed in systematic ways. The unobserved characteristics that led some Hispanic students to enroll in DL and others not may also be driving differences in achievement growth. Though this paper does not aim to provide causal estimates, I conduct a series of sensitivity analyses in an effort to account for baseline differences. These include (a) adding cohort dummies to the model and (b) controlling for early literacy assessment scores before 2nd grade. The results are reported in the online supplemental materials.

I observe achievement scores at up to 20 time points for each student, for a total of 156,212 math test scores and 155,459 English reading test scores nested within 14,577 students within 50 schools. Table 1 shows the summary statistics for the full sample, students who enrolled in DL at least one year (“DL”), and students who never enrolled in DL (“non-DL”). About 76% of the students in the pooled sample were ever eligible for EL services, 92% ever eligible for FRPL services, and 16% ever eligible for SPED services. Students in DL had higher rates of ever-EL (94% vs 38%) and ever-FRPL (96% vs 83%) than non-DL students. Table 2 shows the number of students tested by cohort and grade. Since only four post-DL cohorts have 7th and 8th grade test scores, I interpret findings pertaining to middle school growth as suggestive rather than definitive evidence.

[Table 1 here]

[Table 2 here]

**Analysis**

I examine the math and reading trajectories of DL and non-DL students in two visual analyses and growth modeling. Prior literature primarily relied on z-scores to compare students’ spring achievement across grades (e.g., Valentino & Reardon, 2015). Following this line of research and to provide comparisons to previous findings, my first approach plots spring achievement z-scores, calculated using MAP Growth norms by subtracting the grade-level mean and dividing by the grade-level standard deviation, for DL and non-DL students between 2nd and 7th grade (Thum & Kuhfeld, 2020). Then, I leverage vertically-scaled test scores from the fall, winter, and spring to interrogate growth within and across grades. First, I plot the mean achievement by DL participation for each test term between the fall of 2nd grade and the winter of 8th grade. The plots illustrate changes in achievement score not only between but also within grades.

Finally, I use a piecewise multilevel growth model to estimate growth (e.g., von Hippel, Workman, & Downey, 2018). Exposure to school in each grade and each summer varied. Students were not tested on the first and last days of school each year; even within school, students’ test dates varied depending on the availability of electronic devices used for testing. An important advantage of the piecewise multilevel model is its ability to account for variation in test administration dates within the school year and allow for separate growth terms in each school year and summer (e.g., Quinn et al., 2016). By separately specifying growth terms for each school year, we can test whether any differences in growth rates between DL and non-DL students expand, stay the same, or diminish across grade levels.

The model accounts for variations in test dates and estimates students’ academic growth as a linear function of their “months of exposure” to each school year and summer.7 Months of exposure is calculated based on school start and end dates and the test administration dates (see Appendix A1 in the supplemental materials for details). For example, a student testing at the end of August in 3rd grade may have 9.7 months of exposure to 2nd grade, 2.3 months of exposure to the summer following 2nd grade, and one week of exposure to 3rd grade.

At level 1, I model achievement conditional on exposure to school during the academic year for each grade level (e.g., G2ij = Grade 2 academic year) and exposure to summer after each grade level (e.g., S2ij = summer after Grade 2). Appendix A1 in the supplemental materials details the calculation of each of the level-1 predictors (G2ij through G8ij).

Level 1 (time (*t*) within student (*i*) within school (*j*)):

|  |  |
| --- | --- |
| $$y\_{tij}=π\_{0ij}+ π\_{1ij}G2\_{ij}+ π\_{2ij}S2\_{ij}+ π\_{3ij}G3\_{ij}+ π\_{4ij}S3\_{ij}+ π\_{5ij}G4\_{ij}+ π\_{6ij}S4\_{ij}+ π\_{7ij}G5\_{ij}+ π\_{8ij}S5\_{ij}+ π\_{9ij}G6\_{ij}+ π\_{10ij}S6\_{ij}+ π\_{11ij}G7\_{ij}+ π\_{12ij}S7\_{ij}+ π\_{13ij}G8\_{ij}+e\_{tij}$$ |  |

As von Hippel and colleagues (2018) stated, this model “implicitly extrapolates beyond the test dates to the scores that would have been achieved on the first and last day of the school year” (p. 335). The intercept $(π\_{0ij})$ therefore is the predicted score for student *i* in school *j* testing on the first day of 2nd grade, even if the student tested in the third week of the school year. The slopes $(π\_{1ij},…,π\_{13ij})$ are the monthly learning rates of student *i* during each school year and summer. Each test score $y\_{tij} $is viewed as a linear function of the months that student *i* in school *j* has been exposed to 2nd grade ($G2\_{ij})$, 3rd grade ($G3\_{ij})$, etc., through 8th grade ($G8\_{ij})$; and the number of months that the student has been exposed to the summers after 2nd ($S2\_{ij}$) through 7th grade ($S7\_{ij}).$

At level 2, I include a random intercept to allow students’ starting achievement in fall of 2nd grade to vary by student; slopes are treated as fixed.8 I start with a baseline model that only includes the time predictors, the indictor for ever participating in the program, *DL*, and an interaction term between *DL* and each time predictor, which will yield estimates for additional growth DL students experienced above and beyond the growth rate of non-DL students (Model 1). After that, I run another model that additionally includes student-level covariates (Model 2). Equation (2) below shows my preferred specification. *EL*, *FRPL*, and *SPED* are student-level covariates representing ever being eligible for each service during the students’ entire time in the district; *female* is an indicator for the student’s biological sex reported as female; *refuse* is an indicator for ever refusing EL service before the fall of 2nd grade.

|  |  |
| --- | --- |
| Level 2 (student (*i*) within school (*j*)):$$π\_{0ij}=β\_{00j}+β\_{01j}DL\_{ij}+β\_{02j}EL\_{ij}+β\_{03j}FRPL\_{ij}+β\_{04j}SPED\_{ij}+β\_{05j}female\_{ij}+β\_{06j}refuse\_{ij}+r\_{0ij}$$$$π\_{1ij}=β\_{10j}+β\_{11j}DL\_{ij}$$$$\vdots $$$$π\_{13ij}=β\_{130j}+β\_{131j}DL\_{ij}$$ |  |

At level 3, a random intercept allows starting achievement in the fall of 2nd grade to vary by school; slopes are treated as fixed. I run the model without and with *%EL*, and *%FRPL*, which are grand-mean centered school-level covariates for the first school in which the student enrolled (Model 3). Equation (3) below shows the preferred specification that includes these covariates.

|  |  |
| --- | --- |
| Level 3 (school (*j*)): $$β\_{00j}=γ\_{000}+γ\_{001}\left(\%EL\_{j}\right)+γ\_{002}\left(\%FRPL\_{j}\right)+u\_{00j}$$$$β\_{10j}=γ\_{100}$$$$\vdots $$$$β\_{130j}=γ\_{1300}$$ |  |

*Variance component specification:*

$e\_{tij} \~ N\left(0, σ\_{tij}^{2}\right), r\_{ij}\~MVN\left(0, T\_{St}\right), u\_{j}\~MVN\left(0, T\_{Sch}\right)$.

 Models are estimated using full-information maximum likelihood in HLM 8.0 software (Raudenbush et al., 2019). All available test scores are used in estimation, regardless of whether a student had test scores in every test term. I apply the 3-level model to the full sample of Hispanic students.

Since there is a large difference between the percentages of ever-EL students in the DL and non-DL groups in the sample, ever-EL status may explain a large part of the difference between DL and non-DL students’ outcomes. To further address this potential confounding effect, I run another analysis after restricting the sample to ever-EL students (N=11,066) and compare the growth rates of DL and non-DL students. In this EL-only analysis, I drop the student-level predictor *EL* from the model.

**Sensitivity Checks**

In order to account for variation by cohort (the year students entered 2nd grade), I add cohort dummies to the level-2 equation and rerun the models. For a subsample (about 59% of the students), I also observe Fountas & Pinnell Literacy Gradient (2020) scores measured prior to the fall of 2nd grade. I include this pre-treatment literacy measure as an additional covariate in a sensitivity check.

**Findings**

*Research Question 1: How do Hispanic dual language participants’ movements in achievement status across grades compare to nonparticipants’?*

 Figure 1 shows Hispanic students’ achievement z-score movement between 2nd and 7th grade for math and English reading. In math, DL students scored in the fall of 2nd grade slightly more than 0.3 SD below the national average and non-DL students scored higher than DL students by about 0.1 SD. Spring math z-scores for DL students were fairly stable from 2nd to 6th grade but dipped considerably between 6th and 7th grade. Non-DL students’ spring math z-scores fluctuated between 2nd and 7th grade. By the end of 7th grade, DL students were about 0.45 SD below the national average, non-DL students about 0.2 SD below the national average, with a difference of about .25 SD between DL and non-DL students.

In English reading, the DL students scored about 0.65 SD below the national average in 2nd grade and 0.5 SD below the national average in 7th grade. Non-DL students scored about 0.15 SD below the national average in 2nd grade and about 0.1 SD below the national average in 7th grade. Between DL and non-DL students, the gap was about 0.5 SD in 2nd grade and about 0.4 SD in 7th grade. DL students appeared to make greater gains than non-DL students between 3rd and 4th grade and between 5th and 6th grade.

It is important to note that these movements in z-scores provide information about DL students’ academic positions relative to non-DL students and the national average but cannot speak to the progress students themselves are making within and across years.

[Figure 1 here]

*Research Question 2: How does Hispanic dual language participants’ achievement growth within and across grades compare to nonparticipants’?*

**Achievement Means**

 Figure 2 shows trajectories in math and English reading achievement between the fall of 2nd grade and winter of 8th grade. The horizontal axis shows the 20 test terms, with labels and national average markers for the winter test term of each year (Thum & Kuhfeld, 2020), which the district uses for school accountability.

In math, non-DL students scored at or slightly below the national average each winter, while DL students scored slightly below non-DL students. DL students started the fall of 2nd grade scoring slightly lower than non-DL students. Between the fall and spring of each grade between 2 and 5, DL students appeared to grow slightly more, shrinking the gap between themselves and non-DL peers; however, DL students seemed to lose more learning over the summers in these years (shown by the larger dips in the solid red line) than non-DL students, so that the gap is somewhat restored by fall. In grades 6 to 8, the achievement trends for DL and non-DL students are mostly parallel, suggesting similar academic-year growth and summer loss rates.

In English reading, both non-DL and DL students scored consistently below the national average in each winter in earlier grades; however, the gaps between both groups and the national average shrank as students moved to higher grades. Similar to math, DL students started the fall of 2nd grade scoring below non-DL students in reading. In each subsequent academic year and summer, the growth trajectories of DL and non-DL students look parallel, indicating that the two groups had similar academic-year growth and summer loss rates. It is also noteworthy that neither DL nor non-DL students seemed to lose much learning during each summer, as suggested by the flat segments in the graph.

[Figure 2 here]

**Monthly Growth Estimates**

Monthly growth rates for DL students, non-DL students, and the national average (calculated using MAP Growth norms from Thum & Kuhfeld, 2020) in each academic year and subsequent summer are presented in Figure 3. Estimates are from the preferred model specification which includes student and school covariates (Table 3, Columns 3 and 6). The black markers at the tip of each bar represents 95% confidence intervals for the estimates.

[Figure 3 here]

In math, between grades 2 and 5, DL students appeared to grow significantly more than non-DL students during each academic year; but they also seemed to have greater monthly loss during each summer. For instance, in the 2nd grade year, DL students grew about 2 RITs per month, while non-DL students grew about 1.8 RITs per month. Over the school year, 2nd grade DL students would close the gap between themselves and their non-DL peers by about 1.8 RITs. In the subsequent summer, DL students lost more per month, undoing the amount they caught up over the year. In 6th to 8th grade, there was no significant difference between the two groups in terms of growth or summer loss. Across all years, both DL and non-DL students in the district had similar or higher growth rates compared to the national average, and DL summer loss rates tended to be higher than the national average.

The results for English reading are quite different. DL students grew significantly more than non-DL students during their 4th, 5th, and 7th grade years and lost more than non-DL students during the summer after 2nd grade. For other school years and summers, growth rates did not differ significantly between DL and non-DL students. With the exception of 2nd and 8th grade, students in the district tended to grow slightly slower than the national average. Summer loss rates in the district were also significantly and substantively lower than the national average in grades 2 to 5.

Regression results corresponding to Figure 3 are shown in Table 3. Columns (1) (math) and (4) (English reading) show results from the baseline model that includes ever-DL participation as a predictor and interaction terms between DL and each time predictor (exposure to 2nd grade academic year through exposure to 8th grade academic year). The intercept represents student achievement on the first day of 2nd grade. The average achievement of non-DL students was 171.6 RITs for math and 167.6 RITs for reading. In comparison, DL students scored a significant 3.07 RITs and 8.38 RITs lower in math and reading, respectively. Columns (2) and (5) show results from adding biological sex and educational service eligibility. The intercepts (179.5 and 176.0) represent starting RITs for male, non-DL students who were never eligible to receive EL, SPED, or subsidized lunch services. Being female is associated with significantly lower starting achievement in math but higher in reading. Being eligible to receive EL, SPED, and subsidized lunch services is each associated with significantly lower starting achievement in both subjects. Holding all other factors constant, having refused EL service in kindergarten or 1st grade is associated with higher achievement in both subjects. This would make sense if families were opting out of EL service with the belief that their child would succeed even without language support. When school-level percentages of EL and FRPL enrollment are added to the model (columns 3 and 6), percentage of FRPL eligibility is associated with lower starting achievement in math and English reading (though the estimate for reading is not significant), and the coefficients for the percentage of EL eligibility are not significant. Growth estimates in all academic years and summers are stable in both magnitude and statistical significance across the three models for both math and reading. For instance, as reported in column (1), non-DL students grew 1.79 RITs in math during 2nd grade, and DL students grew an additional 0.18 RITs, which is statistically significant. Estimates in columns (2) and (3) are very similar. Results from the preferred model specification are in columns (3) and (6) and correspond to those reported in Figure 3.

[Table 3 here]

**EL Growth Estimates**

Table 4 shows the findings from multilevel growth models applied to only Hispanic ever-ELs. Again, estimates are stable across models in both magnitude and significance. I interpret results from columns (3) and (6), which are for the preferred specification that includes student and school covariates. EL growth estimates for both math and English reading are very similar to the full sample. In math, ELs in DL appeared to grow significantly more than ELs not in DL during each academic year in grades 2 to 5, but ELs in DL also seemed to have lost more math learning during summers after grades 2 to 5 compared to non-DL ELs. There are no significant differences between the two groups in 7th and 8th grade. In English reading, DL and non-DL ELs grew at similar rates during the school years and lost similar amounts during the summers. Like in the full sample, school-year growth rates in 4th, 5th, and 7th grade for DL EL students were higher than non-DL EL students, but the estimates are not significant, perhaps due to low statistical power in this smaller sample.

[Table 4 here]

**Sensitivity Checks**

The series of models I estimated, including and excluding student- and school-level covariates as controls, yielded very similar results. Appendix A2 and A3 report findings from two additional sensitivity checks. The addition of cohort dummies to the full sample resulted in similar estimates (Table A2). For a subsample (about 59% of the full sample) for whom I observe Fountas and Pinnell literacy assessment data, I repeated the analysis including baseline literacy score measured prior to the fall of 2nd grade; estimates are similar to the full sample (Table A3). These findings offer some reassurance that any differences observed between DL participants and nonparticipants are not solely due to pre-existing differences in student or school demographics, cohort, or baseline literacy.

**Discussion**

 This study leverages rich longitudinal data to estimate growth within and across grade levels for Hispanic participants and nonparticipants of a dual language program. Achievement z-score movements between grades suggest that in English reading, both DL and non-DL students moved closer to the national average; in math, however, non-DL students maintained the gap between themselves and the national average while the gap between DL students and the national average widened, driven by a drop between 6th and 7th grade.

 At first glance, my findings on students’ achievement relative grade-level peers seem to contradict some of the results from earlier studies. For instance, Valentino and Reardon (2015) found that longer-term dual language ELs outperformed ELs in English-only and short-term bilingual programs in 5th grade and reached the state average in 6th grade in ELA achievement. Latinx ELs in their study who were in longer-term dual language programs made ELA gains at twice the rate of Latinx ELs in short-term bilingual programs. Similarly, Steele et al. (2017) found that dual language immersion students, both EL and non-EL, outperformed their peers in reading at the end of 5th grade. Bibler (2021) also reported a small positive effect on reading for native English users and suggestive evidence that dual language education may improve reading outcomes for ELs (most of whom were Hispanic in his sample). In my data, even at the end of 7th grade, Hispanic DL students did not completely close the reading achievement gap between themselves and Hispanic non-DL students in the same district; nor did Hispanic DL students in my district attain the average score of all students across the nation (Figure 1). In math, Valentino and Reardon (2015) and Steele et al. (2017) found that dual language students did better or at least no worse than students in other programs in terms of achievement relative to their state average across grades. In my data, the state average was not available for comparison; however, I found that DL students did not get closer to the national average for their grade (Figure 1).

One potential explanation for the differences between my findings and previous research is study context. Valentino and Reardon (2015) used an ethnically and linguistically diverse EL sample from a district in California, about 33% whom were Latinx. Steele et al.’s (2017) study took place in Portland and included ELs and non-ELs in dual language programs that offered several partner languages. Bibler (2021) looked at school choice lotteries from Charlotte-Mecklenburg School District, and ELs comprised a minority of his sample. In contrast, my partner district only offers Spanish-English dual language education; the vast majority of program participants are Hispanic, 94% of whom were ever classified as ELs.

A closer look at seasonal learning provides additional potential explanations for how the achievement patterns in my data developed. DL and non-DL students exhibited different growth patterns during academic years and summers breaks, at least between the start of 2nd and start of 6th grade (Figure 3). In math, both DL and non-DL students grew faster compared to the national average, and DL students appeared to grow faster than non-DL students. In the summers, however, non-DL students lost math learning at a rate similar to the national average, while DL students seemed to have lost math learning at a significantly higher rate. The difference in math summer loss rates could offer a hypothesis for explaining why DL students were not approaching the national average over time. Unfortunately, previous literature did not examine summer learning and therefore did not provide a point of comparison for my findings. The summer after 6th grade stands out as a period in which the loss rate difference between DL students and the national average seemed especially large (Figure 3); this may have contributed to the large dip in z-scores (Figure 1). In English reading, both DL and non-DL students grew slightly less than the national average during the academic years; however, both groups (DL students in particular) experienced significantly less summer English reading loss than the national average, which would explain their approaching the national achievement average across grades. Turning to ever-EL students in my sample (Table 4), the magnitudes of school-year growth estimates are similar to the findings for 2nd, 3rd, and 4th grade from Johnson (2022), as are math summer loss rates between those grades. However, Johnson (2022) reported significant summer reading loss for ever-ELs and other EL subgroups, while the ever-ELs in my sample experienced very little setback. As I discuss later in this section, this may be attributed to my partner districts’ supports for summer reading.

Focusing only on academic years, my descriptive findings corroborate those of previous experimental and descriptive studies; that is, students in dual language education grow at least as much as, if not more than, their peers in other programs while school is in session. Dual language programs and educators deserve credit for the amount of learning they foster, which is extraordinary, especially considering the many hardships faced by historically-marginalized students, their families, and their communities. It is when services are removed, while students are away from their dual language program in the summer, that the benefits of the program are reduced or undone.

 The analysis of seasonal learning patterns in this study contributes significantly to the extant literature on dual language education. Prior research tended to find participation in dual language instruction to have positive associations with (or impact on) downstream achievement relative to their state average or improvements in z-scores across grades, but none was able to decompose these general patterns. This study shows that academic-year growth and summer maintenance are two key issues that merit consideration by my partner district, and other districts may wish to address summer learning maintenance as well. Districts and schools should be commended for what they accomplish during the school years. But learning and achievement are often evaluated in a cumulative manner. When students leave the K-12 education system, colleges and employers evaluate the knowledge and skills accumulated up to that time, with no distinction for school-year and summer progress. Separately identifying school-year and summer growth rates is vital for informing policy and practice because periods of loss can set students back and ultimately hinder their postsecondary opportunities. In designing programs to support vulnerable and marginalized student populations, policymakers and educators should consider ways to maximize learning not only within academic years. Summers can and should be leveraged to reinforce learning gains.

In my research partner district, Hispanic DL students experienced little summer loss in English reading in each of the early grades. The districts might reflect on its successful practices that led to minimal reading summer loss and share with other programs. In math, however, DL students, who appeared to grow more during the school year, seemed to experience significantly larger summer loss compared to Hispanic non-DL students across 2nd to 5th grade. This is consistent with patterns that previous research observed for students across the nation: students who grow more during the school year compared to their peers also tend to lose more during summer (Kuhfeld, 2019). The district aims to break from this national trend and is considering aspects of DL that could be improved to help students sustain their learning gains. Many current district efforts focus on supporting summer reading, and they seem to be paying off, as reflected by the very low reading summer loss rates. In contrast, no organized curriculum or materials are provided for summer math learning. A viable first step would be to make bilingual print or online materials available for students to work on at home or organizing regular review sessions with bilingual support from adults in the community. Upon seeing these findings, the district’s EL and math departments have initiated discussions around potential approaches to minimize math summer loss.

**Limitations**

 The findings of this study should be interpreted in light of a few limitations. First, the data for this study come from a single district with a long history of serving a large Hispanic student population, the majority of whom were eligible for EL service at some point. Its Spanish-English DL is unique in many ways, including content instruction in Spanish across multiple subjects through middle school and into high school. Due to the uniqueness of the programs and the study context, my findings may not generalize well to other districts with different student demographics or different approaches to program design or implementation.

 Second, the results of this study are descriptive and do not support causal inference. Some factors do alleviate concerns for selection bias to a certain extent (e.g., program eligibility applied to entire cohorts of students, and participation was very high among Hispanic EL students in post-program cohorts). Still, the differences between participants and nonparticipants of dual language instruction should not be construed as causal estimates of program impact.

Third, this study was not able to identify the mechanisms driving achievement growth trends for DL and non-DL students. For instance, I was not able to distinguish the instructional environments and services (e.g., English-monolingual, transitional bilingual) for the non-DL students in my sample like Valentino and Reardon (2015) did. In addition, there is potential variability unaccounted for in this study, including school-level factors such as dual language program implementation and student-level factors such as Spanish and English language proficiency. My hope is for this study to catalyze further research that explain the factors contributing to disparities in achievement and identify effective strategies for addressing those disparities and facilitating student growth.

Finally, this study focuses on math and English reading achievement and does not look at other key outcomes such as Spanish proficiency and literacy, EL reclassification, or social emotional development. These outcomes are also crucial to students’ academic and overall wellbeing. The relation between DL and Spanish proficiency and literacy especially merit rigorous research since bilingualism and biliteracy—potential benefits unique to long-term dual language instruction and unfound in other EL service programs—are at the center of DL’s mission and purpose.

**Conclusion**

 This study compares the academic growth of Hispanic students who did and did not participate in Spanish-English dual language instruction. Prior research was only able to compare achievement status or explore changes over time relative to other students. Building on this body of research, I show that seasonal patterns of learning differ between participants and nonparticipants of dual language instruction. As such, this study is the first to demonstrates *how* achievement gaps develop within and across grades.

As dual language programs continue to develop across the nation, program implementation and evaluation should consider academic growth, in addition to static achievement scores, as a key outcome. In order to maximize learning throughout the student’s academic trajectory, progress needs to be monitored through regular formal or informal assessments both within and across grades. This requires collaboration between teachers across subjects and grade levels, as well as unwavering support from school- and district-level leadership. Students’ families and adults in the community are also indispensable resources for maintaining learning and growth during out-of-school time, such as summer breaks.

In the spirit of Castañeda v. Pickard (1981), as with all other programs that serve ELs and other marginalized student populations, dual language programs should be examined, formally or informally, on a regular basis. Program implementation needs to be interrogated for equity in opportunities to learn and key outcomes like academic growth. As Valdés (2018) and Chávez-Moreno (2021) pointed out, dual language educators are charged with the critical task of managing the resources that diverse groups of students bring from their homes and communities and the resources to which ELs are entitled through federal education policy. To continuously evaluate and improve dual language programs in an iterative fashion, experts on literacy, content areas, language acquisition, and data analysis all need to work together. This research partnership is an example of a collaborative effort to study and improve dual language education. I hope this study contributes to scholarly discussions that lead to more research collaborations for improving learning opportunities and outcomes for diverse student populations.

**Notes**

1 The term “Hispanic” is used throughout this paper because the partner school district used this term in their administrative data shared with me for this study; it is also the race/ethnicity category used by the National Center for Education Statistics. I acknowledge the limitations of this term in reflecting students’ individual racial/ethnic identity. Some students may have preferred “Latinx” or more specific identities such as “Mexican-American.” I did not have access to individual students and was not able to collect data on their preferred categories. Rather than making assumptions about individual students’ preferences, I am using “Hispanic” as does reporting by the school district and the federal government.

2 Non-participants of dual language include students in general education and students receiving all other types of language services, such as transitional bilingual education and pull-out support. Details on the district’s language programs and services can be found in the Study Context section.

3 In this paper, I use the federal term “English Learners (ELs)” to refer to students who are eligible to receive services due to developing English proficiency. The term ELs uniquely identifies students to whom the education system has legal obligations to provide linguistic supports. I acknowledge the linguistic and cultural assets students bring to their schools and classrooms as they develop their multilingual skills. Wherever eligibility for services is not a necessary distinction, I use the term “multilingual students.”

4 Eligibility for Special Education services is a proxy for students with disabilities. While “students with disabilities” is a more affirming term, many disabilities are not identified in the early grades, if ever during K-12. At the same time, many multilingual students can be mis-identified as having a disability, especially in the early grades. For these reasons, I use the SPED eligibility flag rather than “student with disabilities” to avoid the implication that the data included full and accurate information about students with disabilities.

5 Thum and Kuhfeld (2020) state, “MAP Growth score scales are maintained to preserve both stability and the interval scale properties of scores (e.g., Mislevy, 1987). Wang, Jiao, and Zhang (2013) and Wang, McCall, Jiao, and Harris (2013) conclude that the factorial structure of MAP Growth mathematics and reading are each unidimensional across testing terms, grades, years, and states. A recent study of MAP Growth mathematics and reading scores from Grades 3-5 students suggests that the scores possess equal-interval properties and rest on a vertical scale (Thum, 2018). Concurrent validity studies comparing MAP Growth assessments with state and consortium tests are consistently high, at about 0.8, for mathematics, reading, and science. The same studies yield classification accuracy estimates that hover above 0.8 (NWEA, 2019, Table 8.1). These findings strongly support the use of MAP Growth results to monitor student achievement and learning, across grade levels and states.” P.11-12

6 I use school covariates for the first school for each student and model them as time-invariant for theoretical and practical reasons. First, switching schools may be endogenous to the quality of and the student’s experiences at their first school; in other words, the first school may affect the probability of switching. Second, school characteristics do not vary much over time. Third, my data do not support a model that includes time-varying school-level covariates.

7 A linear model is used because my data do not support the inclusion of quadratic or higher order polynomial terms in the model. Using MAP Growth assessment data for a larger, broader sample, Kuhfeld and Soland (2021) showed that growth estimates in most parts of the year are not sensitive to functional form. Future research should investigate the sensitivity of growth estimates to functional form for students who are ELs or in dual language programs.

8 Slopes are treated as fixed because (a) variations in growth rates is of secondary interest for this study and (b) the data do not support adding random effects to the growth terms.

**References**

Alvear, S. A. (2019). The additive advantage and bilingual programs in a large urban school district. *American Educational Research Journal*, *56*(2), 477 -513. <https://doi.org/10.3102/0002831218800986>

Atteberry, A., & Mangan, D. (2020). The sensitivity of teacher value-added scores to the use of fall or spring test scores. *Educational Researcher*, *49*(5), 335-349. https://doi.org/10.3102/0013189X20922993

August, D. & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth.* Lawrence Erlbaum Associates.

Bialystok, E. (2018). Bilingual education for young children: review of the effects and consequences, *International Journal of Bilingual Education and Bilingualism, 21*(6), 666-679. DOI: 10.1080/13670050.2016.1203859

Bibler, A. (2021). Dual language education and student achievement. *Education Finance and Policy*, *16*(4), 634-658. <https://doi.org/10.1162/edfp_a_00320>

Castañeda v. Pickard*,* 648 F.2d 989. 5th Cir. (1981).

Chávez-Moreno, L. (2021). Dual language as White property: Examining a secondary bilingual-education program and Latinx equity. *American Educational Research Journal*, *58*(6), 1107-1141. DOI: 10.3102/00028312211052508

Collier, V. P. & Thomas, W. P. (2017). Validating the power of bilingual schooling: Thirty-two years of large-scale, longitudinal research. *Annual Review of Applied Linguistics* *(ARAL)* 2017.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 221–251.

Dual Language Schools. (2020). https://duallanguageschools.org/

Fountas and Pinnell. (2020). Literacy Gradient. Retrieved from <https://www.fountasandpinnell.com/assessment/>

Garcia, E. E. (2002). Bilingualism and schooling in the United States. *International Journal of the Sociology of Language*, *155–156*, 1–92. doi:10.1515/ijsl.2002.028

Goldenberg, C. (2008). Teaching English Language Learners: What the research does—and does not—say. *American Educator*, *32*(2), 8-44*.*

Howard, E. R., Lindholm-Leary, K. J., Rogers, D., Olague, N., Medina, J., Kennedy, B., Sugarman, J., & Christian, D. (2018). *Guiding Principles* *for Dual Language Education* (3rd ed.). Center for Applied Linguistics.

Hult, F. M., & Hornberger, N. H. (2016). Revisiting orientations in language planning: Problem, right, and resource as an analytical heuristic. The Bilingual Review/La Revista Bilingüe, 33(3), 30-49. <https://repository.upenn.edu/gse_pubs/476>

Johnson, A. (2022). Achievement and growth for English Learners. *Educational Policy*, *37*(4), 1044-1069. <https://doi.org/10.1177/08959048211049419>

Kuhfeld, M. (2019). Surprising new evidence on summer learning loss. *Phi Delta Kappan, 101* (1), 25-29.

Kuhfeld, M. & Soland, J. (2021). The learning curve: Revisiting the assumption of linear growth during the school year. *Journal of Research on Educational Effectiveness*, *14*(1), 143-171. DOI: 10.1080/19345747.2020.1839990

Lindholm-Leary, K. & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism, 13*(1), 43-60. DOI: 10.1080/13670050902777546

Lindholm-Leary, K. & Hernández, A. (2011). Achievement and language proficiency of Latino students in dual language programmes: native English speakers, fluent English/previous ELLs, and current ELLs. *Journal of Multilingual and Multicultural Development, 32*(6), 531-545. DOI: 10.1080/01434632.2011.611596

McCormick, M. P., Pralica, M., Guerrero-Rosada, P., Weiland, C., Hsueh, J., Condliffe, B., Sachs, J., & Snow, C. (2021). Can center-based care reduce summer slowdown prior to kindergarten? Exploring variation by family income, race/ethnicity, and dual language learner status. *American Educational Research Journal*, *58*(2), 420-455. DOI: 10.3102/0002831220944908

Phillips Galloway, E., Uccelli, P., Aguilar, G., & Barr. C. D. (2020). Exploring the cross-linguistic contribution of Spanish and English academic language skills to English text comprehension for middle-grade dual language learners. *AERA Open*, *6*(1), 1-20.

Quinn, D. M., Cooc, N., McIntyre, J., & Gomez, C. J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: Updating and extending past research with new national data. *Educational Researcher*, *45*, 443–453.

Raudenbush, S.W., Bryk, A.S, Cheong, Y.F. & Congdon, R. (2019). HLM 8 for Windows [Computer software]. Scientific Software International, Inc.

Ruíz, R. (1984). Orientations in language planning. *NABE Journal*, *8*(2), 15-34. DOI: [10.1080/08855072.1984.10668464](https://doi.org/10.1080/08855072.1984.10668464)

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal, 54*(1S), 282S-306S.

Stoehr, K., Salazar, F., & Civil, M. (2022). The power of mothers and teachers engaging in a mathematics bilingual collaboration. Teachers College Record, 124(5), 30-48. DOI: 10.1177/01614681221103947.

Thomas, W. P. & Collier, V. P. (2012). *Dual language education for a transformed world*. Fuerte Press.

Thum, Y. M. & Kuhfeld, M. (2020). *NWEA 2020 MAP Growth norms for students and school achievement status and growth*. NWEA.

**Umansky, I.** (2016). "To Be or Not to Be EL: An Examination of the Impact of Classifying Students as English Learners." Educational Evaluation and Policy Analysis, 38(4), 714-737.

Valdés, G. (2018). Analyzing the curricularization of language in twoway immersion education: Restating two cautionary notes. *Bilingual Research Journal, 41*(4), 388-412. DOI: 10.1080/15235882.2018.1539886

Valentino, R. A. & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English Learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis, 37*(4), 612-637.

Vera, E. M., Israel, M. S., Coyle, L., Cross, J., Knight-Lynn, L., Moallem, I., Bartucci, G., & Goldberger, N. (2012). Exploring the educational involvement of parents of English Learners. *School Community Journal*, *22*(2), 183-202.

von Hippel, P. T., & Hamrock, C. (2019). Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science, 6,*43–80.

von Hippel, P. T., Workman, J., & Downey, D. B. (2018). Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of “Are schools the great equalizer?” *Sociology of Education, 91*(4), 323-357.

Table 1. Sample Characteristics

|  |  |  |  |
| --- | --- | --- | --- |
|  | DL (N=9,973) | Non-DL (N=4,604) | Total (N=14,577) |
|   | Mean | SD | Mean | SD | Mean | SD |
| **Student Characteristics** |  |  |  |  |  |  |
| Female | 0.49 | 0.50 | 0.50 | 0.50 | 0.49 | 0.50 |
| Ever-EL | 0.94 | 0.25 | 0.38 | 0.48 | 0.76 | 0.43 |
| Ever-FRPL | 0.96 | 0.20 | 0.83 | 0.38 | 0.92 | 0.28 |
| Ever-SPED | 0.14 | 0.35 | 0.19 | 0.39 | 0.16 | 0.36 |
| Refused EL Service K or G1 (unconditional) | 0.05 | 0.22 | 0.26 | 0.44 | 0.12 | 0.32 |
| Refused EL Service K or G1 (cond. Ever-EL) | 0.05 | 0.22 | 0.66 | 0.47 | 0.15 | 0.36 |
|  |  |  |  |  |  |  |
| **School Characteristics** |  |  |  |  |  |  |
| % Female | 0.49 | 0.02 | 0.49 | 0.02 | 0.49 | 0.02 |
| % EL | 0.29 | 0.15 | 0.22 | 0.16 | 0.27 | 0.15 |
| % FRPL | 0.75 | 0.16 | 0.61 | 0.24 | 0.71 | 0.20 |
| % SPED | 0.11 | 0.04 | 0.11 | 0.04 | 0.11 | 0.04 |
| Notes: DL = ever enrolled in dual language program. Non-DL = never enrolled in dual language program. SD = standard deviation. EL= English Learner. FRPL = free or reduced-price lunch. SPED= special education. K = kindergarten. G1 = grade 1. Sample includes students whose reported race/ethnicity was Hispanic and who took MAP Growth math or reading assessments between 2011-12 and winter 2020. Ever-EL indicates having ever been eligible for EL services while enrolled in the district. Ever-FRPL indicates having ever been eligible for free or reduced-price lunch services. Ever-SPED indicates having ever been eligible for Special Education services. Refused EL Service K or G1 indicates having been eligible for but refused to receive EL service during kindergarten or 1st grade. |

Table 2. Number of Students in Sample by 2nd Grade Cohort and Grades Tested

|  |  |
| --- | --- |
|  | Grades Tested |
| 2nd Grade Cohort (Fall) | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2011 | 1,676 | 1,684 | 1,649 | 1,653 | 1,685 | 1,680 | 1,660 |
| 2012 | 1,642 | 1,616 | 1,591 | 1,628 | 1,643 | 1,638 | 1,611 |
| 2013 | 1,636 | 1,619 | 1,649 | 1,665 | 1,678 | 1,670 | 1,572 |
| 2014 | 1,597 | 1,636 | 1,633 | 1,629 | 1,635 | 1,513 |  |
| 2015 | 1,492 | 1,509 | 1,521 | 1,527 | 1,432 |  |  |
| 2016 | 1,600 | 1,627 | 1,605 | 1,064 |  |  |  |
| 2017 | 1,434 | 1,416 | 1,008 |  |  |  |  |
| 2018 | 1,515 | 1,047 |   |   |   |   |   |

Table 3. Estimated Monthly Growth Rates for Full Sample

|  |  |  |
| --- | --- | --- |
|   | Math | Reading |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
|   |  |  |  |  |  |  |
| Starting RIT | 171.609\*\*\* | 179.503\*\*\* | 179.473\*\*\* | 167.584\*\*\* | 175.961\*\*\* | 175.938\*\*\* |
|  | (0.620) | (0.664) | (0.604) | (0.839) | (0.768) | (1.085) |
| Gr 2 AY | 1.788\*\*\* | 1.785\*\*\* | 1.784\*\*\* | 1.839\*\*\* | 1.833\*\*\* | 1.833\*\*\* |
|  | (0.033) | (0.033) | (0.033) | (0.027) | (0.027) | (0.027) |
| **DL x Gr 2 AY** | **0.182\*\*\*** | **0.184\*\*\*** | **0.184\*\*\*** | **0.089\*** | **0.093\*** | **0.093\*** |
|  | (0.063) | (0.063) | (0.063) | (0.049) | (0.049) | (0.049) |
| Gr 2 Summer | -0.489\*\*\* | -0.474\*\*\* | -0.473\*\*\* | -0.183\*\* | -0.160\* | -0.160\* |
|  | (0.106) | (0.105) | (0.105) | (0.088) | (0.087) | (0.087) |
| **DL x Gr 2 Summer** | **-1.115\*\*\*** | **-1.126\*\*\*** | **-1.126\*\*\*** | **-0.293\*\*** | **-0.312\*\*** | **-0.312\*\*** |
|  | (0.148) | (0.148) | (0.148) | (0.126) | (0.125) | (0.125) |
| Gr 3 AY | 1.441\*\*\* | 1.439\*\*\* | 1.439\*\*\* | 1.148\*\*\* | 1.146\*\*\* | 1.146\*\*\* |
|  | (0.032) | (0.032) | (0.032) | (0.032) | (0.032) | (0.032) |
| **DL x Gr 3 AY** | **0.177\*\*\*** | **0.178\*\*\*** | **0.178\*\*\*** | **0.057** | **0.059\*** | **0.059\*** |
|  | (0.037) | (0.037) | (0.037) | (0.036) | (0.036) | (0.036) |
| Gr 3 Summer | -0.872\*\*\* | -0.865\*\*\* | -0.865\*\*\* | -0.141 | -0.133 | -0.133 |
|  | (0.094) | (0.094) | (0.094) | (0.094) | (0.094) | (0.094) |
| **DL x Gr 3 Summer** | **-0.531\*\*\*** | **-0.538\*\*\*** | **-0.538\*\*\*** | **0.070** | **0.059** | **0.058** |
|  | (0.104) | (0.105) | (0.105) | (0.116) | (0.116) | (0.116) |
| Gr 4 AY | 1.125\*\*\* | 1.123\*\*\* | 1.123\*\*\* | 0.840\*\*\* | 0.838\*\*\* | 0.837\*\*\* |
|  | (0.031) | (0.031) | (0.031) | (0.025) | (0.025) | (0.025) |
| **DL x Gr 4 AY** | **0.259\*\*\*** | **0.261\*\*\*** | **0.261\*\*\*** | **0.083\*\*** | **0.086\*\*** | **0.086\*\*** |
|  | (0.036) | (0.036) | (0.036) | (0.038) | (0.038) | (0.038) |
| Gr 4 Summer | -0.758\*\*\* | -0.754\*\*\* | -0.754\*\*\* | -0.177\*\* | -0.173\*\* | -0.173\*\* |
|  | (0.076) | (0.075) | (0.075) | (0.086) | (0.085) | (0.085) |
| **DL x Gr 4 Summer** | **-0.642\*\*\*** | **-0.647\*\*\*** | **-0.647\*\*\*** | **0.118** | **0.108** | **0.108** |
|  | (0.110) | (0.110) | (0.110) | (0.094) | (0.094) | (0.094) |
| Gr 5 AY | 0.891\*\*\* | 0.891\*\*\* | 0.891\*\*\* | 0.613\*\*\* | 0.613\*\*\* | 0.613\*\*\* |
|  | (0.028) | (0.028) | (0.028) | (0.031) | (0.031) | (0.031) |
| **DL x Gr 5 AY** | **0.282\*\*\*** | **0.282\*\*\*** | **0.282\*\*\*** | **0.089\*\*** | **0.089\*\*** | **0.089\*\*** |
|  | (0.038) | (0.038) | (0.038) | (0.041) | (0.041) | (0.041) |
| Gr 5 Summer | -1.427\*\*\* | -1.431\*\*\* | -1.431\*\*\* | 0.252\*\*\* | 0.246\*\*\* | 0.246\*\*\* |
|  | (0.082) | (0.082) | (0.082) | (0.095) | (0.095) | (0.095) |
| **DL x Gr 5 Summer** | **-0.618\*\*\*** | **-0.615\*\*\*** | **-0.615\*\*\*** | **-0.048** | **-0.046** | **-0.047** |
|  | (0.102) | (0.102) | (0.102) | (0.113) | (0.113) | (0.113) |
| Gr 6 AY | 0.878\*\*\* | 0.878\*\*\* | 0.878\*\*\* | 0.489\*\*\* | 0.490\*\*\* | 0.490\*\*\* |
|  | (0.029) | (0.029) | (0.029) | (0.023) | (0.023) | (0.023) |
| **DL x Gr 6 AY** | **0.034** | **0.035** | **0.035** | **0.042** | **0.043** | **0.043** |
|  | (0.035) | (0.035) | (0.035) | (0.037) | (0.037) | (0.037) |
| Gr 6 Summer | -0.957\*\*\* | -0.957\*\*\* | -0.958\*\*\* | -0.356\*\*\* | -0.356\*\*\* | -0.357\*\*\* |
|  | (0.079) | (0.079) | (0.079) | (0.064) | (0.064) | (0.064) |
| **DL x Gr 6 Summer** | **-0.312\*\*\*** | **-0.312\*\*\*** | **-0.311\*\*\*** | **-0.006** | **-0.009** | **-0.009** |
|  | (0.103) | (0.103) | (0.103) | (0.108) | (0.109) | (0.109) |
| Gr 7 AY | 0.613\*\*\* | 0.613\*\*\* | 0.613\*\*\* | 0.315\*\*\* | 0.316\*\*\* | 0.316\*\*\* |
|  | (0.035) | (0.035) | (0.035) | (0.026) | (0.026) | (0.026) |
| **DL x Gr 7 AY** | **-0.054** | **-0.055** | **-0.055** | **0.086\*\*** | **0.085\*\*** | **0.085\*\*** |
|  | (0.043) | (0.043) | (0.043) | (0.038) | (0.038) | (0.038) |
| Gr 7 Summer | -0.410\*\*\* | -0.411\*\*\* | -0.412\*\*\* | 0.239\*\* | 0.238\*\* | 0.238\*\* |
|  | (0.089) | (0.089) | (0.089) | (0.102) | (0.102) | (0.102) |
| **DL x Gr 7 Summer** | **0.139** | **0.138** | **0.139** | **-0.263\*\*** | **-0.264\*\*** | **-0.264\*\*** |
|  | (0.107) | (0.106) | (0.106) | (0.113) | (0.113) | (0.113) |
| Gr 8 AY | 0.741\*\*\* | 0.740\*\*\* | 0.740\*\*\* | 0.662\*\*\* | 0.663\*\*\* | 0.663\*\*\* |
|  | (0.061) | (0.061) | (0.061) | (0.078) | (0.078) | (0.078) |
| **DL x Gr 8 AY** | **-0.114** | **-0.112** | **-0.112** | **0.081** | **0.080** | **0.081** |
|  | (0.081) | (0.081) | (0.081) | (0.104) | (0.104) | (0.104) |
| Ever-DL | -3.067\*\*\* | -0.414 | -0.328 | -8.384\*\*\* | -3.790\*\*\* | -3.713\*\*\* |
|  | (0.574) | (0.569) | (0.572) | (0.594) | (0.603) | (0.597) |
| Female |  | -1.634\*\*\* | -1.627\*\*\* |  | 1.054\*\*\* | 1.060\*\*\* |
|  |  | (0.151) | (0.152) |  | (0.171) | (0.172) |
| Ever-FRPL |  | -3.955\*\*\* | -3.843\*\*\* |  | -4.455\*\*\* | -4.356\*\*\* |
|  |  | (0.308) | (0.309) |  | (0.428) | (0.436) |
| Ever-EL |  | -5.284\*\*\* | -5.269\*\*\* |  | -8.512\*\*\* | -8.500\*\*\* |
|  |  | (0.414) | (0.412) |  | (0.459) | (0.454) |
| Ever-SPED |  | -11.421\*\*\* | -11.417\*\*\* |  | -14.580\*\*\* | -14.576\*\*\* |
|  |  | (0.352) | (0.352) |  | (0.350) | (0.350) |
| Refused EL Service |  | 0.832\*\* | 0.843\*\* |  | 2.159\*\*\* | 2.171\*\*\* |
|  |  | (0.350) | (0.349) |  | (0.402) | (0.398) |
| School % FRPL |  |  | -8.362\*\* |  |  | -11.799 |
|  |  |  | (3.349) |  |  | (7.216) |
| School % EL |  |  | 4.613 |  |  | 9.340 |
|  |  |  | (7.576) |  |  | (16.336) |
| Student Covariates | no | yes | yes | no | yes | yes |
| School Covariates | no | no | yes | no | no | yes |
| Test Scores | 156212 | 156212 | 156212 | 155459 | 155459 | 155459 |
| Students | 14577 | 14577 | 14577 | 14577 | 14577 | 14577 |
| Schools | 50 | 50 | 50 | 50 | 50 | 50 |
| Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. EL=English Learner. DL = dual language. RIT = Rasch Unit. SPED = Special Education. FRPL = Free or reduced-price lunch. Gr = grade. AY = Academic year. Sample includes Hispanic students who took MAP Growth assessments between 2011-12 and winter 2020 in 2nd to 8th grade. Each column is a separate 3-level piecewise growth model regression with RIT score as the dependent variable. Starting RIT is the intercept and represents test scores on the first day of 2nd grade. Ever-DL is an indicator for ever enrolling in the dual language program. Female, Ever-FRPL, Ever-EL, Ever-SPED, and refused EL service are student-level covariates. Refused EL service represents having refused EL service in either kindergarten or 1st grade. School % FRPL and school % EL are school-level covariates. Main effects represent growth estimates for non-DL students. Interaction effects representing differential growth for dual language participants are bolded. |

Table 4. Estimated Monthly Growth Rates for Ever-ELs

|  |  |  |
| --- | --- | --- |
|   | Math | Reading |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
|   |  |  |  |  |  |  |
| Starting RIT | 167.547\*\*\* | 174.216\*\*\* | 174.264\*\*\* | 161.479\*\*\* | 168.316\*\*\* | 168.381\*\*\* |
|  | (0.828) | (0.991) | (0.960) | (1.626) | (1.265) | (1.379) |
| Gr 2 AY | 1.777\*\*\* | 1.777\*\*\* | 1.777\*\*\* | 1.794\*\*\* | 1.792\*\*\* | 1.792\*\*\* |
|  | (0.039) | (0.038) | (0.038) | (0.038) | (0.038) | (0.038) |
| **DL x Gr 2 AY** | **0.196\*\*\*** | **0.195\*\*\*** | **0.195\*\*\*** | **0.123\*\*** | **0.124\*\*** | **0.124\*\*** |
|  | (0.068) | (0.068) | (0.068) | (0.057) | (0.056) | (0.056) |
| Gr 2 Summer | -0.592\*\*\* | -0.583\*\*\* | -0.583\*\*\* | -0.181\* | -0.165 | -0.165 |
|  | (0.115) | (0.114) | (0.114) | (0.106) | (0.106) | (0.106) |
| **DL x Gr 2 Summer** | **-1.070\*\*\*** | **-1.078\*\*\*** | **-1.078\*\*\*** | **-0.303\*\*** | **-0.316\*\*** | **-0.316\*\*** |
|  | (0.144) | (0.142) | (0.142) | (0.136) | (0.135) | (0.135) |
| Gr 3 AY | 1.494\*\*\* | 1.490\*\*\* | 1.490\*\*\* | 1.164\*\*\* | 1.161\*\*\* | 1.161\*\*\* |
|  | (0.041) | (0.042) | (0.042) | (0.042) | (0.042) | (0.042) |
| **DL x Gr 3 AY** | **0.141\*\*\*** | **0.144\*\*\*** | **0.144\*\*\*** | **0.049** | **0.053** | **0.053** |
|  | (0.046) | (0.046) | (0.046) | (0.046) | (0.046) | (0.046) |
| Gr 3 Summer | -0.950\*\*\* | -0.939\*\*\* | -0.939\*\*\* | -0.218\* | -0.207\* | -0.207\* |
|  | (0.129) | (0.129) | (0.129) | (0.116) | (0.116) | (0.116) |
| **DL x Gr 3 Summer** | **-0.507\*\*\*** | **-0.517\*\*\*** | **-0.517\*\*\*** | **0.119** | **0.105** | **0.105** |
|  | (0.144) | (0.145) | (0.145) | (0.136) | (0.136) | (0.136) |
| Gr 4 AY | 1.056\*\*\* | 1.056\*\*\* | 1.056\*\*\* | 0.918\*\*\* | 0.918\*\*\* | 0.919\*\*\* |
|  | (0.050) | (0.050) | (0.050) | (0.040) | (0.040) | (0.040) |
| **DL x Gr 4 AY** | **0.345\*\*\*** | **0.345\*\*\*** | **0.344\*\*\*** | **0.022** | **0.021** | **0.021** |
|  | (0.048) | (0.048) | (0.048) | (0.051) | (0.051) | (0.051) |
| Gr 4 Summer | -0.751\*\*\* | -0.748\*\*\* | -0.749\*\*\* | -0.257\*\* | -0.254\*\* | -0.254\*\* |
|  | (0.094) | (0.094) | (0.094) | (0.120) | (0.120) | (0.120) |
| **DL x Gr 4 Summer** | **-0.692\*\*\*** | **-0.694\*\*\*** | **-0.694\*\*\*** | **0.163** | **0.158** | **0.158** |
|  | (0.127) | (0.127) | (0.127) | (0.129) | (0.129) | (0.129) |
| Gr 5 AY | 0.926\*\*\* | 0.925\*\*\* | 0.925\*\*\* | 0.686\*\*\* | 0.686\*\*\* | 0.686\*\*\* |
|  | (0.037) | (0.037) | (0.037) | (0.043) | (0.044) | (0.044) |
| **DL x Gr 5 AY** | **0.247\*\*\*** | **0.248\*\*\*** | **0.248\*\*\*** | **0.027** | **0.028** | **0.028** |
|  | (0.042) | (0.042) | (0.042) | (0.050) | (0.050) | (0.050) |
| Gr 5 Summer | -1.428\*\*\* | -1.428\*\*\* | -1.428\*\*\* | 0.108 | 0.105 | 0.105 |
|  | (0.114) | (0.114) | (0.114) | (0.111) | (0.112) | (0.112) |
| **DL x Gr 5 Summer** | **-0.616\*\*\*** | **-0.617\*\*\*** | **-0.617\*\*\*** | **0.076** | **0.074** | **0.074** |
|  | (0.131) | (0.131) | (0.131) | (0.124) | (0.125) | (0.125) |
| Gr 6 AY | 0.847\*\*\* | 0.847\*\*\* | 0.847\*\*\* | 0.545\*\*\* | 0.547\*\*\* | 0.547\*\*\* |
|  | (0.034) | (0.034) | (0.034) | (0.039) | (0.039) | (0.039) |
| **DL x Gr 6 AY** | **0.069\*** | **0.069\*** | **0.069\*** | **-0.008** | **-0.009** | **-0.009** |
|  | (0.036) | (0.036) | (0.036) | (0.046) | (0.046) | (0.046) |
| Gr 6 Summer | -0.968\*\*\* | -0.966\*\*\* | -0.966\*\*\* | -0.411\*\*\* | -0.410\*\*\* | -0.411\*\*\* |
|  | (0.110) | (0.110) | (0.110) | (0.141) | (0.141) | (0.141) |
| **DL x Gr 6 Summer** | **-0.332\*\*\*** | **-0.334\*\*\*** | **-0.334\*\*\*** | **0.028** | **0.026** | **0.026** |
|  | (0.119) | (0.118) | (0.118) | (0.152) | (0.152) | (0.152) |
| Gr 7 AY | 0.597\*\*\* | 0.598\*\*\* | 0.598\*\*\* | 0.376\*\*\* | 0.378\*\*\* | 0.378\*\*\* |
|  | (0.045) | (0.045) | (0.045) | (0.043) | (0.043) | (0.043) |
| **DL x Gr 7 AY** | **-0.050** | **-0.051** | **-0.051** | **0.034** | **0.032** | **0.032** |
|  | (0.053) | (0.053) | (0.053) | (0.059) | (0.059) | (0.059) |
| Gr 7 Summer | -0.330\*\*\* | -0.330\*\*\* | -0.330\*\*\* | 0.083 | 0.081 | 0.081 |
|  | (0.119) | (0.119) | (0.119) | (0.168) | (0.168) | (0.168) |
| **DL x Gr 7 Summer** | **0.080** | **0.079** | **0.079** | **-0.100** | **-0.098** | **-0.098** |
|  | (0.132) | (0.131) | (0.131) | (0.177) | (0.177) | (0.177) |
| Gr 8 AY | 0.691\*\*\* | 0.687\*\*\* | 0.687\*\*\* | 0.588\*\*\* | 0.586\*\*\* | 0.585\*\*\* |
|  | (0.124) | (0.124) | (0.124) | (0.144) | (0.145) | (0.145) |
| **DL x Gr 8 AY** | **-0.068** | **-0.064** | **-0.064** | **0.137** | **0.140** | **0.141** |
|  | (0.126) | (0.126) | (0.126) | (0.164) | (0.164) | (0.164) |
| Ever-DL | -0.675 | -1.792\*\* | -1.707\*\* | -4.519\*\*\* | -5.673\*\*\* | -5.586\*\*\* |
|  | (0.665) | (0.715) | (0.719) | (0.621) | (0.729) | (0.728) |
| Female |  | -1.642\*\*\* | -1.635\*\*\* |  | 0.921\*\*\* | 0.928\*\*\* |
|  |  | (0.178) | (0.178) |  | (0.196) | (0.197) |
| Ever-FRPL |  | -3.443\*\*\* | -3.368\*\*\* |  | -4.419\*\*\* | -4.350\*\*\* |
|  |  | (0.518) | (0.523) |  | (0.642) | (0.648) |
| Ever-SPED |  | -11.301\*\*\* | -11.295\*\*\* |  | -14.366\*\*\* | -14.360\*\*\* |
|  |  | (0.441) | (0.441) |  | (0.451) | (0.452) |
| Refused EL Service |  | 0.251 | 0.251 |  | 0.891\* | 0.897\* |
|  |  | (0.457) | (0.456) |  | (0.484) | (0.483) |
| School % FRPL |  |  | -12.343\*\*\* |  |  | -22.529\*\*\* |
|  |  |  | (4.620) |  |  | (8.533) |
| School % EL |  |  | 17.800\* |  |  | 32.515\* |
|  |  |  | (9.837) |  |  | (18.730) |
| Student Covariates | no | yes | yes | no | yes | yes |
| School Covariates | no | no | yes | no | no | yes |
| Test Scores | 120591 | 120591 | 120591 | 119877 | 119877 | 119877 |
| Students | 11066 | 11066 | 11066 | 11066 | 11066 | 11066 |
| Schools | 50 | 50 | 50 | 50 | 50 | 50 |
| Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. EL=English Learner. DL = dual language. RIT = Rasch Unit. SPED = Special Education. FRPL = Free or reduced-price lunch. Gr = grade. AY = Academic year. Sample includes Hispanic students who were ever eligible for EL service and took MAP Growth assessments between 2011-12 and winter 2020 in 2nd to 8th grade. Each column is a separate 3-level piecewise growth model regression with RIT score as the dependent variable. Starting RIT is the intercept and represents test scores on the first day of 2nd grade. Ever-DL is an indicator for ever enrolling in the dual language program. Female, Ever-FRPL, Ever-SPED, and refused EL service are student-level covariates. Refused EL service represents having refused EL service in either kindergarten or 1st grade. School % FRPL and school % EL are school-level covariates. Main effects represent growth estimates for non-DL students. Interaction effects representing differential growth for dual language participants are bolded. |

Figure 1. Spring Standardized Achievement Scores Across Grades



Notes: Sample includes Hispanic students who took MAP Growth assessments between 2011-12 and winter 2020 in 2nd to 8th grade. DL represents students who ever enrolled in the dual language program. Non-DL represents students who never enrolled in the dual language program. Z-scores are calculated using spring test scores and NWEA MAP Growth Norms 2020 for each grade and subject.

Figure 2. Mean Achievement Scores by Dual Language Participation





Notes: DL represents students who ever enrolled in the dual language program. Non-DL represents students who never enrolled in the dual language program. Trends represent unconditional means for each group of students at each test grade and term. Trends include test scores from fall of 2nd grade to winter of 8th grade. Winter terms are labeled because the district reports winter scores for accountability.

Figure 3. Estimated Monthly Growth Rates in School Year and Summer

Notes: DL represents students who ever enrolled in the dual language program. Non-DL represents students who never enrolled in the dual language program. Monthly growth rates are estimated using 3-level piecewise growth models and correspond to Model (3) and Columns (3) and (6) in Table 3. Sample includes Hispanic students who took MAP Growth assessments between 2011-12 and winter 2020 in 2nd to 8th grade.