ACHIEVEMENT EFFECTS OF DUAL LANGUAGE IMMERSION IN ONE-WAY AND TWO-WAY PROGRAMS: EVIDENCE FROM A STATE SCALE-UP IN UTAH

Jennifer L. Steele, Associate Professor of Education and Faculty Affiliate in the Dept. of Public Administration and Policy, American University, 4400 Massachusetts Ave., NW, Washington, DC, 20016-8030, (202) 885-3762, steele@american.edu *(Corresponding Author)*

Johanna Watzinger-Tharp, Associate Professor of Linguistics, University of Utah, 255 Central Campus Dr #2300, Salt Lake City, UT 84112, (801) 585-7148 j.tharp@utah.edu

Robert O. Slater, Co-Director of the American Councils Research Center, American Councils for International Education, 1828 L St NW Suite 1200, Washington, DC 20036, (202) 833-7522, rslater@americancouncils.org

Gregg Roberts, Director of Dual Language Studies, American Councils for International Education, 1828 L St NW Suite 1200, Washington, DC 20036, (202) 833-7522, groberts@americancouncils.org

Karl Bowman, Dual Language Immersion and World Language Specialist, Utah State Board of Education, 250 E 500 S, Salt Lake City, UT 84111, (801) 538-7500, Karl.Bowman@schools.utah.gov

RUNNING HEAD: ACHIEVEMENT EFFECTS OF DUAL LANGUAGE IMMERSION

ABSTRACT

The rising U.S. demand for dual-language immersion (DLI) programs, which offer core instruction in two languages from early grands onward, has raised questions about program design and access. We leverage the rapid expansion of DLI schools across Utah to estimate effects of DLI program availability on the academic achievement of native English speakers and English learners (ELs) in programs that serve mainly the former (one-way) and those comprising one-to-two thirds of the latter (two-way). Adjusting for school fixed effects, cross-grade intent-to-treat estimates in one-way programs are largely null, but those in two-way programs reach 0.10-0.11 standard deviations in math and English and show higher EL reclassification rates by grade 5. Estimates may suggest an advantage of cultural adjacency in program design.

KEYWORDS: dual language immersion, English learners, student achievement, state policy, culturally relevant instruction, school fixed effects

ACKNOWLEDGEMENTS

The authors are grateful to the U.S. Department of Education's Institute of Educational Sciences for supporting this work through Research Practice Partnership Grant #R305H170005. This research has benefitted from feedback from IES program officers Molly Faulkner-Bond and Allen Ruby and from attendees at annual meetings of the Association for Public Policy Analysis and Management and the Association of Education Finance and Policy. We are very grateful for data preparation and support from Kristin Campbell at the Utah State Board of Education and to Dual Language Immersion directors across the state who provided survey and DLI enrollment data. The views expressed here are those of the authors and do not represent the views of IES or any of the research partnership organizations.

ACHIEVEMENT EFFECTS OF DUAL LANGUAGE IMMERSION IN ONE-WAY AND TWO-WAY PROGRAMS: EVIDENCE FROM A STATE SCALE-UP IN UTAH

Abstract

The rising U.S. demand for dual-language immersion (DLI) programs, which offer core instruction in two languages from early grades onward, has raised questions about program design and access. We leverage the rapid expansion of DLI schools across Utah to estimate effects of DLI program availability on the academic achievement of primary English speakers and English learners (ELs) in programs that serve mainly the former (one-way) and those comprising one-to-two thirds of the latter (two-way). Adjusting for school fixed effects, cross-grade intent-to-treat estimates in one-way programs are largely null, but those in two-way programs reach 0.10-0.11 standard deviations in math and English and show higher EL reclassification rates by grade 5. Estimates suggest an advantage of cultural adjacency in program design.

Introduction

Policy considerations around language education are often fraught. The United Nations recommends that young children in linguistically diverse societies have access to education in their primary language (UNESCO, 2016a), but fulfillment of this recommendation has proven difficult both logistically and politically in many nations (UNESCO, 2016b). In the United States, advocacy for "English only" policies in schools and other public domains, linked closely to anti-immigrant ideology (Padilla et al., 1991), yielded voter-initiated bans on bilingual education for English learners (ELs) in California, Arizona, and Massachusetts from the late 1990s through the mid-2010s (Lam & Richards, 2020; Mora, 2009). In the past decade, however, the U.S. has seen a surge of public interest in bilingual and dual-language education programs as a means not only of supporting the roughly 5 million English learners (ELs) in public schools, but also of promoting world language proficiency in the U.S., and bilingual education bans were overturned in California and Massachusetts (Commission on Language Learning, 2017; Kamenetz, 2016). In particular, dual language immersion (DLI)—an instructional model that delivers core content instruction in two languages to primary English speakers and ELs alike from early grades onward—has gained prominence as the public has become aware of the cognitive and economic advantages of bilingualism (Fabián Romero, 2017; Maxwell, 2014).[1]

DLI programs offer general academic instruction in two languages beginning in early grades and often extending into middle or high school. They include both *two-way programs*, in which at least a third of classroom students are primary speakers of each of the two classroom languages (in the U.S., typically English and a non-English "partner" language), and *one-way programs*, in which most students in the classroom share a common primary language and are immersed in a non-primary partner language. Both types of programs are designed to move students toward bilingualism and biliteracy, regardless of their primary or home languages (Fortune, 2012). But by design, two-way programs can facilitate rapid access to content and communication for ELs alongside their primary English-speaking counterparts. In contrast, one-way programs operate under the assumption that students share a common primary language (in our study, English) and are new to the partner language, though in practice, they may still serve some language minority students, including primary speakers of the partner language.[2]

[1] We use the term "primary English speaker" to refer to students who enter school proficient in English, even though some may be primary speakers of other languages as well. Throughout the paper, we use the term "primary language" to describe what parents report as the child's first or home language, and we use the term "EL" to refer to students who enter school without English proficiency.

[2] These terms can have different meanings depending on context. In some places, one-way immersion connotes "English-only" immersion for English learners, or what we refer to in this paper as monolingual English instruction. The commonality in terms is that in one-way programs, a large majority of students in the classroom share a primary language and are working to learn the same partner language. Of course, whether the primary language they share is the socially dominant language matters to their experience and

In 2008, aiming to prepare its young people for a competitive global economy, Utah became the first U.S. state to invest in dual-language education statewide (Utah Senate, 2016). It established a common DLI curriculum and teacher professional development program and provided schools with $10,000 for each new grade level in which they offered DLI. By the 2019-2020 academic year, the state featured 244 DLI schools dispersed across 22 of its 41 districts and enrolling about 57,900 DLI students, including 75 programs in Mandarin Chinese, 32 in French, 1 in German, 13 in Portuguese, 1 in Russian, and 113 in Spanish. Thirty-one of the Spanish programs were classified as two-way, meaning that at least one-third and no more than two-thirds of students had reported Spanish as their home or primary language at the time of enrollment.

This study estimates plausibly causal effects of that scale-up effort on the academic achievement of students in schools that launched DLI programs by comparing before-and-after academic performance within the same schools, net of observed student and cohort attributes. We estimate the effects of DLI access expansion on core academic performance in grades 3-6 across the 22 Utah school districts that eventually adopted DLI. We also estimate DLI-access effects on the reclassification of ELs as English proficient.

Our study contributes to the research on DLI in several important ways. First, our school fixed-effects approach—otherwise understood as a difference-in-differences approach with staggered introductions of treatment—sheds new light on the efficacy of DLI programs as large-scale school reform vehicles. Second, by leveraging the insights of a federally funded research-practice partnership, we consider estimated achievement effects alongside the broader goals of DLI programs to prepare graduates who are bilingual and biliterate. Third and most importantly,

to their position within the social context. In this paper, most students in one-way programs are primary English speakers.

our examination of the differential effects of one-way and two-way DLI programs illuminates possible cultural and linguistic mechanisms that previous studies have been unable to examine in depth.

Our analysis finds null to negative intent-to-treat (ITT) effects of one-way programs on student achievement in core content areas, and null to positive effects of two-way programs, reaching about a third of a standard deviation (SD) in sixth grade in English language arts (ELA), math, and science. These findings are robust to sensitivity tests for change over time in sorting into treated schools, for differences in the middle school transition grades, and for a placebo test on schools' DLI conversion years to test for endogenous school-level selection into DLI.[3] Differences in estimated effects between one-way and two-way programs are not explained by the use of Spanish as a partner language or by the demographics of the school in the last pre-DLI year, but estimates do covary strongly with the share of students in the school whose primary language matches the partner language. Findings are generally consistent for both primary English speakers and ELs. Because effects differ by the level of alignment between the home and school languages of the student body, they provide suggestive evidence for the role of cultural adjacency in support of student achievement.

In this article, we briefly summarize the existing literature on DLI program effects and how our current study contributes to the knowledge base. We then describe Utah's DLI program and our dataset. Next, we present our analytic approach, followed by our descriptive and ITT results, including the results of sensitivity tests. We also present evidence about differential

[3] They are also robust to a test of weighting for attrition probability described in Online-Only Appendix II. The rates of overall and differential attrition that we report in that appendix are low and fall below the most conservative threshold maintained by the U.S. Department of Education's What Works Clearinghouse (2020).

effects in one-way and two-way programs. We conclude with a discussion of implications for policymakers.

Literature Review

Recent estimates place the number of public dual-language immersion schools at about 3,000, implying that they account for about 2% of public schools in the United States (Lam & Richards, 2020). Though this figure remains modest, it represents a five-fold increase from nine years ago, when the leading estimate was 600 programs nationally (Center for Applied Linguistics, 2011a, 2011b). In addition, public demand for these programs is strong in many cities across the U.S., yielding long wait lists and raising concerns about equitable access (Lam & Richards, 2020; Williams, 2017). The growing embrace of dual language education may be driven, at least in part, by economic concerns. Rigorous estimates of the earnings returns to bilingualism in North America range from 2-3% for non-English languages in the U.S. (Saiz & Zoido, 2005), to 4-6% percent for French in Anglophone Canada (Christofides & Swidinsky, 2010), and demand for bilingual workers in many sectors of the U.S. economy appears to be growing (Committee for Economic Development, 2006).

Mounting evidence suggests that bilingualism carries cognitive advantages. Bilinguals outperform monolinguals on many tests of cognition, including those of working memory, attention control, and task switching (e.g., Bialystok, 2011; Bialystok & Craik, 2010), and such skills have shown links to academic achievement (Gathercole, Alloway, Willis, & Adams, 2006). Bilingualism also appears to benefit metalinguistic awareness (Cenoz, 2003; Keshavarz & Astaneh, 2004) and children's social perceptive-taking skills (Fan, Liberman, Keysar, & Kinzler, 2016; Greenberg, Bellana, & Bialystok, 2013). Bilingualism can also be important for connecting students to the family cultures of their heritage languages (Potowski, 2004).

Our growing understanding of bilingual benefits has raised questions about whether schools should be providing earlier and more intensive exposure to multiple languages, as through DLI programs. Studies of French DLI programs serving primary English speakers in Canada and the U.S. have shown that immersion students perform as well as or better than their peers in English-tested content by about fifth grade (Barik & Swain, 1978; Caldas & Boudreaux, 1999; Lapkin, Hart, & Turnbull, 2003; Marian, Shook, & Schroeder, 2013), and some of these studies used baseline matching on pre-interevention characteristics (Lambert, Genesee, Holobow, & Chartrand, 1993; Lambert, Tucker, & d'Anglejan, 1973). More recently, Authors (2016) matched DLI students from 26 elementary schools to similar non-DLI students in matched non-DLI schools in Utah, finding no significant differences in math performance in grade 3, but three additional percentile points of math growth from grades 3 to 4 among DLI students.

Most studies of DLI in the U.S., however, have focused on the academic performance of ELs, comparing DLI programs to other types of language support programs. Researchers have often examined differences in outcomes between ELs taught in English-only or transitional bilingual programs, which focus on English language development, versus those taught in developmental bilingual or DLI programs, which promote maintenance of the students' non-English home language. These studies have sometimes shown vastly better performance by ELs enrolled in two-way immersion programs than in transitional bilingual or English-only programs. But they have typically failed to adjust for the selection of families into programs (Collier & Thomas, 2004; De Jong, 2004; Lindholm-Leary & Block, 2010).

Newer work has attempted to provide plausibly causal estimates of DLI effects on ELs and primary English speakers using econometric methods. Employing data from a large urban district and using extensive statistical controls, Umansky and Reardon (2014) examined EL reclassification rates of about 5,400 Spanish-speaking ELs assigned to DLI, transitional or developmental bilingual programs, or monolingual English programs. They found that cumulative reclassification rates were highest for monolingual English programs until grade 7, at which point DLI programs surpassed them, reaching a 13-point advantage by the end of high school. Examining nearly 14,000 students from the same district, and adding fixed effects for parent program preferences alongside student and school controls, Valentino and Reardon (2015) found that ELs placed in DLI programs grew at a faster rate in ELA than their peers placed in transitional bilingual, developmental bilingual, and monolingual English programs, and their ELA performance exceeded that of similar peers in developmental bilingual and monolingual English programs by grade 6.

Leveraging the launches of English-Spanish DLI programs in Spain, where most students were primary Spanish speakers learning the non-dominant partner language (English), Anghel, Cabrales, and Carro (2016) examined the sixth-grade achievement of about 4,000 students whose preschools were selected to begin offering DLI programs when the students reached first grade. Comparing the sixth-grade exam scores of students in treated versus untreated schools across two years of DLI program launches, similar to the approach we adopt in the current study, the authors found no statistically significant achievement effects on subjects taught in Spanish (math and reading) or on those taught in English (science, history, and geography).

Other recent studies have used data from oversubscribed DLI school lotteries to identify causal program effects. Steele and colleagues (2017) focused on 1,625 students randomized through pre-K or kindergarten lotteries in Portland, Oregon, finding higher ELA achievement among DLI lottery winners of 0.13 SD in grade 5 and 0.22 SD in grade 8. They found no statistically significant differences in effects between ELs and primary English speakers or between one-way and two-way programs, but the study was not powered to detect subgroup effects. They also found that ELs randomly assigned to DLI began to be reclassified at higher rates than their non-DLI peers by grade 6. The cross-grade average benefit was 0.09 SD, at an average cost of about 2.8% of per-pupil spending per DLI student (Steele et al., 2018). Employing data from 510 kindergarten lottery applicants to two two-way-DLI programs in Charlotte-Mecklenburg, North Carolina, Bibler (2020) estimated per year ITT effects of 0.037 SD in reading for primary English speakers and 0.055 SD in math for ELs, with Local Average Treatment Effects about 25% larger. Such estimates imply large cumulative ITT benefits of 0.22 SD in reading for primary English speakers and 0.33 SD in math for ELs by grade 8.

An important nuance is that one-way and two-way DLI programs may provide very different student experiences, especially for English learners and other students whose home language is not English. For an EL whose primary language matches the partner language, both types of programs offer access to at least half-time instruction in the primary language, facilitating access to academic content, but two-way programs may also offer greater affirmation of the partner language among peers and teachers in the school. For primary English speakers as well, two-way programs may offer a more complete language immersion experience among peers who are already fluent in the partner language. In addition, research on culturally relevant instruction suggests that cultural alignment between the partner language and a critical mass of students in the school could influence the effects of DLI programs. For instance, in describing the practices of culturally relevant instruction used by successful teachers of African American students, Ladson-Billings (1992, p. 387) noted that “[s]tudents’ real life experiences are legitimated as part of the ‘official curriculum.’” Moll and González (1994) described how schools in four language-minority communities helped students draw on the "funds of knowledge" in their communities, "taking full advantage of social and cultural resources in the service of academic goals" (p. 441). Paris and Alim (2014) build on this idea, calling for "culturally sustaining pedagogy" (p. 85) that supports students' home languages and cultures to promote democratic ideals. Still, despite a substantial body of literature discussing the facets of culturally relevant instruction, only a few studies have sought to estimate achievement effects on a large scale (Sleeter, 2012). To address the question causally, Dee and Penner (2016) undertook a regression discontinuity study of high school ethnic studies courses in San Francisco, finding large positive effects on attendance, grade point averages, and credit acquisition among ninth graders identified as academically at risk. Their work also builds on a large-scale study that linked exposure to Mexican American studies courses in Arizona high schools to higher graduation and exit examination pass rates, even after accounting for an extensive set of student background characteristics (Cabrera, Milem, Jaquette, & Marx, 2014).

Because our current study examines one-way and two-way program effects separately across many schools in Utah, it contributes toward disentangling language-access effects from cultural adjacency effects, where both program types provide the former, and two-way programs may provide the latter. We cannot definitively say that any differences in effects between one-way and two-way programs in Utah are attributable to cultural adjacency of the DLI programs, because other differences may exist in how the programs are run and taught. But with about 202,000 unique students in our analysis (a quarter in ever-two-way schools), we can comment on these differences in a way that prior studies have been less able to do because of design or sample size constraints.

Setting and Data

Setting and Policy Context

With the 2008 passage of Senate Bill 41, Utah became the first U.S. state to launch a DLI expansion initiative, followed by Delaware in 2011 and North Carolina in 2013 (Delaware Department of Education, 2011; North Carolina Department of Public Instruction, 2020). The current analysis stems from a federally funded research-practice partnership designed to identify insights from Utah's DLI scale-up.

Because kindergarten is optional in Utah, schools typically started new DLI programs with first grade and then added a grade each year (Utah State Board of Education, 2020). As noted, schools received $10,000 for each new grade they established, and an additional $5,000 per year in program maintenance thereafter. The funding was designed to incentivize DLI program launches across the state, but decisions about launching programs and allocating DLI slots were ultimately made by districts and schools. Most DLI districts reported to our team that they used a lottery process when DLI slots were oversubscribed, but because districts did not systematically track lottery applicants, we were unable leverage random assignment in our study design. Districts also varied in the extent to which they prioritized slots for students in a school's residential school zone.

Guided by promising practices in other localities (Lyster, 2007; Met, 1994), Utah employs a 50/50, two-teacher model for grades 1-6, meaning elementary school students spend 50% of their time in each language, switching teachers and languages midday. In grades 1-3, partner-language instruction focuses on math and social studies. In grades 4-5, it focuses on science and some math, and in grade 6, it focuses on science and social studies. Language arts in the partner language is taught in all grades but is emphasized in grades 4-6. As the programs expand into middle school, students typically take two classes per day in the partner language. In high school, Utah makes college-level coursework available in the partner language for students who pass an Advanced Placement exam in that language.

To promote high instructional consistency across DLI schools, Utah developed uniform curricula for DLI programs and provides common professional development to DLI teachers. Teachers are hired from local labor markets where possible, and through international guest worker programs as needed (Authors, 2016). One-way programs and two-way programs operate similarly, with common curriculum and teacher professional development opportunities. From a policy perspective, the key difference between them is in the primary language composition of the students they serve. This difference is of interest because it could affect the extent to which schools organize themselves around the needs of English learners and their families, and the extent to which the cultural heritages of students with primary languages other than English are honored within the school.

Analytic Sample

Our study uses an administrative dataset provided by the Utah State Board of Education that includes all public school students in the state of Utah in the academic years 2000-2001 through 2017-2018. We observe test score outcomes from grades 3 onward for 14 cohorts, including the kindergarten cohorts of 2000-2001 through 2014-2015. We exclude charter schools from the analysis because they were not part of the state's DLI scale-up policies. Because our analytic strategy compares cohorts of students who attended eventual DLI schools in years before and after the schools' launch of DLI programs, we focus on 146,739 unique students who attended ever-one-way DLI schools, and 55,826 students who attended ever-two-way DLI schools in grade 6 or below. We treat the students' first observed year in a Utah public school as his or her base year. The base year represents kindergarten for 59% of the sample, and first grade for 9%. Table 1 presents descriptive statistics for members of the analytic sample in their base observation year, as well as for 689,548 unique students who attended Utah public schools that never launched DLI programs.

<Insert Table 1 about here>

Table 1 shows that white students constituted 76% of students in never-DLI schools and 83% in ever one-way schools, but only 48% in ever two-way schools. Still, Hispanic or Latinx students constituted a substantial share of public school students in the state, representing 16% in never-DLI schools, 10% in ever-one-way schools, and 41% in ever-two-way schools. Students in ever two-way schools showed much higher rates of subsidized meal eligibility than their peers in ever one-way or never-DLI schools, at 58% versus 25% and 36%, respectively, and much higher rates of EL enrollments, at 38% versus 8% and 13%, respectively. Students who attended ever two-way schools also lived in moderately less-educated zip codes than those in ever one-way schools, and in neighborhoods with higher shares of Limited English Proficiency (5% vs. 2%) and eligibility for the federal Supplemental Nutrition Assistance Program (SNAP) (11% versus 7%).

Table 1 also shows the distribution of the ITT variable (i.e., DLI slots per first grader in the individual's first grade year and base school), as well as a dichotomous indicator of whether any slots were available. Even in years in which schools offered DLI programs, the average number of slots per first grader was about 0.52, and thus the average number of slots per first grader in Table 1 (0.22 for ever one-way and 0.21 for ever two-way) is just over half the fraction for whom slots were available. In addition, Table 1 shows the distribution of DLI languages among students who attended ever-DLI schools, illustrating that among students who attended ever one-way schools, about 76% attended schools that eventually offered Spanish or Mandarin Chinese, whereas French, Portuguese, and German programs accounted for a smaller share. Among ever two-way schools, all DLI programs were offered in Spanish.

Finally, Table 1 presents students' average test scores on state accountability tests across all observed grades. Utah administered the Utah Criterion Referenced Tests (CRTs) in ELA, math, and science through spring 2013. In 2014, it transitioned to the Student Assessment of Growth and Excellence (SAGE). To make the assessment scales consistent across years, we standardize all test scores to have a mean of 0 and SD of 1 within subject, grade, and year. We observe that students attending ever one-way schools performed about a tenth of an SD above the mean, whereas those at ever two-way schools performed between 0.29 and 0.34 SDs below the mean, on average, pooled across grades and years.

Analytic Methods

Our analysis begins with a purely descriptive comparison of academic performance between DLI and non-DLI students in a given year, net of a rich set of baseline characteristics, as shown in equation 1.

$$y_{icst} = \alpha_1 + \beta_1 DLI_{icst} + \boldsymbol{\lambda}_1{}'\mathbf{C}_\mathbf{c} + \boldsymbol{\delta}_1{}'\mathbf{S}_\mathbf{s} + \boldsymbol{\varphi}_1{}'\mathbf{X}_{\mathbf{ics}} + \boldsymbol{\eta}_1{}'\mathbf{K}_{\mathbf{cst}} + \varepsilon_{1icst} \quad (1)$$

Using ordinary least squares (OLS) regression, equation 1 estimates the relative performance, y_{icst}, of student *i* in cohort *c* from baseline school *s* at time *t*, as a function of whether the student is enrolled in DLI. The average difference in y_{icst} between DLI and non-DLI students in the same school and cohort is given by β_1, holding constant vectors of fixed effects for kindergarten cohort ($\boldsymbol{\lambda}_1$), initial school ($\boldsymbol{\delta}_1$), and baseline student characteristics $\mathbf{X}_{\mathbf{ics}}$, which include gender, race/ethnicity, subsidized meal eligibility at baseline, whether the student was ever classified as an EL, having a home language other than English (regardless of EL classification), special education status at baseline, and migrant status at baseline. Vector $\mathbf{X}_{ics}$ also includes controls for the resources in a student's baseline residential communities, including the share of residents with at least a bachelor's degree and with a professional degree, the share of limited English-speaking households, and the share of households receiving Supplemental Nutrition Assistance Program (SNAP) benefits. Vector $\mathbf{K}_{\mathbf{cst}}$ captures school-by-grade attributes of cohort *c* of baseline school *s* in time *t*, including the percent who are white, subsidized-meal eligible at baseline, ever classified as EL, and special education eligible at baseline. The dependent variable y_{icst} is a test score in ELA, math, or science for student *i* in year *t*, standardized statewide by subject, grade, and year to mean 0 and SD 1. The error term is given by ε_{1icst} . We cluster standard errors at the base school level. Note that we estimate equation 1 only within the final two academic years in the dataset, 2016-2017 and 2017-2018, because these are the only two years in which the DLI enrollment variable is available statewide.

Because virtually all schools in Utah that offer DLI programs also offer parallel non-DLI strands, students attending a DLI school must choose whether to take part in the DLI program at his or her school. This may happen through a choice process at the student's residentially zoned school or by the family's application to a DLI program, depending on district policy. Of course, ample evidence suggests that families who do and do not take advantage of DLI programs may differ in terms of both observed and unobserved attributes (Lindholm-Leary & Block, 2010; Marian et al., 2013; Steele et al., 2017). Relevant unobserved attributes may include families' education levels and values, perceptions of their children's motivations and aptitudes, and access to information about local dual-language options. To avoid confounding by unobserved differences between DLI and non-DLI students in the same schools and cohorts, we use the availability of DLI slots in the student's first-attended Utah public school in the student's first grade year as a measure of their baseline access to DLI. Utah schools launched DLI programs in different years and varied in the percentage of first grade enrollment slots dedicated to DLI. This fact lets us leverage variation in first graders' access to DLI within schools over time to estimate the ITT effect of a school's launch of DLI on subsequent achievement in the school, assuming other school characteristics remained constant—an assumption we discuss in greater detail below. Rather than treating the presence or absence of DLI in a school as dichotomous, we use the number of slots offered per first grader to estimate the linear effect of each per-pupil DLI slot. The ITT effect of first-grade slots per pupil for cohort *c* in base school s (SPP_{sc}) is thus estimated as follows:

$$y_{icst} = \alpha_2 + \beta_2 SPP_{sc} + \boldsymbol{\lambda}_2{}'\mathbf{C}_c + \boldsymbol{\delta}_2{}'\mathbf{S}_s + \boldsymbol{\varphi}_2{}'\mathbf{X}_{ics} + \boldsymbol{\eta}_2{}'\mathbf{K}_{ict} + \varepsilon_{2icst} \quad (2)$$

where coefficient β_2 is the average difference in y_{icst} associated with a school changing from offering 0% to 100% of its first-grade seats as DLI slots. In practice, the student-by-year mean of SPP_{sc} when non-zero ranged from 0.26 to 0.99, with a mean of 0.52, and with half of the observations falling between 0.44 and 0.58 slots per pupil. Thus, our estimates assume a linear effect of SPP_{sc}. We refer to SPP_{sc} as an ITT variable because it indicates students' access to DLI slots as a function of their base school and cohort year, but it does not indicate their take-up of such slots, since take-up within the cohort and school is much more likely to be affected by selection on unobservable attributes.

Other terms in the model are interpreted as in equation 1. Unlike equation 1, which pertains only to the 2016-17 and 2017-2018 academic years, equation 2 is estimated for academic years 2000-2001 onward. Districts for which historical DLI enrollments are available show SPP_{sc} to be quite consistent within schools over time, so we extrapolate backward based on ratios in the 2017-2018 academic year. Any measurement error in SPP_{sc} due to this choice would bias estimates toward 0.

To identify plausibly causal ITT effects on individual students with equation 2, we must assume that students who begin their educations in a given Utah school are, on average, the same from year to year, net of their observed baseline characteristics and the observed baseline attributes of their school-by-grade peers. If the population of students who enroll in a school systematically changes in response to the availability of a DLI program, and the changes on unobserved attributes are associated with the outcomes of interest, these unobserved attributes may bias our estimates of the slots-per-pupil effect. To test for changes in *observable* characteristics in response to DLI slots per pupil, we modify equation 2, sequentially regressing each school-by-grade characteristic in vector $\mathbf{K}_{ict}$ on the ITT slots-per-pupil measure, controlling for cohort fixed effects. The results yield the differences in observed school-by-grade student attributes associated with an increase in the DLI slots per first grader from 0% to 100%.

After examining potential school-by-cohort selection and then estimating ITT effects for one-way and two-way programs by grade as in equation 2, we disaggregate estimates by students' ever-EL classification. Among ELs, we disaggregate further between students whose primary language matches and does not match the partner language of the base school. In addition, we examine ITT effects on the EL status over time of students ever classified as EL.

Emulating the causal identification strategy of Anghel et al. (2016) in Spain, we then relax the assumption of homogeneous school composition over time in a robustness test that limits the treatment group only to the initial first-grade cohort eligible for DLI in the school. The rationale is that the first cohort (most of whom entered their base schools in kindergarten, before

DLI was offered) would have had the least time to anticipate or respond to the first-grade program launch.

If students who attend DLI programs in elementary school obtain access to higher-quality middle schools due to different middle school feeder patterns for DLI versus non-DLI students, the quality of the middle schools could mediate DLI program effects in grade 6. To test the strength of middle school transitions as a plausible mechanism, we conduct a robustness test examining whether estimates differ depending on the highest grade offered in the baseline elementary school (grade 6 for about two-thirds of students, and grade 5 for about a third).

The next part of our analysis aims to interpret differences in estimates for one-way and two-way programs. Because all two-way programs in Utah are Spanish programs, and because two-way programs serve less-affluent students than one-way programs on average, we try restricting one-way program estimates first to Spanish-language programs and then to schools that were demographically similar to two-way schools in the year before their DLI program launch. We estimate similarity in the pre-DLI year through a logistic regression model, in which the probability that a student who enrolled at baseline in an ever-DLI school was attending an ever two-way school is predicted as a function of the share of grade-level peers in the student's base year who were white, subsidized-meal eligible, EL, or special education eligible. These fitted probabilities, p_{ics}, are then extrapolated to be constant within student. Using the predicted probabilities, and following Austin (2011), we calculate average treatment on the treated (ATT) weights as in equation 3:

$$w_{ics.ATT} = Z_s + \frac{(1 - Z_s)\, p_{ics}}{1 - p_{ics}} \tag{3}$$

where Z_s is dichotomously coded as 1 if the student's baseline ever-DLI school became a two-way school and 0 if it became a one-way school, and where the fitted probability of its becoming a two-way school is represented as p_{ics}. We then re-estimate the ITT model for one-way schools with the ATT weights so that the estimates apply to one-way schools that were demographically similar to two-way schools in the year before they launched DLI.

To test whether linguistic alignment of the student body with the partner language moderates estimated ITT effects, we then interact the SPP_{sc} indicator with a measure of the fraction of students, m_s, in an ever-DLI school whose primary language matched that of the school's soon-to-be DLI partner language in the year before DLI launch:

$$y_{icst} = \alpha_4 + \beta_4 SPP_{sc} + \chi_4 (SPP * m)_{sc} + \boldsymbol{\lambda}_4{}'\mathbf{C}_c + \boldsymbol{\delta}_4{}'\mathbf{S}_s + \boldsymbol{\varphi}_4{}'\mathbf{X}_{ics} + \boldsymbol{\eta}_4{}'\mathbf{K}_{ict} + \varepsilon_{4icst} \quad (4)$$

In fitting equation 4, we estimate not only the main effect, β_4, of first-grade slots per pupil, but also its differential effect, χ_4, for each unit difference (from 0 to 1) in the fraction of students whose home language matches the DLI partner language in the last pre-DLI year. If the effect of SPP differs by m_s, this suggests that cultural adjacency between students' school and home languages may shape the effects of DLI access on student learning.

Finally, we run a placebo test to examine whether ITT and interaction parameters β_4 and χ_4 in equation 4 would differ if we redefined the DLI launch year as the year *before* the program actually launched in each school. This model is estimated only within the true pre-DLI cohorts in each school, so that the last pre-launch cohort is defined as the placebo treatment cohort. Finding similar effects in the placebo and true models may suggest endogeneity in schools' decision to offer DLI programs, implying that pre-existing trends in these schools, rather than the conversion to DLI, were responsible for any estimated DLI effects.

Results

Descriptive Within-Cohort Estimates

We begin with descriptive results from the within-cohort regression model described in equation 1. As noted, these estimates pertain only to 2016-2017 and 2017-2018 data because those are the years in which reliable DLI enrollment estimates are available statewide. Within the same cohorts of students who began in the same base schools, students enrolled in DLI classrooms outperformed those in non-DLI classrooms by a large margin, as shown in Table 2, where estimates appear on the left for one-way programs and on the right for two-way programs. Coefficients are reported in state-by-grade-by-year SD units and are shown separately for grades 3 to 6 in ELA, math, and science. Across grades levels and subject areas, the outperformance in one-way schools was roughly a quarter of an SD, and in two-way schools, it was even larger, ranging from a third to a half of an SD across subjects, with the largest relative performance across subjects in grade 6. All estimates in Table 2 are statistically significant at the 0.001 level. Thus, even after adjusting for observed student demographic characteristics at baseline and for peer-by-grade demographic characteristics, DLI students were notably outperforming their peers from the same base schools.[4]

<Insert Table 2 about here>

[4] Inclusion of individual covariates as shown here reduces unadjusted one-way DLI participation coefficients, which aren't shown, by about a quarter, but it increases unadjusted two-way coefficients by about a fifth, suggesting, as further discussed below, that individual selection patterns may be positive for one-way programs and negative for two-way programs.

When estimates are disaggregated between primary English speakers and ELs in the top and bottom rows of Figure 1 and in the three corresponding panels of Appendix Table A1, we find similar patterns. Estimates for primary English speakers in the top row of Figure 1 strongly track those in Table 2. Estimates for English learners are further disaggregated between students whose primary languages did and did not match the primary languages of their schools, with circles in Figure 1 representing the former, and triangles the latter. For those with language matches, results are similar in magnitude and statistical significance (denoted with solid markers) as for primary English speakers. For students without language matches, they are even larger, though selection on unobservable attributes is an even larger concern for this group, whose families have chosen immersion in a language other than English and the primary language.

<Insert Figure 1 about here>

For Utah families with children in DLI programs, these descriptive within-cohort results are useful, because they show that DLI students perform very well on standards-based tests administered in English, even though they spend half their school hours studying core content in the partner language. Of course, from a policy perspective, we cannot attribute this outperformance strictly to the DLI programs themselves, since DLI and non-DLI families in the same base schools and cohorts may differ in their academic preparedness, values, and awareness in ways that our control variables do not fully capture.

Between-Cohort Selection into DLI Schools

In Table 3, we turn to a selection question that underpins our ITT design: to what extent do schools' compositions change when they launch DLI programs? We first place this question in context by noting that students' migrations out of their initial schools are relatively (and similarly) common in ever one-way, ever two-way, and never DLI schools in the state. As shown in Appendix Figure A1, about 6-15% of students migrate away from their base school each year through grade 4, and these rates increase modestly by grades 5 and 6, as some systems begin transitioning students into middle school.

We now consider whether there is evidence that migration is systematic in response to DLI program launches. To do so, we regress the school-by-grade characteristics of ever-DLI schools on first grade slots per pupil in students' first grade year. Coefficients in Table 3 represent the average difference in the given peer-by-grade fraction (ranging from 0 to 1) resulting from converting all first-grade enrollment slots from regular to DLI slots. Our estimates are disaggregated for one-way programs (the left side of Table 3) and two-way programs.

<Insert Table 3 about here>

In fact, we do find evidence of systematic differences in peer-by-grade characteristics associated with DLI program availability, and the direction of change differs importantly between one-way and two-way programs. In ever one-way schools, moving from 0 to 100 DLI slots per pupil is linked to a two percentage-point increase in the share of white students. It is also linked to decreases in subsidized meal eligibility rates of about 7 percentage points, of EL enrollment rates by about 3-5 percentage points, and of the share of special education students by about 3 percentage points, alongside an increase of about 1-2 in the number of zip codes represented in the school. In other words, peer groups in one-way schools become somewhat more white, affluent, primary English-dominant, and general education-dominant as DLI programs are opened. In two-way programs, we see mostly the opposite patterns. Changing from offering 0 to 100 DLI slots per first grader in two-way schools reduces the share of white students by 13 to 18 percentage points and raises the share of students qualifying for subsidized meals by 7 to 15 percentage points, though we still see decreases in special education rates of about 3 percentage points. We see few statistically significant changes in the share of ELs and number of zip codes, perhaps because two-way programs were intentionally placed in schools that already served substantial numbers of EL students. The non-significant zip code estimates in most years may suggest that families were less likely to transfer across zip codes into two-way programs than into one-way programs.

Thus, there is evidence that grade-by-year demographic attributes did change in association with the opening of DLI programs. To minimize between-cohort selection effects, our ITT analysis includes controls not only for individual baseline attributes but also for the base school grade-by-year peer attributes in Table 3, anticipating that the grade-by-year controls will largely absorb unobserved between-cohort differences associated with them (Altonji, Elder, & Taber, 2005; Shadish, Clark, & Steiner, 2008). To strengthen inferences of plausible causality, we subject our ITT estimates to robustness and placebo tests in the ensuing subsections.

Intent-to-Treat Estimates of DLI Access Effects

Table 4 summarizes key ITT estimates in ELA, reading, and mathematics for grades 3 through 6, where the ITT variable is first grade slots per pupil in the student's first grade year. Here, results are far more muted than in the preceding analyses that focuses on within-cohort differences.[5] In one-way-programs, ITT estimates trend negative but are generally indistinguishable from 0, save for a negative estimate of 0.063 SD in fourth grade math ($p<0.1$) and of -0.12 SD in sixth grade science ($p<0.05$). In two-way programs, on the other hand, we find some indication of gradual benefits over time. Estimates trend positive in ELA and math and are statistically significant in ELA in grades 3 and 6, with magnitudes of 0.092 and 0.362 SD,

[5] Estimates in Table 4 include all sample years, unlike those in Table 2, which include only 2016-17 and 2017-18. Still, estimates in Table 4 remain substantively similar when the slots-per-pupil analysis is restricted to only those final two years of data.

respectively, and in math and science in grade 6, with respective magnitudes of 0.388 and 0.311 SD. Top-row estimates show pooled cross-grade effects. For one-way programs, these are null in ELA and math and -0.049 SD in science. For two-way programs, they are 0.114 and 0.102 SD in ELA and math, respectively, and null in science. Given that these estimates pertain to a unit difference in slots per pupil (from 0 to 1), and that the average first-grade slots per pupil in DLI schools and years was 0.52 rather than 1, the effects for a typical DLI school can be approximated by multiplying the coefficients by 0.52, or (roughly) dividing by 2.

<Insert Table 4 about here>

When we disaggregate these estimates for primary English speakers and ELs in Figure 2 and Appendix Table A2, we find statistically significant benefits of two-way programs for both language-match ELs and primary speakers in grade 6, and a statistically significant negative effect for language-match ELs in ELA in grade 6 in one-way programs. Here it is worth bearing in mind that non-language match ELs may under-enroll in DLI programs when they become available, and this would lead their subgroup estimates to be null in ITT models.

<Insert Figure 2 about here>

Also focusing on ELs, Table 5 presents estimated slots-per-pupil effects on the probability that a student ever classified as EL in Utah public schools remains classified as such in grades 1 through 6. For one-way programs, we find no statistically significant differences in rates of EL classification. In other words, ELs appear to be reclassified at similar rates before and after the launch of one-way DLI programs in their base schools, regardless of whether their primary language matches the partner language. For two-way programs, we find no differences in EL classification rates for ELs whose primary languages *do not* match the partner language. For those whose home languages *do* match, we find similar rates of EL status persistence until grades 5 and 6, at which time students ever classified as EL have lower rates of EL classification by 8 and 10 percentage points, respectively. The finding that language-aligned DLI access raises EL reclassification rates after several years of exposure is consistent with Steele et al. (2017), who find the same starting in grade 6, and Umansky and Reardon (2014), who find it from grade 7 onward.

<Insert Table 5 about here>

Robustness Tests

As noted, the possibility of unobservable sorting across schools in response to the opening of DLI programs remains a threat to causal inference. Expecting that children in the first DLI eligible cohorts in each school would have had less time to respond to a newly launched program, we run a robustness test in which, like Anghel et al. (2016) in Spain, we limit the ITT group to just the first DLI-eligible treatment group in each school. In Utah, more than half of these students were already enrolled in kindergarten in their base schools in the year before their DLI programs launched, meaning they would have needed foreknowledge of program launches to sort systematically. Results of this test, shown in Table 6, are similar to those in the main analysis in Table 4. This suggests that our main results are not driven by increased sorting over time as families acclimated to the presence of DLI programs, and that such sorting on unobservable characteristics would have had to occur with little lead time.

<Insert Table 6 about here>

To address whether results are driven by access to higher-quality middle schools for students whose elementary schools end before grade 6, Table 7 presents results of the ITT analyses for students whose base elementary school continued through grade 6 or higher (left panel), and for those whose base elementary school offered no grade higher than 5 (right panel).

About two-thirds of students had base schools that continued through grade 6, and results for these students are robust, with continued evidence of positive sixth-grade effects for two-way programs, implying that transitions to middle school are not driving the main results. Estimates for students from two-way base schools that ended at grade 5 or below show greater heterogeneity, with statistically significant estimates that are negative in grades 4 and 5 and positive in grade 6. The question underlying Table 7 is whether grade 6 estimates in the main analysis appear driven by middle school transitions, and the evidence across columns suggests that they are not.

<Insert Table 7 about here>

Interpreting One-Way versus Two-way Program Estimates

In Utah, both one-way and two-way programs use the same 50/50 instructional model. Programs also receive common dual-language curriculum and teacher professional development. From that perspective, we would expect similar average achievement effects in both program types, but the estimates do appear to differ, trending from null to modestly negative in one-way programs, and from null to substantially positive in two-way programs. Estimates in Table 8 consider two possible explanations. One is that all two-way programs are Spanish programs, whereas one-way programs comprise Spanish, Mandarin, French, and German. Because Spanish is arguably the most phonetically accessible and English-adjacent of the partner languages, it is possible that the two-way effects are actually Spanish effects. In the left panel of Table 8, we estimate ITT effects for one-way programs in Spanish only. However, we continue to find one-way estimates that are null in ELA and math and modestly negative in science in some grades.

<Insert Table 8 about here>

Another possible difference lies in the demographics of one-way and two-way schools. As we have seen, one-way schools in Utah are more affluent and white than their two-way counterparts. To address whether program differences are driven by the baseline demographic attributes of the schools, we weight the one-way programs by their similarity to two-way schools in their pre-DLI opening year, in terms of the percentage of students who are white, subsidized-meal eligible, ever-ELs, and special-education eligible. With this weighting, which improves balance on pre-DLI school characteristics by up to two-thirds, we find one-way program estimates that are modestly negative and statistically significant in several grades in ELA, math, and science. This suggests that the differences between two-way and one-way estimates in the main analysis are not driven by the relative socioeconomic advantages of one-way and two-way schools.

<Insert Table 9 about here>

Finally, in the left panel of Table 9, we consider the extent to which the ITT effect of first grade DLI slots per pupil differs by the fraction of students in the school whose primary languages matches the partner language. The analysis includes all schools that eventually opened one-way or two-way programs, and it defines the language match in pre-treatment years based on the language that eventually became the partner language in the school. We interact the slots-per-pupil ITT variable with the fraction of language-match students in the school in the pre-DLI year, which ranges from 0 to 0.83. On the left of Table 9, we report the main effect of the slots per pupil variable (the ITT coefficient when the language-match fraction is 0), and the interaction effect, which is interpreted as the additional, differential effect of slots per pupil for each unit difference (here, from 0 to 1) in the fraction of language-match students in the last pre-DLI year.

In the left panel of Table 9, we find null-to-negative main effects but positive and significant interaction effects. The fitted cross-grade predictions for students in base schools with no primary speakers of the partner language are effectively 0 for ELA and math and -0.07 SD for science, but for schools (hypothetically) with 100% primary speakers of the partner language, they are as high as 0.298 SD (the sum of the main effect and interaction term) for ELA and 0.343 SD for math. For an average two-way school, in which only about 0.34 of students were primary speakers of the partner language, the corresponding effect would be 0.099 SD in ELA and 0.129 SD in math. The cross-grade interaction term for science does not reach statistical significance. In other words, ITT effects on ELA and math rise substantially with the share of partner-language speakers in the school.

Testing Endogenous School Selection into DLI

To test further for school-level and student-level bias due to systematic selection into DLI, we refit the language-match interaction model after redefining the DLI launch year as the year *before* the DLI program actually launched. Results of this placebo test—essentially a test for parallel near-term trends pre-DLI—are shown in the right panel of Table 9. Similar to the first-cohort test in Table 6, this placebo test focuses on the performance of the first DLI cohort in a given base school relative to the performance of the preceding cohorts. The purpose of the test is to examine whether other time-varying attributes of the treated schools, such as pre-existing trends in leadership, teaching quality, or changes in unobserved student attributes may account for the DLI estimates in the main analysis. These are uniformly null in terms of both the main effects and the interaction coefficients. This suggests that our main results in Table 4 and those on the left side of Table 9 are capturing effects of the launch of DLI programs and not of pre-existing secular trends due to schools or families systematically opting into DLI conversions.

Discussion and Conclusion

As demand grows for public school programs that are both culturally inclusive and academically challenging, DLI programs show clear appeal. Demand for these programs is growing, with lotteries and wait lists in many localities, raising concerns about gentrification and the crowding out of students whose primary languages match the schools' partner languages (Lam & Richards, 2020; Williams, 2017). This study adds to the growing research on DLI programs by examining the effects of DLI program launches on schools' subsequent achievement across a large state scale-up effort. Though Utah's DLI students spent half of their elementary instructional hours learning in a language other than English, we find mainly null-to-modest ITT effects, positive or negative, on schools' achievement in core content areas tested in English. In one-way programs, cross-grade ITT estimates are null in ELA and math but modestly negative in science, with a negative science estimate in grade 6. In two-way programs, cross-grade effects are modestly positive in ELA and math and null in science, with large positive effects in all three subjects in grade 6, and results are robust to numerous sensitivity tests. Also, by grade 5, ever-ELs with access to two-way DLI programs that matched their primary languages were reclassified as English proficient at rates up to 8 points higher than those without access. Reclassification rates for ELs with access to one-way language-match programs or non-matching DLI programs did not differ from those of their peers without access.

A clear limitation of the study is that we do not have data on students' residentially zoned schools. We must treat the schools in which they first enrolled as their base schools, meaning our estimates could capture sorting on unobserved family characteristics over time. Still, there are a few reasons to believe that our estimates are not strongly affected by unobservable between-cohort selection. First, the changes we observe in grade-by-year peer characteristics linked to

DLI slots per pupil in Table 3 fall in opposite directions of the estimated slots-per-pupil effects, with evidence of more-advantaged students selecting into one-way programs and of less-advantaged students entering two-way programs. For selection on unobservable characteristics to inflate our ITT estimates*, we would have to assume that selection on unobservable characteristics fell in the opposite direction as selection on observable characteristics*. The notion that selection bias on unobserved and observed characteristics falls in opposite directions seems unlikely since our controls include a host of individual and aggregate sociodemographic factors that are well-known to be key predictors of student achievement (Altonji et al., 2005; Shadish et al., 2008), and we would therefore anticipate that attributes that are not fully captured by these observed variables would operate a similar direction. Second, focusing on within-school changes in DLI access over time yields much more conservative estimates than comparing DLI and non-DLI students within schools. This fact comports with our *a priori* expectation that selection on unobservable attributes is of much greater concern between same-cohort, same-school students than between DLI-eligible and ineligible cohorts entering a given school over time. Third, and very importantly, our estimates are similar if we narrow them to the first DLI-eligible cohorts in each school, including mainly kindergarteners, whose families would have had difficulty anticipating DLI launches. Focusing on the initial cohort is the causal identification strategy adopted by Anghel et al. (2016) in Spain, so the robustness of our findings to that approach offers reassurance. Fourth, our placebo test redefining the first treatment year as the year before the true launch year shows no effects, suggesting that our DLI estimates are not capturing other recent changes—organizational or demographic, including anticipatory residential migration—in schools that soon opened DLI programs. For these reasons, we adopt the phrasing of Valentino and Reardon (2015) in classifying our findings as likely to be "largely, but not completely, unbiased" (p. 633) by selection on unobservable characteristics.

In interpreting the results, it is important to bear in mind that they are not the effects of *an individual's participation in a DLI program over time*, but of the expected achievement of a student over time if her base school offered only DLI slots when she was in first grade, as opposed to the scenario in which it did not offer any DLI slots. Because the average share of slots per pupil in DLI schools, when not 0, was 0.52, such estimates *overstate* the average program launch effect in a given school, but they retain a consistent definition, assuming that any effects of slots per pupil are roughly linear. Also, because these are intent-to-treat estimates, they may simultaneously *understate* the effects of DLI access on slots-per-pupil compliers—that is, students who persist over time in the base schools to which they were zoned when they entered Utah public schools. Still, because they are intent-to-treat effects, they are also policy-relevant, as they examine what happens to student achievement in a real-world context where some school transfers over time are typical, as referenced in Appendix Figure A1.

It is worth acknowledging that our dependent variables are not the only goals of DLI programs in Utah or elsewhere. The state's stated intention in rapidly scaling DLI was to prepare a bilingual and biliterate workforce. Because students not enrolled in DLI were not tested in bilingualism or biliteracy, our analysis focuses on the effects of program launches on students' achievement in core content tested in English. Fortunately, given that the study is part of a broader research-practice partnership, we can interpret these estimates alongside companion research in Utah. Specifically, Authors (2018) found that Utah students in Chinese, French, and Spanish DLI programs were meeting or exceeding partner-language performance benchmarks in grades 3, 6, and 8, with average eighth-grade skill attainment of Intermediate Mid-to-High in

Spanish and French and Intermediate Low in Chinese. These levels already exceed what would be expected in traditional secondary school language electives (Burkhauser et al., 2016; Xu, Padilla, & Silva, 2015). In a follow-up study, the team found that well over 80% of ninth graders reached all four of the state's proficiency benchmarks in Spanish and French, and over 60% achieved listening and reading benchmarks in Chinese (Authors, 2021). In other words, Utah DLI students appear to meet the state's goals of moving students toward bilingualism and biliteracy. Given this progress, future work should examine ITT effects on AP language credit completion, high school graduation, postsecondary attainment, and even labor market outcomes.

Our study finds clearer evidence of academic benefits from two-way than from one-way program access, and this is true for language-matched ELs as well as for primary English speakers. In contrast, there is little evidence that ELs benefitted from or were harmed by access to language-matched one-way programs in terms of academic achievement, despite spending only half of their instructional time in English, which is the language of the academic accountability tests. (They may also have benefitted substantially terms of partner language or cultural skills, which this article does not examine.) Still, we find that estimated effect of DLI access increased linearly as students' primary languages aligned with the partner languages of their schools. These findings comport with quantitative evidence about the academic benefits of culturally relevant instruction (Cabrera et al., 2014; Dee & Penner, 2016) and suggest a need to better understand language and cultural practices in these schools. For example, our finding that overall effects are correlated with the share of primary partner-language speakers in the school suggests that benefits may exist for more than just ELs. Schools that offer two-way DLI may be more responsive to the needs of language minority students and families in general, creating a more culturally and linguistically sustaining environment. Of course, from a policy perspective, two-way programs depend on a critical mass of students who share a common, non-English language, and they may not be feasible in communities that serve students from diverse language backgrounds or from mostly English-speaking backgrounds. Thus, future research should examine how Utah’s two-way and one-way programs may differ, and the extent to which effects are correlated with differences in school cultural norms, parent communication practices, racial/ethnic alignment of teachers and students, and other factors. In the interim, our study may be seen as reflecting the entwined nature of language and culture, and the complex ways in which they may reinforce one another.

Bibliography

Altonji, Joseph G., Elder, Todd E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy, 113*(1), 151-184. doi: 10.1086/426036

Anghel, B., Cabrales, A., & Carro, J. M. (2016). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning. *Economic Inquiry*, *54*(2), 1202-1223. doi: 10.1111/ecin.12305

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424. doi: 10.1080/00273171.2011.568786

Authors. (2016). Academic achievement of students in dual language immersion. *International Journal of Bilingual Education and Bilingualism, 21*(8), 913-928. doi: 10.1080/13670050.2016.1214675

Authors. (2018). Linguistic performance of dual language immersion students. *Foreign Language Annals, 51*(3), 575-595. doi: 10.1111/flan.12354

Authors. (2021). Sustaining dual language immersion: Partner language outcomes in a state-wide program. *Modern Language Journal, 105*(1), 194-217. doi: https://doi.org/10.1111/modl.12694

Barik, H. C., & Swain, M. (1978). Evalutation of a French immersion program: the Ottawa study through grade five. *Canadian Journal of Behavioural Science, 10*(3), 192-201. doi: 10.1037/h0081548

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Science, 65*(4), 229-235. doi: 10.1037/a0025406

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, *19*(1), 12-23. doi: 10.1177/0963721409358571

Bibler, A. (2020). Dual language education and student achievement. *Education Finance and Policy,* (online first). doi: https://doi.org/10.1162/edfp_a_00320

Burkhauser, S., Steele, J. L., Li, J., Slater, R. O., Bacon, M., & Miller, T. (2016). Partner-language learning trajectories in dual-language immersion: Evidence from an urban district. *Foreign Language Annals, 49*(3), 415-433.

Cabrera, N. L., Milem, J. F., Jaquette, O., & Marx, R. W. (2014). Missing the (student achievement) forest for all the (political) trees: Empiricism and the Mexican American studies controversy in Tucson. *American Educational Research Journal, 51*(6), 1084-1118.

Caldas, S. J., & Boudreaux, N. (1999). Poverty, race, and foreign language immersion: Predictors of math and English language arts performance. *Learning Languages, 5*, 4-15.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism, 7,* 1-87.

Center for Applied Linguistics. (2011a). Directory of foreign language immersion programs in U.S. schools. Retrieved from http://www.cal.org/resources/immersion/

Center for Applied Linguistics. (2011b). Directory of two-way bilingual immersion programs in the U.S. . Retrieved from www.cal.org/twi/directory/

Christofides, L. N., & Swidinsky, R. (2010). *The economic returns to a second official language: English in Quebec and French in the Rest-of-Canada* (2010-04). Nicosia, Cyprus: Institute for the Study of Labor. Retrieved from https://core.ac.uk/download/files/153/6625815.pdf

Collier, V. P., & Thomas, W. P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and Practice, 2*(1), 1-20.

Commission on Language Learning. (2017). *America's languages: Investing in langauge education for the 21st century*. Cambridge, MA: American Academy of Arts & Sciences.

Committee for Economic Development. (2006). *Education for global leadership: The importance of international studies and foreign language education for U.S. economic and national security*. Washington, DC: Committee for Economic Development.

De Jong, E. (2004). L2 proficiency development in a two-way and a developmental bilingual program. *NABE Journal of Research and Practice, 2*(1), 77-108.

Dee, T. S., & Penner, E. K. (2016). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal, 54*(1), 127-166. doi: 10.3102/0002831216677002

Delaware Department of Education. (2011). Dual language immersion education in Delaware. Retrieved from https://www.doe.k12.de.us/immersion

Fabián Romero, E. (2017, October 25). With new research, policy shifts, bilingual education on rise. *Education Writer's Association Blog: Latino Ed Beat*. Retrieved from https://www.ewa.org/blog-latino-ed-beat/new-research-policy-shifts-bilingual-education-rise

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2016). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science, 26*(7), 1090-1097. doi: 10.1177/0956797615574699

Fortune, T. W. (2012). What the research says about immersion. In *Chinese language learning in the early grades: A handbook of resources and best practices for Mandarin immersion* (pp. 9-14). New York: Asia Society.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology, 93*(3), 265-281.

Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive development, 28*(1), 41-50. doi: 10.1016/j.cogdev.2012.10.002

Kamenetz, A. (2016, November 29). 6 potential brain benefits of bilingual education. *National Public Radio*. Retrieved from http://ripr.org/post/6-potential-brain-benefits-bilingual-education

Keshavarz, M. H., & Astaneh, H. (2004). The impact of bilinguality on the learning of English vocabulary as a foreign language (L3). *Bilingual Education and Bilingualism, 7,* 295-302.

Ladson-Billings, G. (1992). Liberatory consequences of literacy: A case of culturally relevant instruction for African American students. *The Journal of Negro Education, 61*(3), 378-391. doi: 10.2307/2295255

Lam, K., & Richards, E. (2020). More US schools teach in English and Spanish, but not enough to help Latino kids. *USA Today*. Retrieved from https://www.usatoday.com/in-depth/news/education/2020/01/06/english-language-learners-benefit-from-dual-language-immersion-bilingual-education/4058632002/

Lambert, W. E., Genesee, F., Holobow, N., & Chartrand, L. (1993). Bilingual education for majority English-speaking children. *European Journal of Psychology of Education, 8*(1), 3-22. doi: 10.1007/BF03172860

Lambert, W. E., Tucker, G. R., & d'Anglejan, A. (1973). Cognitive and attitudinal consequences of bilingual schooling: The St. Lambert Project through grade five. *Journal of Educational Psychology*, *65*(2), 141-159.

Lapkin, S., Hart, D., & Turnbull, M. (2003). Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: Building explanations. *Alberta Journal of Educational Research, 49*(1), 6-23.

Lindholm-Leary, K. J., & 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

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamins.

Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal*, *36*, 167-186. doi: 10.1080/15235882.2013.818075

Maxwell, L. A. (2014, October 15). Successes spur push for dual-language classes. *Education Week*, pp. 14-15. Retrieved from https://www.edweek.org/ew/articles/2014/10/15/08dual.h34.html

Met, M. (1994). Teaching content through a second language. In F. Genesee (Ed.), *Educating second language children: The whole child, the whole curriculum, the whole community* (pp. 159-182). New York: Cambridge University Press.

Moll, L. C., & González, N. (1994). Lessons from research with language-minority children. *Journal of Reading Behavior, 26*(4), 439-456. doi: 10.1080/10862969409547862

Mora, J. K. (2009). From the ballot box to the classroom. *Educational Leadership*, *66*(7), 14-19.

North Carolina Department of Public Instruction. (2020). Dual language immersion. Retrieved from https://www.dpi.nc.gov/districts-schools/classroom-resources/k-12-standards-curriculum-and-instruction/programs-and-initiatives/dual-language-immersion

Padilla, A. M., Lindholm, K. J., Chen, A., Durán, R., Hakuta, K., Lambert, W. E., & Tucker, G. R. (1991). *The English-only movement--Myths, reality, and implications for psychology*. Washington, DC: American Psychological Association.

Paris, D., & Alim, H. S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review, 84*(1), 85-100. doi: 10.17763/HAER.84.1.982L873K2HT16M77

Potowski, K. (2004). Student Spanish use and investment in a dual immersion classroom: Implications for second language acquisition and heritage language maintenance *The Modern Language Journal*, *88*(1), 75-101. doi: 10.1111/j.0026-7902.2004.00219.x

Saiz, A., & Zoido, E. (2005). Listening to what the world says: Bilingualism and earnings in the United States. *Review of Economics and Statistics, 87*(3), 523-538.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. *Journal of the American Statistical Association, 103*(484), 1334-1356.

Sleeter, C. E. (2012). Confronting the marginalization of culturally responsive pedagogy. *Urban Education, 47*(3), 562-584. doi: 10.1177/0042085911431472

Steele, J. L., Slater, R. O., Li, J., Zamarro, G., Miller, T., & Bacon, M. (2018). Dual-language immersion education at scale: An analysis of program costs, mechanisms, and moderators. *Educational Evaluation and Policy Analysis, 40*(3), 420-445. doi: https://doi.org/10.3102/0162373718779457

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*(1), 282S-306S. doi: 10.3102/0002831216634463

Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal, 51*(5), 879-912. doi: 10.3102/0002831214545110

UNESCO. (2016a). *Education 2030: Incheon declaration and framework for action for the implementation of Sustainable Development Goal 4*. New York: United Nations Educational, Scientific, and Cultural Organization.

UNESCO. (2016b). *If you don't understand, how can you learn? (Global Education Monitoring Report, Policy Paper 24)*. New York: United Nations Educational, Scientific, and Cultural Organization.

Utah Senate. (2016). Dual language immersion: Origin story. Retrieved from http://senatesite.com/utahsenate/dual-language-immersion/

Utah State Board of Education. (2020). Welcome to kindergarten: Frequently asked questions. Retrieved from https://www.schools.utah.gov/curr/kindergarten?mid=1179&tid=3

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English Learners: Variations by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis, 37*(4), 612-637. doi: 10.3102/0162373715573310

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, Version 4.1*. Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Williams, C. (2017, December 28). The intrusion of white families into bilingual schools. *The Atlantic*.

Wooldridge, J. M. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal, 1*(2), 117-139. doi: 10.1007/s10258-002-0008-x

Xu, X., Padilla, A. M., & Silva, D. M. (2015). Learner performance in Mandarin immersion and high school world language programs: A comparison. *Foreign Language Annals, 48*(1), 26-38. doi: 10.1111/flan.12123

Table 1. Characteristics of students in their first observed year, by base school category

	Ever One-Way	Ever Two-Way	Never DLI	*Pooled SD*
N Unique Students	***145,739***	***55,826***	***689,548***	***891,113***
Individual Characteristics				
Female	0.49	0.49	0.49	*0.50*
Asian	0.03	0.06	0.03	*0.18*
Black	0.01	0.02	0.02	*0.13*
Hispanic	0.10	0.41	0.16	*0.37*
American Indian	0.01	0.01	0.02	*0.12*
White	0.83	0.48	0.76	*0.43*
Race Other/Missing	0.01	0.01	0.01	*0.12*
Base Free/Red. Lunch	0.25	0.58	0.36	*0.48*
Primary Not English	0.09	0.40	0.14	*0.35*
Ever EL	0.08	0.38	0.13	*0.35*
Primary/Partner Lang. Match	0.03	0.34	.	*0.18*
Base Special Education	0.09	0.09	0.11	*0.31*
Ever Migrant	0.00	0.01	0.00	*0.07*
Residential Zip Code Characteristics				
Pct. Bach. Deg.	34.81	26.80	29.62	*12.35*
Pct. Grad. Deg.	11.75	8.66	9.46	*5.68*
Pct. Limited Eng. Prof.	1.59	5.17	2.33	*2.63*
Pct. SNAP	6.71	10.64	8.54	*4.54*
Peer Attributes in Base School and Grade				
Pct. White	0.85	0.49	0.78	*0.21*
Pct. Free/Red. Lunch	0.25	0.58	0.35	*0.24*
Pct. Base EL	0.04	0.25	0.07	*0.14*
Pct. Base Special Ed.	0.09	0.09	0.11	*0.08*
DLI Access				
Slots available in gr. 1 (y/n)	0.42	0.41	0.00	*0.29*
Slots per first grader in gr. 1	0.22	0.21	0.00	*0.15*
Base School DLI Language				
Spanish	0.37	1.00	.	*0.50*
Chinese	0.39	0.00	.	*0.45*
French	0.13	0.00	.	*0.29*
Portuguese	0.09	0.00		*0.24*
German	0.02	0.00	.	*0.13*
N with ELA Scores	***111,643***	***42,784***	***539,488***	***693,915***
Mean Scores Across Obs. Years (Standardized within Subject, Grade, Year)				
ELA	0.09	-0.30	-0.04	*0.92*
Math	0.11	-0.29	-0.03	*0.92*
Science	0.11	-0.34	-0.04	*0.91*

Table 2. Descriptive within-cohort estimates disaggregated by one-way versus two-way programs

	One-way			Two-way		
	ELA	Math	Science	ELA	Math	Science
Grade	(1)	(2)	(3)	(4)	(5)	(6)
All	0.252***	0.227***	0.246***	0.354***	0.389***	0.347***
	(0.017)	(0.017)	(0.019)	(0.044)	(0.052)	(0.059)
3	0.270***	0.222***		0.383***	0.387***	
	(0.024)	(0.024)		(0.052)	(0.057)	
4	0.246***	0.209***	0.233***	0.339***	0.398***	0.355***
	(0.024)	(0.025)	(0.028)	(0.063)	(0.074)	(0.069)
5	0.291***	0.266***	0.290***	0.381***	0.444***	0.383***
	(0.024)	(0.028)	(0.027)	(0.050)	(0.064)	(0.075)
6	0.248***	0.265***	0.246***	0.438***	0.517***	0.428***
	(0.027)	(0.026)	(0.029)	(0.075)	(0.076)	(0.073)
Schools base gr.	494	494	497	360	359	369
Obs. base gr.	16,194	16,174	16,308	5,395	5,390	5,261
R-sq base gr.	0.101	0.095	0.111	0.189	0.167	0.181
Schools gr. 6	539	539	539	292	290	290
Obs. gr. 6	20,371	20,289	20,432	2,957	2,955	2,959
R-sq gr. 6	0.127	0.107	0.098	0.154	0.140	0.147

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level. Within-cohort estimates pertain only to 2016-17 and 2017-18, for which clean DLI enrollment data are available statewide.

Table 3. Selection test: Regressing school-by-grade characteristics on first grade DLI slots-per-pupil

	One-way					Two-way				
Grade	Fract. White	Fract. FRL	Fract. EL at Base	Fract. Sped at Base	N Zip Codes	Fract. White	Fract. FRL	Fract. EL at Base	Fract. Sped at Base	N Zip Codes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.024*	-0.090***	-0.030**	-0.018**	1.651**	-0.131**	0.071*	0.085	-0.040***	0.799
	(0.010)	(0.017)	(0.010)	(0.006)	(0.600)	(0.040)	(0.033)	(0.051)	(0.008)	(0.758)
2	0.020*	-0.097***	-0.025**	-0.025***	1.857**	-0.160**	0.096*	0.125~	-0.034***	1.168
	(0.009)	(0.016)	(0.009)	(0.006)	(0.613)	(0.045)	(0.036)	(0.061)	(0.008)	(0.865)
3	0.019*	-0.093***	-0.027**	-0.022***	0.966	-0.166**	0.092*	0.123~	-0.030***	1.002
	(0.009)	(0.014)	(0.010)	(0.005)	(0.667)	(0.053)	(0.036)	(0.069)	(0.008)	(0.694)
4	0.017~	-0.086***	-0.023*	-0.025***	0.923	-0.170**	0.095*	0.109	-0.028**	1.262~
	(0.010)	(0.013)	(0.011)	(0.005)	(0.684)	(0.056)	(0.039)	(0.077)	(0.009)	(0.706)
5	0.011	-0.096***	-0.026*	-0.031***	0.425	-0.175**	0.117**	0.120	-0.037***	1.449*
	(0.011)	(0.014)	(0.012)	(0.005)	(0.779)	(0.056)	(0.037)	(0.089)	(0.009)	(0.694)
6	0.021~	-0.091***	-0.047**	-0.027***	-0.123	-0.162***	0.154**	0.129	-0.027~	-0.124
	(0.011)	(0.017)	(0.015)	(0.005)	(0.861)	(0.044)	(0.052)	(0.077)	(0.014)	(0.988)
Schools base gr.	81	81	81	81	81	30	30	30	30	30
Obs. base gr.	87,221	87,221	87,221	87,221	87,221	33,478	33,478	33,478	33,478	33,478
R-sq base gr.	0.102	0.081	0.145	0.048	0.015	0.061	0.100	0.496	0.096	0.031
Schools gr. 6	103	103	103	103	103	29	29	29	29	29
Obs. gr. 6	77,559	77,559	77,559	77,559	77,559	28,659	28,659	28,659	28,659	28,659
R-sq gr. 6	0.024	0.062	0.242	0.239	0.004	0.018	0.043	0.499	0.309	0.009

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, with standard errors clustered at the base school level. *N Zip Codes* is number of residential zip codes represented, to capture changes in the geographic composition. *Fract.* indicates fraction or proportion in each group, ranging from 0 to 1.

Table 4. ITT estimates of DLI first grade slots-per-pupil effects, by grade and program type

	One-way			Two-way		
Grade	ELA (1)	Math (2)	Science (3)	ELA (4)	Math (5)	Science (6)
All	-0.004	-0.038	-0.049~	0.114*	0.102~	0.006
	(0.031)	(0.028)	(0.030)	(0.051)	(0.055)	(0.077)
3	0.035	-0.052		0.092~	0.050	
	(0.037)	(0.035)		(0.045)	(0.054)	
4	-0.032	-0.063~	-0.048	0.104	0.056	-0.032
	(0.031)	(0.035)	(0.041)	(0.069)	(0.078)	(0.070)
5	-0.004	-0.031	0.008	0.017	0.023	-0.031
	(0.045)	(0.047)	(0.045)	(0.089)	(0.088)	(0.113)
6	-0.092	-0.028	-0.120*	0.362**	0.388**	0.311**
	(0.061)	(0.072)	(0.053)	(0.127)	(0.115)	(0.095)
Schools base gr.	82	82	82	30	30	30
Obs. base gr.	79,121	79,084	76,630	31,411	31,436	30,639
R-sq base gr.	0.095	0.085	0.105	0.132	0.116	0.145
Schools gr. 6	100	100	101	29	29	29
Obs. gr. 6	73,829	72,786	73,822	27,309	27,019	27,290
R-sq gr. 6	0.129	0.105	0.117	0.147	0.119	0.143

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level.

Table 5. Estimated ITT effects on the probability of being classified as EL in each year among those ever classified

	One-way		Two-way	
Grade	Home/ School Language Match (1)	No Language Match (2)	Home/ School Language Match (3)	No Language Match (4)
1	0.031	0.044	0.019	0.065
	(0.027)	(0.031)	(0.012)	(0.044)
2	0.038	0.029	0.016	0.068
	(0.053)	(0.048)	(0.029)	(0.052)
3	-0.027	0.036	0.022	0.093
	(0.064)	(0.064)	(0.035)	(0.060)
4	-0.080	-0.006	-0.021	0.106
	(0.092)	(0.052)	(0.033)	(0.067)
5	-0.014	0.025	-0.082~	0.084
	(0.085)	(0.058)	(0.041)	(0.105)
6	-0.018	-0.006	-0.098*	0.143
	(0.141)	(0.072)	(0.041)	(0.160)
Schools gr. 1	68	79	29	29
Obs. gr. 1	2,545	3,585	11,574	1,695
R-sq gr. 1	0.067	0.106	0.118	0.107
Schools gr. 6	77	94	29	29
Obs. gr. 6	2,483	3,472	9,182	1,742
R-sq gr. 6	0.219	0.240	0.252	0.292

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level.

Table 6. Robustness test against confounding due to selection over time

	Estimated Effects on First DLI Cohort Only					
	One-way			Two-way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	0.000	-0.052		0.060	0.020	
	(0.047)	(0.053)		(0.046)	(0.064)	
4	-0.029	-0.071~	-0.087~	0.066	0.022	-0.128
	(0.040)	(0.040)	(0.046)	(0.089)	(0.089)	(0.080)
5	0.031	-0.026	-0.027	0.020	0.014	-0.039
	(0.049)	(0.052)	(0.047)	(0.112)	(0.097)	(0.131)
6	-0.099	-0.016	-0.069	0.410*	0.372*	0.285*
	(0.065)	(0.073)	(0.070)	(0.160)	(0.153)	(0.116)
Sch. base gr.	82	82	82	30	30	30
Obs. base gr.	51,701	51,707	56,401	22,902	22,922	24,350
R-sq base gr.	0.099	0.088	0.110	0.125	0.112	0.144
Sch. gr. 6	100	100	101	29	29	29
Obs. gr. 6	65,556	64,551	65,554	24,932	24,637	24,900
R-sq gr. 6	0.133	0.108	0.122	0.149	0.120	0.143

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level.

Table 7. Robustness test of extent to which middle school transitions may drive sixth grade ITT estimates

	Base School Ends at Grade 6 or Higher						Base School Ends at Grade 5 or Lower					
	One-way			Two-way			One-way			Two-way		
Grade	ELA (1)	Math (2)	Science (3)	ELA (4)	Math (5)	Science (6)	ELA (7)	Math (8)	Science (9)	ELA (10)	Math (11)	Science (12)
3	0.037	-0.052		0.155*	0.108		0.123~	0.013		0.004	-0.105	
	(0.044)	(0.042)		(0.064)	(0.087)		(0.072)	(0.076)		(0.060)	(0.073)	
4	-0.051	-0.067	-0.048	0.255*	0.214~	0.139	0.099	-0.016	-0.024	-0.099~	-0.084**	-0.215**
	(0.037)	(0.043)	(0.050)	(0.095)	(0.109)	(0.132)	(0.062)	(0.070)	(0.075)	(0.052)	(0.023)	(0.053)
5	-0.030	-0.057	-0.020	0.102	0.123	0.190	0.105	-0.008	0.084	-0.195~	-0.127	-0.372**
	(0.051)	(0.052)	(0.050)	(0.130)	(0.132)	(0.182)	(0.102)	(0.124)	(0.093)	(0.100)	(0.114)	(0.114)
6	-0.099	-0.029	-0.164**	0.365*	0.376**	0.294**	0.133	-0.003	-0.127	0.444**	0.722**	0.507**
	(0.064)	(0.075)	(0.053)	(0.140)	(0.128)	(0.103)	(0.273)	(0.218)	(0.317)	(0.143)	(0.224)	(0.127)
Sch. base gr.	64	64	64	21	21	21	29	29	29	13	13	13
Obs. base gr.	55,001	54,953	53,676	20,802	20,811	20,262	24,120	24,131	22,954	8,511	8,525	8,601
R-sq base gr.	0.091	0.079	0.095	0.116	0.101	0.124	0.112	0.104	0.135	0.175	0.160	0.200
Sch. gr. 6	83	83	84	21	21	21	28	28	28	12	12	12
Obs. gr. 6	54,939	54,293	54,970	18,782	18,724	18,803	18,890	18,493	18,852	7,545	7,313	7,507
R-sq gr. 6	0.130	0.104	0.115	0.130	0.108	0.125	0.136	0.115	0.129	0.198	0.156	0.197

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level.

Table 8. One-way ITT estimates for Spanish programs and weighted by similarity to two-way programs

	Spanish Only			**ATT Weighted by Similarity to Two-way in Pre-DLI Year**		
Grade	ELA (1)	Math (2)	Science (3)	ELA (4)	Math (5)	Science (6)
3	-0.001	-0.059		-0.035	-0.182***	
	(0.057)	(0.051)		(0.049)	(0.052)	
4	-0.025	-0.054	-0.114~	-0.105~	-0.145*	-0.143**
	(0.043)	(0.057)	(0.063)	(0.053)	(0.057)	(0.052)
5	-0.036	-0.144	-0.117	-0.063	-0.132*	-0.039
	(0.088)	(0.086)	(0.078)	(0.058)	(0.060)	(0.060)
6	-0.019	-0.017	-0.052	-0.193~	-0.202	-0.227*
	(0.097)	(0.120)	(0.069)	(0.103)	(0.139)	(0.102)
Schools base gr.	27	27	28			
Obs. base gr.	28,413	28,391	27,299	79,121	79,084	76,630
R-sq base gr.	0.108	0.098	0.125	0.198	0.189	0.261
Schools gr. 6	41	41	41			
Obs. gr. 6	28,405	28,030	28,408	73,829	72,786	73,822
R-sq gr. 6	0.153	0.126	0.137	0.311	0.322	0.284

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level.

Table 9. Estimated interactions between ITT indicator and fraction of students in school with home/partner language match for true and placebo models

	True Interaction Models			Placebo Interaction Models		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
One-Way Program Main Effect Coefficients						
All	0.006	-0.035	-0.069*	0.046	0.036	0.006
	(0.031)	(0.029)	(0.031)	(0.032)	(0.037)	(0.037)
3	0.056	-0.050		0.036	0.010	
	(0.037)	(0.037)		(0.041)	(0.042)	
4	-0.039	-0.082*	-0.091*	0.059	0.070	0.006
	(0.030)	(0.035)	(0.043)	(0.042)	(0.049)	(0.052)
5	-0.021	-0.041	-0.035	0.055	0.051	0.019
	(0.044)	(0.048)	(0.044)	(0.044)	(0.049)	(0.049)
6	-0.090	-0.019	-0.114*	0.038	0.028	-0.018
	(0.064)	(0.075)	(0.055)	(0.040)	(0.055)	(0.042)
Interaction Coefficients: Differential Effects for Unit Diff. in Language Match						
All	0.292*	0.378*	0.344	-0.167	-0.128	0.032
	(0.139)	(0.145)	(0.283)	(0.132)	(0.155)	(0.169)
3	0.114	0.339~		-0.190	-0.133	
	(0.131)	(0.193)		(0.174)	(0.159)	
4	0.457*	0.516**	0.465~	-0.231	-0.206	0.259
	(0.187)	(0.185)	(0.253)	(0.173)	(0.223)	(0.233)
5	0.216	0.207	0.256	-0.182	0.036	-0.119
	(0.264)	(0.232)	(0.369)	(0.155)	(0.180)	(0.173)
6	1.160**	0.918***	0.995***	-0.130	-0.175	0.004
	(0.409)	(0.269)	(0.243)	(0.150)	(0.167)	(0.217)
Sch. base gr.	100	100	100	100	100	100
Obs. base gr.	106,648	106,644	104,125	65,652	65,656	72,222
R-sq base gr.	0.106	0.093	0.116	0.105	0.092	0.118
Sch. gr. 6	119	119	120	119	119	120
Obs. gr. 6	99,252	97,934	99,219	84,385	83,105	84,354
R-sq gr. 6	0.133	0.107	0.123	0.136	0.109	0.128

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects, as well as individual and school-by-grade controls, with standard errors clustered at the base school level.

Figure 1. Within-cohort descriptive estimates, disaggregated for primary English speakers vs. ELs

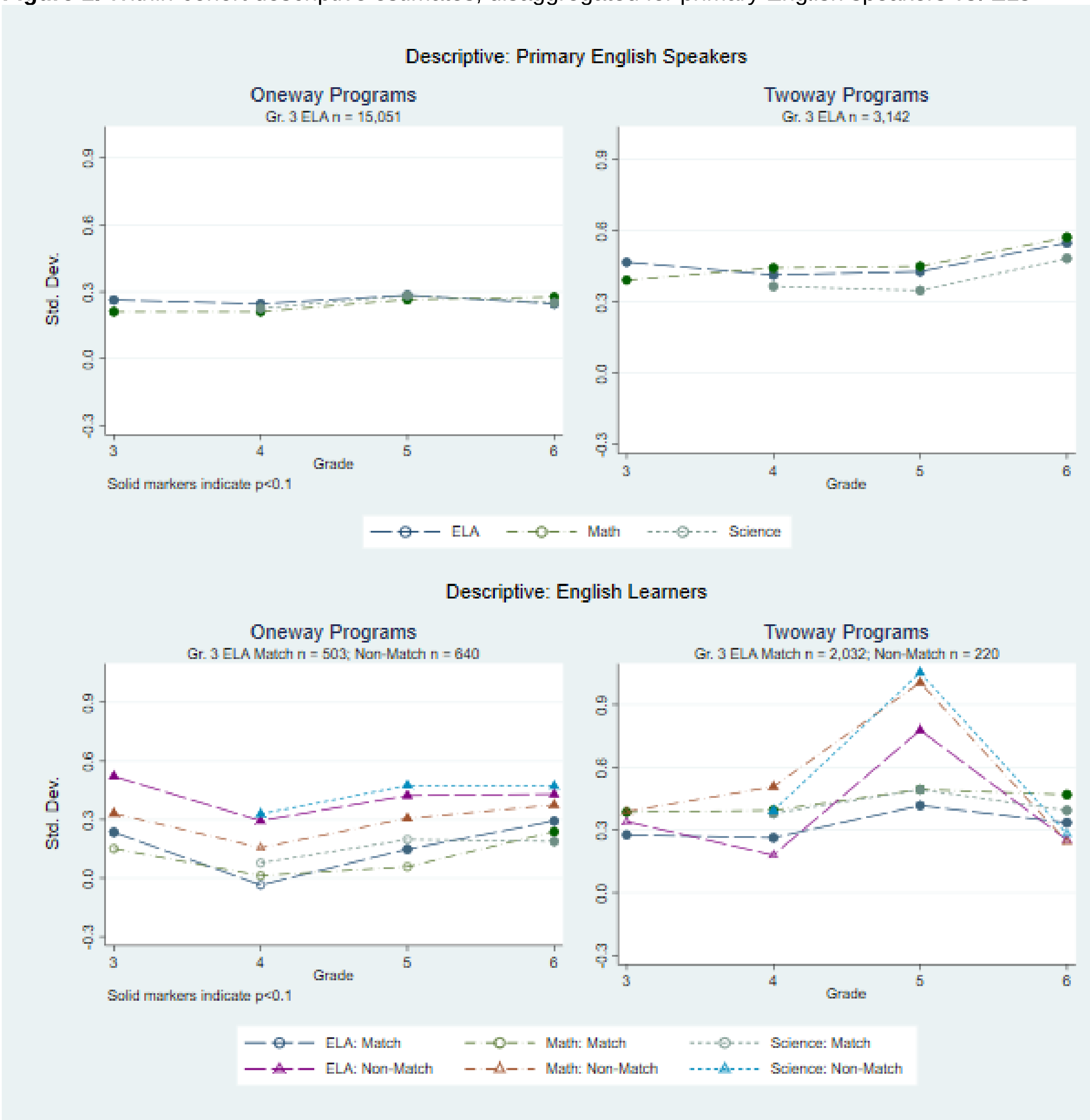
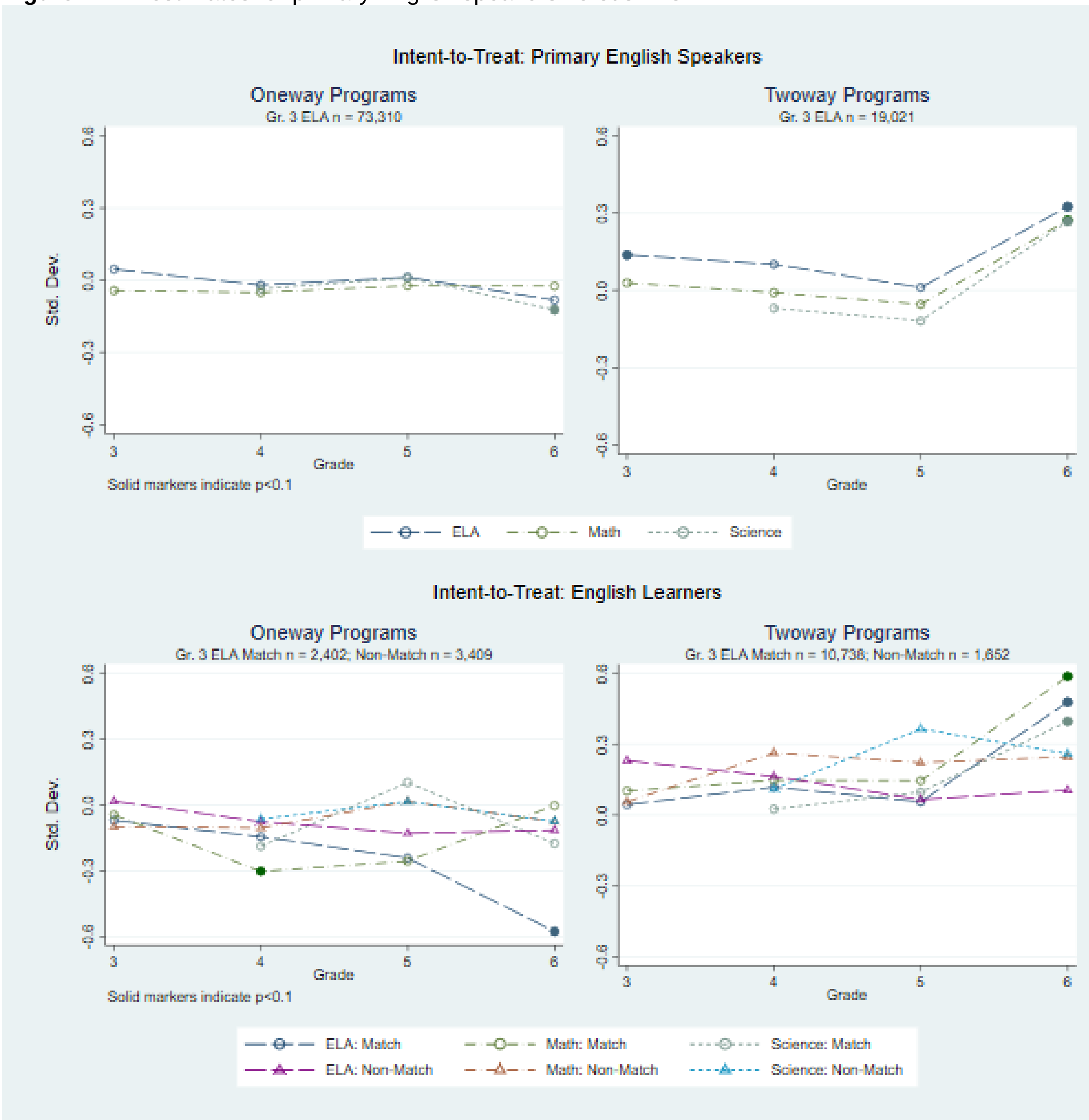


Figure 2. ITT estimates for primary English speakers versus ELs



Note: Y-axis scale differs between Figure 1 and Figure 2.

Appendix I (for Print and Online Versions)

Appendix Table A1. Descriptive within-cohort estimates corresponding to Figure 1 for Never-ELs, Language-Match ELs, and ELs with No Language Match

A. Never EL	One-Way			Two-Way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	0.263***	0.211***		0.467***	0.392***	
	(0.025)	(0.024)		(0.055)	(0.065)	
4	0.246***	0.211***	0.226***	0.414***	0.443***	0.365***
	(0.024)	(0.025)	(0.029)	(0.075)	(0.081)	(0.074)
5	0.285***	0.265***	0.283***	0.428***	0.450***	0.348***
	(0.024)	(0.028)	(0.026)	(0.061)	(0.070)	(0.074)
6	0.246***	0.277***	0.251***	0.548***	0.572***	0.483***
	(0.027)	(0.025)	(0.029)	(0.136)	(0.136)	(0.125)
Schools base gr.	490	490	495	338	337	357
Obs. base gr.	15,051	15,025	15,078	3,143	3,135	3,057
R-sq base gr.	0.077	0.068	0.077	0.158	0.144	0.121
B. Ever-EL: Language Match	One-Way			Two-Way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	0.235**	0.151		0.277***	0.386***	
	(0.082)	(0.111)		(0.073)	(0.080)	
4	-0.035	0.013	0.078	0.263**	0.394***	0.379***
	(0.110)	(0.104)	(0.113)	(0.079)	(0.103)	(0.087)
5	0.148~	0.058	0.198	0.417***	0.494***	0.493***
	(0.087)	(0.132)	(0.128)	(0.083)	(0.090)	(0.115)
6	0.292**	0.238~	0.188~	0.335***	0.469***	0.394***
	(0.105)	(0.128)	(0.110)	(0.085)	(0.090)	(0.091)
Schools base gr.	150	150	164	164	163	155
Obs. base gr.	503	505	530	2,032	2,036	2,012
R-sq base gr.	0.109	0.084	0.078	0.080	0.067	0.064
C. Ever-EL: No Language Match	One-Way			Two-Way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	0.521***	0.331*		0.340	0.389	
	(0.106)	(0.139)		(0.269)	(0.262)	
4	0.294*	0.156	0.329***	0.180	0.507**	0.392*
	(0.126)	(0.110)	(0.094)	(0.219)	(0.165)	(0.191)
5	0.423**	0.306**	0.474***	0.777*	1.005**	1.053***
	(0.132)	(0.106)	(0.110)	(0.344)	(0.302)	(0.139)
6	0.430***	0.375***	0.472***	0.253	0.244	0.282
	(0.121)	(0.095)	(0.104)	(0.371)	(0.490)	(0.379)
Schools base gr.	200	201	211	67	67	75
Obs. base gr.	640	644	700	220	219	192
R-sq base gr.	0.181	0.129	0.172	0.115	0.110	0.220

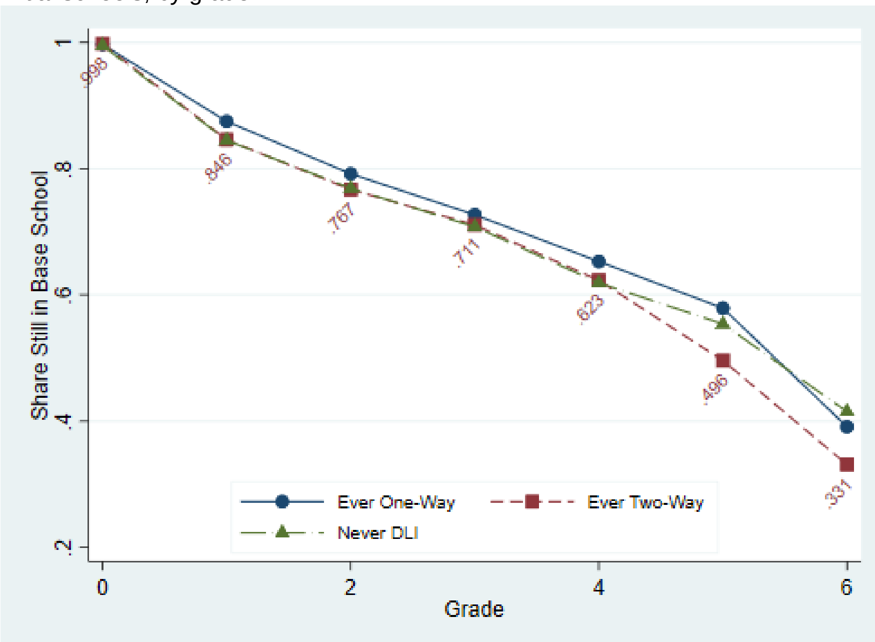
*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects and individual and school-by-grade controls, with standard errors clustered at the base school level.

Appendix Table A2. ITT estimates corresponding to Figure 2 for Never-ELs, Language-Match ELs, and ELs with No Language Match

A. Never EL	One-Way			Two-Way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	0.046	-0.043		0.137*	0.029	
	(0.038)	(0.037)		(0.058)	(0.053)	
4	-0.020	-0.051	-0.036	0.101	-0.009	-0.069
	(0.031)	(0.037)	(0.041)	(0.075)	(0.079)	(0.069)
5	0.013	-0.023	0.010	0.011	-0.054	-0.118
	(0.045)	(0.047)	(0.045)	(0.087)	(0.091)	(0.083)
6	-0.082	-0.024	-0.122*	0.325*	0.272*	0.267*
	(0.062)	(0.071)	(0.055)	(0.123)	(0.129)	(0.108)
Schools base gr.	82	82	81	30	30	30
Obs. base gr.	73,310	73,246	70,978	19,021	19,011	18,637
R-sq base gr.	0.071	0.060	0.069	0.094	0.078	0.086
B. Ever-EL: Language Match	One-Way			Two-Way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	-0.071	-0.045		0.044	0.102	
	(0.120)	(0.218)		(0.058)	(0.081)	
4	-0.145	-0.303~	-0.189	0.118	0.146	0.025
	(0.148)	(0.171)	(0.143)	(0.083)	(0.097)	(0.108)
5	-0.242	-0.256	0.102	0.056	0.145	0.097
	(0.213)	(0.227)	(0.213)	(0.123)	(0.120)	(0.183)
6	-0.577*	-0.003	-0.176	0.478**	0.587***	0.396**
	(0.270)	(0.268)	(0.166)	(0.169)	(0.116)	(0.134)
Schools base gr.	71	72	69	29	29	29
Obs. base gr.	2,402	2,415	2,293	10,738	10,766	10,332
R-sq base gr.	0.081	0.060	0.069	0.056	0.036	0.031
C. Ever-EL: No Language Match	One-Way			Two-Way		
Grade	ELA	Math	Science	ELA	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
3	0.017	-0.099		0.230	0.056	
	(0.094)	(0.103)		(0.138)	(0.139)	
4	-0.077	-0.105	-0.064	0.163	0.263	0.110
	(0.084)	(0.106)	(0.100)	(0.175)	(0.175)	(0.143)
5	-0.129	0.016	0.015	0.066	0.222	0.365
	(0.118)	(0.123)	(0.123)	(0.266)	(0.248)	(0.354)
6	-0.115	-0.076	-0.075	0.105	0.246	0.259
	(0.147)	(0.179)	(0.138)	(0.300)	(0.180)	(0.178)
Schools base gr.	76	76	78	29	29	29
Obs. base gr.	3,409	3,423	3,359	1,652	1,659	1,670
R-sq base gr.	0.085	0.073	0.075	0.093	0.089	0.098

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects and individual and school-by-grade controls, with standard errors clustered at the base school level.

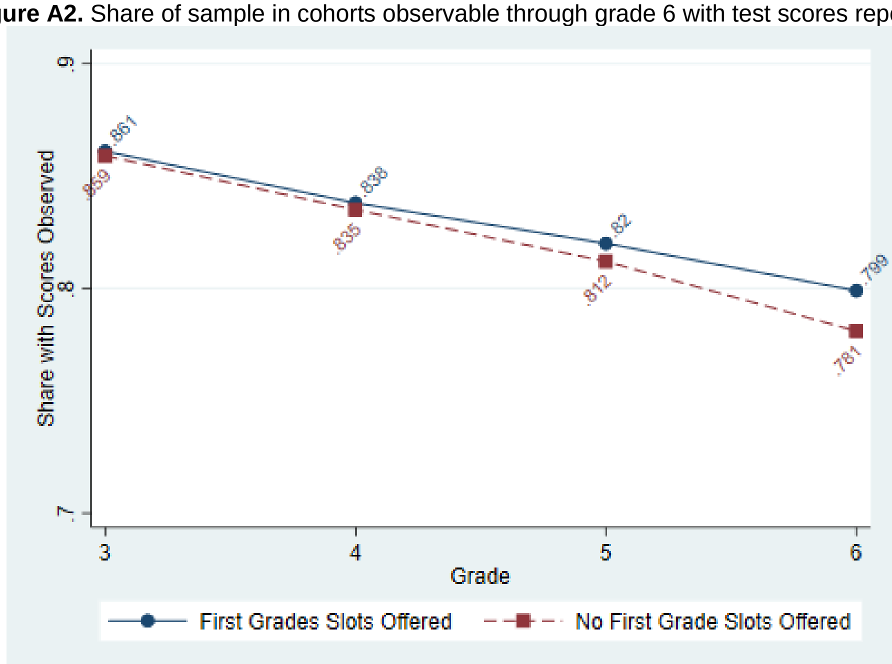
Appendix Figure A1. Share of sample in cohorts observable through grade 6 whose current schools match their initial schools, by grade



Appendix II: Attrition Considerations (Online-Only)

Because access to distinctive school programs like immersion could affect families' attrition from public school in Utah over time, we examine attrition patterns in the data, as well as the sensitivity of our estimates to attrition by grade 6.

Appendix Figure A2. Share of sample in cohorts observable through grade 6 with test scores reported, by grade



Appendix Figure A2 displays the share of students in the sample who entered the Utah public schools by grade 3 and had test scores available in each of the tested grades in our analysis—grades 3 through 6. We see that persistence rates (1 minus the rate of attrition) are very similar regardless of whether first-grade DLI slots were available in their base school in their first-grade year. Persistence in the dataset by grade 6 is about 79.9% for treated students and about 78.1% for control students, a difference of 1.8 percentage points, and the gaps are even narrower in earlier grades. This combination of overall attrition (under 20%) and differential attrition (less than two percentage points) falls easily within the conservative attrition boundary set forth by the U.S. Department of Education's What Works Clearinghouse (2020) that would be sufficient to meet causal evidence standards for low attrition in a randomized, controlled trial.

Appendix Table A3. Robustness test to address the possibility of differential attrition

	Weighted for Probability of Attrition by Grade 6					
	One-way			Two-way		
Grade	ELA (7)	Math (8)	Science (9)	ELA (10)	Math (11)	Science (12)
3	0.018	-0.073		0.041	-0.001	
	(0.044)	(0.046)		(0.061)	(0.059)	
4	-0.036	-0.093*	-0.076	0.128	0.050	-0.086
	(0.034)	(0.040)	(0.047)	(0.082)	(0.091)	(0.071)
5	-0.023	-0.057	-0.040	0.063	0.055	-0.014
	(0.050)	(0.052)	(0.052)	(0.101)	(0.098)	(0.118)
6	-0.098	-0.030	-0.123*	0.358**	0.381**	0.311**
	(0.061)	(0.073)	(0.053)	(0.128)	(0.115)	(0.093)
Sch. base gr.	57,126	57,098	62,120	23,070	23,099	25,031
Obs. base gr.	0.151	0.137	0.164	0.186	0.167	0.228
R-sq base gr.	73,807	72,749	73,799	27,307	27,014	27,288
Sch. gr. 6	0.181	0.156	0.177	0.201	0.166	0.212
Obs. gr. 6	57,126	57,098	62,120	23,070	23,099	25,031
R-sq gr. 6	0.151	0.137	0.164	0.186	0.167	0.228

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1 Models include base school and cohort fixed effects and individual and school-by-grade controls, with SEs clustered at the base school level.

Nevertheless, because many of the slots-per-first-grade effects we observe in our analysis occur in grade 6, we also add a robustness test in which we upweight observed students who have a higher predicted risk of attrition. Specifically, we estimate a student's probability of persistence in the dataset through grade 6 for the fall 2000 through 2011 kindergarten cohorts by regressing a time-invariant dichotomous indicator of whether the student has sixth grade scores on baseline vectors $\mathbf{C}_c$, $\mathbf{S}_s$, and $\mathbf{X}_{ics}$ from equation 2, which reflect their base school, base cohort, available DLI slots per first grader in their school-by-year, and their individual and school-by-grade characteristics. We then re-estimate equation 2 using weights for the inverse of the predicted probability of persistence in the dataset, thus upweighting observations with higher predicted levels of attrition (Wooldridge, 2002). As shown in Appendix Table A3, our findings are largely unaffected by this weighting. In combination with the relatively low levels of overall and differential attrition the dataset, we take this to suggest that our main results are unlikely to be driven by systematic sorting of treatment or control students out of the Utah public schools.