10-1-2015

Effects of Dual-Language Immersion on Students' Academic Performance

Jennifer L. Steele
American University

Robert O. Slater
American Councils for International Education

Gema Zamarro
University of Arkansas, Fayetteville, gzamarro@uark.edu

Trey Miller
RAND Corporation

Jennifer Li
RAND Corporation

See next page for additional authors

Citation

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2015). Effects of Dual-Language Immersion on Students' Academic Performance. *Education Reform Faculty and Graduate Students Publications.* Retrieved from https://scholarworks.uark.edu/edrepub/41

University of Arkansas | The Department of Education Reform

WORKING PAPER SERIES

Effects of Dual-Language Immersion on Students' Academic Performance

Jennifer L. Steele , Robert O. Slater, Gema Zamarro, Trey Miller,

Jennifer Li, Susan Burkhauser, Michael Bacon

October 2015

EDRE Working Paper 2015-09

Effects of Dual-Language Immersion on Students' Academic Performance

October 2015

Jennifer L. Steele
American University
steele@american.edu

Robert O. Slater
American Councils for International Education
rslater@americancouncils.org

Gema Zamarro
University of Arkansas & University of Southern California
gzamarro@uark.edu

Trey Miller
RAND Corporation
tmiller@rand.org

Jennifer Li
RAND Corporation
jennifer@rand.org

Susan Burkhauser
RAND Corporation
sburkhau@rand.org

Michael Bacon
Portland Public Schools
mbacon@pps.net

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through State and Local Policy Programs and Systems grant # R305E120003 to the RAND Corporation, the American Councils for International Education, and the Portland Public Schools. This effort has benefitted enormously from ongoing research collaboration with Deborah Armendariz, Director of Dual Language in the Portland Public Schools, and from ongoing feedback from Allen Ruby at the Institute of Education Sciences. It would not have been possible without outstanding data assistance by Joseph Suggs, Karin Brown, and Jennifer Miller in the Portland Public Schools and Jonathan Wiens in the Oregon Department of Education. The opinions expressed are those of the authors and do not represent views of IES, the U.S. Department of Education, or the research partner organizations.

Abstract

Using data from seven cohorts of language immersion lottery applicants in a large, urban school district, we estimate the causal effects of immersion on students' test scores in reading, mathematics, and science, and on English learners' (EL) reclassification. We estimate positive intent-to-treat (ITT) effects on reading performance in fifth and eighth grades, ranging from 13 to 22 percent of a standard deviation, reflecting 7 to 9 months of learning. We find little benefit in terms of mathematics and science performance, but also no detriment. By sixth and seventh grade, lottery winners' probabilities of remaining classified as EL are three to four percentage points lower than those of their counterparts. This effect is stronger for ELs whose native language matches the partner language.

Keywords: Language immersion**;** English language learners**;** Student achievement**;** Urban education**;** Language education

JEL Classification: I21, C90

Introduction

Dual-language immersion schools, which provide native English speakers and English learners (ELs) with general academic instruction in two languages from kindergarten onward, are proliferating rapidly in the United States. The Center for Applied Linguistics (2011a, 2011b) estimates that the number of immersion schools in the U.S. grew from 278 to 448 between 1999 and 2011, but more-recent extrapolations place the latest number between 1000 and 2000 (Maxwell, 2012; Watanabe, 2011). For instance, through recent statewide efforts, Utah is home to at least 118 language immersion schools, and North Carolina to 94 (North Carolina Department of Education, 2014; Utah State Office of Education, 2014). Meanwhile, the New York City Department of Education more than doubled the number of dual-language immersion programs it offers, from about 82 to 175, between the 2012-13 and 2015-16 school years (New York City Department of Education, 2015; Schneider, 2013).

This swift expansion of an approach that was recently considered boutique seems driven by a few complementary forces: growth in the share of U.S. school children who are ELs (U.S. Department of Education, 2014); observational evidence that ELs in dual-language immersion programs academically outperform ELs in English-only or transitional bilingual programs (Collier & Thomas, 2004; Lindholm-Leary & Block, 2010; Umansky & Reardon, 2014; Valentino & Reardon, 2015); and demand from parents of native English speakers who anticipate benefits of bilingualism within a globally competitive society (Maxwell, 2012). In addition, considerable laboratory evidence suggests that bilinguals outperform monolinguals on some measures of executive function and attention control (e.g., Bialystok & Craik, 2010).

Though a number of studies have examined the performance of students in dual-language immersion versus monolingual education, most have been observational studies that, due to data constraints, cannot fully adjust for unobserved differences between immersion and non-immersion participants. Our study attempts to address this limitation by capitalizing on a lottery that randomly assigns students to language immersion in the Portland Public Schools (PPS) in Portland, Oregon. PPS serves 47,000 students and is among the largest two public school districts in the Pacific Northwest. We find that students randomly assigned to immersion in kindergarten outperform their counterparts in fifth grade reading by 13% of a standard deviation, and in eighth grade reading by more than a fifth of a standard deviation. Conditional on their EL status at school entry, lottery winners are three to four percentage points less likely to be classified as ELs in sixth and seventh grade, and the estimates are larger for students whose native language matches the partner language. The effects of lottery winning on mathematics and science performance appear positive in magnitude but are indistinguishable from zero in most cases.

In this article, we first discuss the literature on dual-language immersion education and how immersion is implemented in Portland. We then describe our sample and analytic methods and present our results. We conclude with a discussion of implications for public education in the globally competitive, 21st century economy.

Background

Substantial research from cognitive psychology points to the cognitive benefits of bilingualism, such as improved working memory and attention control (Bialystok & Craik, 2010;

Bialystok, Craik, & Luk, 2008). These functions appear to play a key role in solving mathematics problems and in comprehending written material (Alloway, 2007; Gathercole, Alloway, Willis, & Adams, 2006). Though researchers disagree about whether achievement can be raised by discrete cognitive training tasks (D'Amico & Guarnera, 2005; Jacob & Parkinson, 2015; McClelland et al., 2007), immersion education is a comprehensive instructional approach that may yield direct academic benefits—proficiency in multiple languages—while also benefitting cognition and generalized academic performance (Esposito & Baker-Ward, 2013). Researchers have also reached different conclusions about the extent to which linguistic similarity mediates a bilingual advantage, with some evidence suggesting that orthographically similar languages confer greater benefits in executive control (Coderre & van Heuven, 2014), and other evidence suggesting little difference (Paap, Darrow, Dalibar, & Johnson, 2014).

Because U.S. education accountability policy focuses on students' proficiency in English language arts, mathematics, and science—all of which are typically tested in English—policymakers and parents are sometimes concerned that dual-language immersion will slow students' ability to master such content (e.g., Hoff et al., 2012). In response, proponents of dual-language immersion point to studies suggesting that instruction in two languages conveys academic advantages over monolingual instruction. These studies are important but in many cases lack a convincing basis for causal inference. One pioneering study of a French immersion program in Canada found that native English-speaking students randomized to French immersion in kindergarten lagged their counterparts on some measures of English language arts until fifth grade, at which point they matched or outperformed their peers in both language arts and mathematics (Lambert, Tucker, & d'Anglejan, 1973). Though the study was rigorously designed, it was conducted on a small scale, with only 63 participants persisting through grade 5. In the United States, one randomized study of dual-language immersion in a preschool found mostly positive benefits on students' Spanish reading skills among native Spanish and native English speakers, and no clear detriment or benefit to reading skills in English, but the study was small, with 150 students, and was able to track students for only one year (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007). In a small study of 124 mostly native English speakers in a Mandarin immersion program, Padilla and colleagues (2013) demonstrated that immersion students outperformed same-school peers on an English language arts examination in grades 3 through 5, but the study did not adjust for baseline differences between groups, and though the immersion group was admitted by lottery, the same-school comparison group was not necessarily randomly assigned. Because the programs described in these studies were contained within single schools, a concern is that their findings may not generalize to larger-scale programs.

Other studies of immersion on native English speakers in U.S. or Canadian contexts also found that students enrolled in immersion programs outperform their peers in English language arts and/or mathematics, especially from the middle elementary grades onward, but in many cases, these studies were not able to employ aggressive controls for selection into the programs (Barik & Swain, 1978; Caldas & Boudreaux, 1999; Marian, Shook, & Schroeder, 2013; Turnbull, Hart, & Lapkin, 2003). Reinforcing the concern that some extant estimates may be due to families' selection of programs or school zones, a recent, well-designed study leveraging the sudden introduction of immersion programs to elementary schools in Spain found negative-to-no achievement effects in core content by grade 6 (Anghel, Cabrales, & Carro, 2012). As with the Canadian studies, however (Barik & Swain, 1978; Lambert et al., 1973; Turnbull et al., 2003), it is possible that studies conducted outside the United States may not fully generalize to U.S. contexts.

Most studies of dual-language immersion in the United States have focused on the outcomes for ELs whose native language matches the partner (i.e., non-English) language, rather than on native English speakers. This is important since dual-language immersion for ELs serves as one possible alternative to monolingual English instruction or to bilingual education programs in which students receive core instruction in their native language until they are able to transition to monolingual English classes in early or later elementary school. (Early-transition programs are sometimes called transitional bilingual, and later-exit programs are sometimes called developmental bilingual programs (Francis, Lesaux, & August, 2006; Valentino & Reardon, 2015). A key distinction of dual-language immersion programs is that they typically include native English speakers, and may therefore provide less segregation of ELs than the other program types, while still supporting these students in learning their native language as well as English. Some dual-language immersion programs—called *two-way* programs--are explicitly designed to serve native speakers of both languages, whereas *one-way* programs primarily serve students who are new to the partner language (Collier & Thomas, 2004).

It is conceivable that immersion effects for native speakers of the partner language may differ from those for native English speakers, since immersion programs allow ELs to receive a substantial share of core academic content instruction in their native language, to share a classroom with native English speakers, and to begin school with a baseline advantage over their monolingual English-speaking peers in terms of knowledge of the partner language.Thus, in an immersion context, ELs may experience their native language skills as a particular academic asset. In numerous studies, ELs placed in immersion programs that match their native language have outperformed their peers placed in monolingual English classes or transitional bilingual programs (Collier & Thomas, 2004; Lindholm-Leary & Block, 2010; Marian et al., 2013; Thomas & Collier, 2011). However, many of these studies include few adjustments for baseline between-group differences. In a notable exception, Umansky and Reardon (2014) employed hazard analysis with extensive statistical controls; finding that Latino ELs placed in Spanish immersion classrooms are reclassified from English learner to English-proficient status more slowly in elementary school but at higher rates by high school.[1] More recently, Valentino and Reardon (2015) compared the academic performance of ELs placed in monolingual English instruction, transitional bilingual education, developmental bilingual education, and dual-language immersion programs. They found that students in all three of the bilingual programs, including dual-language immersion, grew as fast as or faster in English language arts performance than their peers in monolingual English instruction. Importantly, the study disaggregated effects for Latino versus Chinese ELs, finding that dual-language immersion in the native language and English was associated with the most-positive trajectory in English language arts performance between kindergarten and grade 7.

A number of meta-analyses have examined the effects of transitional bilingual education as compared to English-only instruction for ELs, reaching varied conclusions due to differences in their inclusion criteria (Greene, 1997; Rossell & Baker, 1996; Slavin & Cheung,

[1] Reclassification as English-proficient typically means that students no longer qualify for EL support services, but this may increase their access to mainstream academic offerings within the school.

2005; Willig, 1985). In a meta-analysis of 71 effect estimates from 15 studies, Francis and colleagues (2006) found a net benefit of transitional bilingual education on students' reading scores of about 33% of a standard deviation, or 37% among studies that used random-assignment. In a subsequent study that randomly assigned Spanish-dominant kindergarteners to transitional bilingual versus English-only instruction, Slavin and colleagues (2010) found that initial Spanish advantages among the transitional-bilingual group and English advantages among the English-only group became statistically indistinguishable by grade 4.

Taken together, the existing research on dual-language immersion education for ELs and suggests that families who are able to enroll their children in dual-language immersion programs can expect to see equivalent or even outperformance in English language arts by elementary school, but the extent to which selection is driving these estimates is less clear.

The present study contributes to this body of research in several ways: first, it is one of few studies to examine the general academic effects of immersion education on native English speakers as well as ELs in the United States, and to do so longitudinally between kindergarten and (for the oldest cohort) eighth grade. Second, it examines effects at scale in a large urban district, focusing on twelve schools and four partner languages. Finally, it leverages data from a district-wide lottery system in order to estimate causal effects over time, integrating test scores from a state data system to track students who leave the district but remain in the state. As such, it represents the largest random-assignment study of dual-language immersion we are aware of, and it is able to estimate causal effects over time for native English speakers as well as for native speakers of other languages. Our analysis responds to three research questions:

1. What is the causal effect of random assignment to dual-language immersion on student achievement in mathematics, English language arts, and science, and (for students who began as ELs in kindergarten) on students' subsequent classification as ELs?
2. To what extent do immersion effects differ for one-way versus two-way immersion programs and for programs in Spanish versus Mandarin, Japanese, and Russian?
3. To what extent do immersion effects depend on whether a student's first language is English and on whether the student's first language matches the partner language?

Intervention and Setting

Portland Public Schools began implementing dual-language immersion programs in 1986. During the 2012-13 academic year when our study commenced, it maintained programs in 11 elementary schools, 4 middle schools, and 5 high schools, with instruction in Spanish, Mandarin, Japanese, and Russian.[2] In that year, about 8% of percent of Portland's students, or 3,860 individuals, were enrolled in immersion. Key characteristics of these programs are summarized in Table 1, including their instructional models and student composition.

<Table 1 about here >

During the school years in our analysis, the Russian program and all but one of the Spanish programs followed a *two-way* model in which about of half of the students were native speakers of the partner language—Spanish or Russian—and the other half were native speakers of English or another language. The district's other immersion programs (Japanese, Mandarin,

[2] Two of the middle schools are home to two languages each.

and one Spanish program), offered a *one-way* model, in which most students were native English speakers.

Two-way programs. As noted in Table 1, the two-way programs in Portland follow a 90/10 instructional model, meaning that in kindergarten, 90% of the school day is conducted in the partner language, and 10% in English. The partner-language proportion then declines by 10 percentage points per grade. In grades K-3, students receive 75%-100% of mathematics instruction, 56%-100% of language arts instruction, and about 100% of science and social studies instruction in the partner language. In grades 4 and 5, they receive about 25% of mathematics, 58% of language arts, and 100% of science and social studies instruction in the partner language. Middle school students take one language arts class in English, one language arts class in the partner language, and one social studies class in the partner language; the rest of their classes are conducted in English. High school immersion students typically take only one class per day—an advanced language class—in the partner language.

One-way programs. In Portland's one-way programs, instruction of core content (mathematics, language arts, science, and social studies) follows a 50/50 instructional model in each elementary grade. Each day, half of the instruction in each core subject occurs in the partner language, and half occurs in English (see Table 1). In middle and high school, however, one-way and two-way programs operate similarly, with middle school immersion students taking about two classes per day in the partner language, and high school students taking about one per day.

Instructional practice and partner-language learning. Immersion and non-immersion students in the district are held to the same academic content standards, and the district develops or purchases partner-language curricula to make this possible. Still, it is possible that instructional practices would differ between immersion and non-immersion classrooms. In the spring of 2014, our research team conducted observations of 119 forty-five-minute instructional sessions, noting that time allocated to the partner language in each subject and grade (focusing on grades 1, 3, and 5) was reasonably consistent with the aforementioned district guidelines for the 90/10 and 50/50 models. In our observations of 46 immersion and 33 English-only classrooms in the 2012-13 academic year, we recorded similar distributions of on-task student behavior and instructional strategies across languages, though all observations were conducted in schools that had immersion programs. In terms of proficiency in the partner language, district-administered eighth-grade tests of immersion students using the Standards-Based Measurement of Proficiency (STAMP-4S) (Avant Assessment, 2015) suggest that immersion students in Spanish and Chinese reach intermediate-mid-level proficiency (5 to 6 on 9-point scales) by grade 8; students in Japanese reach intermediate-low-level proficiency (4 to 5 on 9-point scales).

Entry to Immersion in Portland

Students receive admission to immersion programs in Portland through a lottery process administered by the school district. In the spring prior to their child's pre-k or kindergarten year, families may apply for up to three school programs of their choice (including immersion and a few other program types), in order of preference. The number of lottery slots available in a given program and year is established by the school principal, and many immersion schools establish multiple preference categories, such as slots for native speakers of the partner language, for students who live in the school's catchment neighborhood, and for students living in other neighborhoods. Students receive a random lottery number for each preference choice, but in practice, all immersion slots are filled in the first lottery round.

Within each lottery round, slots in a given school and preference category are filled first by students who have siblings at the school, then by other applicants who reside with the school district, and then by applicants from outside the district. Consequently, for any given school and preference category, randomization will occur for only one of the three subcategories—co-enrolled siblings, no co-enrolled siblings, or out-of-district. (Most randomization occurs within the in-district, no-sibling category.) We consider a lottery to be binding only if there are winners and losers within a given category and subcategory in a given year. In other words, only a subset of lottery applicants are truly randomized, and we limit our lottery-based analysis to this subset. Students who do not win an immersion slot are assigned to the regular instructional program in their default neighborhood schools.

Data and Sample

The study focuses on the seven cohorts of students who applied to a pre-k or kindergarten immersion slot in Portland for the fall terms of 2004 through 2010.[3] Outcome data are measured through the 2013-14 academic year, so the oldest cohort can be tracked through ninth grade, and the youngest is observed through third grade. The lottery applicant sample

[3] We classify the lottery-winning status of pre-k applicants based on their first application, but results are not sensitive to this decision.

includes 3,457 students, and we have data on 26,018 other Portland students who enrolled as pre-kindergarteners or kindergarteners during the years in question.

<Figure 1 about here >

The CONSORT diagram (Schultz, Altman, & Moher, 2010) shown in Figure 1 describes the randomization process. Of the 3,457 students who applied to Portland immersion lotteries during the study years, 1,946 (56.3%) were truly randomized within a binding lottery category and subcategory. Of those truly randomized, 44.4% won immersion slots (the treatment group), and 1,082 (55.6%) did not (the control group). Working with the Oregon Department of Education, we are able to obtain outcome data (reading, mathematics, or science scores or English language learning status) for 1,625 randomized students, meaning that overall sample attrition is 16.5%.[4] Attrition is 13.0% for the treatment sample and 19.3% for the control sample, yielding differential attrition of 6.3 percentage points.[5] This combination of overall and differential attrition rates lies very near the conservative threshold for meeting What Works Clearinghouse (2014) evidence standards, and it falls easily within the liberal threshold. To provide further assurance of balance--and to improve the precision of our estimates--our models adjust for observed baseline characteristics, as well as for lottery strata fixed effects.

Intent-to-treat effects, which are the estimated effects of winning the lottery, may of course understate the effect of immersion program enrollment. In the analytic sample, compliance with assigned status is 77% for the treatment group and 73% for the control group,

[4] To capture academic outcomes for individuals who enroll in Oregon public schools outside of Portland, we were able to match PPS to Oregon Department of Education data. This augmented the analytic sample by 11% and improved grade-specific samples by 7-24%.
[5] By grade 8, the rate of sample persistence from the point of randomization is 67.9% for the treatment group and 72.5% for the control group, for a 4.6-point differential.

where compliance for winners is defined as kindergarten enrollment in a Portland immersion program, and compliance for those not placed is defined as *not* enrolling in a Portland immersion program in kindergarten. As described below, we use instrumental variables (IV) analyses (Angrist & Pischke, 2008) to recover the effect for those who comply with their random-assignment status.

<Table 2 about here >

The left-hand side of Table 2 presents descriptive statistics for the randomized (binding) analytic sample, and the right-hand side presents comparable information for the full sample of pre-k and kindergarten entrants to Portland. For binding lottery applicants, the intent-to-treat condition is defined as winning or not winning an immersion slot; for all Portland kindergarten entrants, the treatment is enrollment in immersion in kindergarten, and the comparison condition is not enrolling in immersion in kindergarten. Table 2 also presents the difference between groups for each variable, and p-values for t-tests of the differences. Because t-tests are affected by sample size, one might be more concerned with the magnitude of the difference in terms of pooled standard deviation units (What Works Clearinghouse, 2014), which we report at left for the full sample. For the randomized group, the p-values are adjusted for lottery strata fixed effects and thus refer to within-strata differences. Because randomization produces groups that are identical *in expectation* (Rubin, 1974), we might expect chance differences in observable characteristics in finite samples. The bottom panel of Table 2 indicates the number of students in the analytic sample at each grade; it becomes smaller over time primarily because cohorts are observed for different lengths of time. Because the ninth grade sample includes only one cohort, ninth-grade estimates are especially noisy and are not reported in our analysis.

Outcome Measures

Student achievement in reading, mathematics, and science is measured by performance on the state-mandated accountability test, the Oregon Assessment of Knowledge and Skills (OAKS). Mathematics and reading tests are administered annually in grades 3 through 8 and once in high school; science is tested in grades 5 and 8. The tests are administered solely in English. We standardize scores to have mean zero and standard deviation one within grade level, subject, and school year. We also examine a student's status as an English language learner in each academic year after kindergarten, adjusting for his or her status at kindergarten entry. Students in Portland may be classified as EL each year based on their status the prior year and their overall performance on the English Language Proficiency Assessment (ELPA). ELPA tests are typically administered between January and March. We code a student as being an EL until the first full school year in which he/she no longer qualifies for services based on ELPA scores.[6]

Analytic Strategy

Full-Sample Analysis: Generalized Least Squares

To gauge the relationship between immersion and performance in the full sample of kindergarten entrants to Portland, even for those not randomized, we first undertake a

[6] In the data, reclassification is highly consistent with ELPA proficiency, suggesting strong adherence to the policy.

covariate-adjustment approach in the full sample. We compare the outcomes of interest for students who did and did not begin immersion in kindergarten, adjusting for the observed baseline characteristics reported in Table 2. Because we are interested in immersion effects over time, we use generalized least squares (GLS) models with student-level random effects to estimate immersion effects in each observed grade level and to adjust for the nesting of observations within students (Raudenbush & Bryk, 2002). We define the treatment as time-invariant (based on kindergarten enrollment) so that any subsequent movement into and out of immersion programs over time would conservatively bias our treatment estimates toward zero. The estimation model is as follows:

$$y_{it} = a_1 + \tau_1 DLI_i^{kg} + \boldsymbol{\theta_1 G_{it}} + \boldsymbol{\beta_1}(\boldsymbol{DLI_i^{kg} G_{it}}) + \boldsymbol{\delta_1 X_i} + u_{1i} + \varepsilon_{1it} \quad (1)$$

where the dependent variable, y_{it}, represents the outcome of interest for student *i* at time *t.* $\boldsymbol{G_{it}}$ is a vector of dichotomous grade-level dummy variables with effects given by vector $\boldsymbol{\theta_1}$. The predictors of interest are the observed value of immersion enrollment in kindergarten, DLI_i^{kg}, and its interaction with grade level, $\boldsymbol{DLI_i^{kg} G_{it}}$. Vector $\mathbf{X}_i$ contains time-invariant student demographic characteristics observed in kindergarten, including the child's race/ethnicity, gender, subsidized-meal eligibility, whether the child's first language is English, and whether the child is classified in kindergarten as needing special education services. $\boldsymbol{\delta}_1$ is its corresponding parameter vector, and α_1 is an intercept term. The student-level error term is given by u_{1i}, and the observation-level error term is represented by ε_{1it}, both assumed to be normally distributed with zero means and constant variances.[7]

[7] We us a linear probability model for EL status, but results are similar when we use logit models.

Intent-to-Treat Analysis

Given that the full-sample analysis is vulnerable to selection on unobservables, our causal identification strategy capitalizes on students' random assignment to immersion. We estimate the causal effect of winning an immersion lottery using a model that accounts for randomization within blocks that are specific to the student's application year, first-choice school, and preference category and subcategory. We implement this within-block randomization using lottery strata fixed effects in a model specified as follows:

$$y_{it} = a_2 + \tau_2 z_i + \boldsymbol{\theta_2 G_{it}} + \boldsymbol{\beta}_2(\boldsymbol{z_i G_{it}}) + \boldsymbol{\delta_2 X_i} + \boldsymbol{\gamma_2 L_i} + u_{2i} + \varepsilon_{2it} \quad (2)$$

where the terms are as described above, except that the intent-to-treat variable, z_i, is a dichotomous indicator of random assignment to the treatment in the lottery for student *i,* and $\boldsymbol{z_i G_{it}}$ is its interaction with the student's grade level. $\mathbf{L}_j$ is a vector of time-invariant dichotomous cohort × school × randomization subgroup lottery indicators, and $\boldsymbol{\gamma_2}$ is a corresponding vector of lottery fixed effects. The parameters of interest are τ_2, representing the main effect of winning the lottery, and vector $\boldsymbol{\beta_2}$, representing differential effects of lottery winning by grade.

To address the second and third research questions, we assess whether the causal effects of immersion differ by characteristics of the program to which the student applied (one-way vs. two-way models, and Spanish versus other languages), and by key student characteristics (native language other than English, and native speaker of the partner language). We do this by including three-way interactions in the model among the category of interest (C_{it}), the students' random assignment status (z_i), and grade level ($\boldsymbol{G_{it}}$):

$$y_{it} = a_3 + \tau_3 z_i + \boldsymbol{\theta_3 G_{it}} + \boldsymbol{\beta_3}(\boldsymbol{z_i G_{it}}) + \upsilon_3 C_{it} + \kappa_3(z_i C_{it}) + \boldsymbol{\phi_3}(\boldsymbol{G_{it} C_{it}}) +$$

$$\boldsymbol{\eta_3}(\boldsymbol{z_i G_{it} C_{it}}) + \boldsymbol{\delta_3 X_i} + \boldsymbol{\gamma_3 L_i} + u_{3i} + \varepsilon_{3it} \quad (3)$$

The key parameters of interest are the coefficients on the treatment assignment-by-category interaction, κ_3, and on the treatment assignment-by-category-by-grade interaction terms, $\boldsymbol{\eta_3}$.

Instrumental Variables Analysis

To estimate the causal effect of immersion enrollment in kindergarten on those who complied with their initial lottery-assignment status, (known as the local average treatment effect, or LATE), we use lottery assignment status as an instrument for DLI enrollment in kindergarten, specifying a two-stage least squares regression model as follows:

$$DLI_i^{kg} = a_4 + \tau_4 z_i + \boldsymbol{\theta_4 G_{it}} + \boldsymbol{\beta_4}(\boldsymbol{z_i G_{it}}) + \boldsymbol{\delta_4 X_i} + \boldsymbol{\gamma_4 L_i} + u_{4i} + \varepsilon_{4it} \quad (4)$$

$$y_{it} = a_5 + \tau_5 \widehat{DLI}_i^{kg} + \boldsymbol{\theta_5 G_{it}} + \boldsymbol{\beta_5}(\boldsymbol{\widehat{DLI}_i^{kg}} * \boldsymbol{G_{it}}) + +\boldsymbol{\delta_5 X_{ij}} + \boldsymbol{\gamma_5 L_j} + u_{5i} + \varepsilon_{5it} \quad (5)$$

In the first stage (equation 4), the randomly assigned lottery admission status, z_i and its interaction with grade level, $\boldsymbol{z_i G_{it}}$, instrument kindergarten enrollment in an immersion program in the district, DLI_i^{kg}, and its interaction with grade level, $\boldsymbol{\widehat{DLI}_i^{kg}} * \boldsymbol{G_{it}}$. In the second stage (equation 5), the estimated values of $\widehat{DLI}_i^{kg}$ and $\boldsymbol{\widehat{DLI}_i^{kg}} * \boldsymbol{G_{it}}$ from equation 4 become the treatment variable in predicting student achievement. In practice, the first and second stages are estimated simultaneously. Because z_i is randomly assigned, it is presumed to be unrelated to y_{it} except through its effect on DLI program participation, thereby satisfying the exclusion restriction assumption of instrumental variables estimation (Angrist & Pischke, 2008; Imbens & Angrist, 1994). The monotonicity assumption, which specifies that the relationship between z_i and DLI_i^{kg} is positive for all *i*, is also likely satisfied, since randomly-assigned lottery status largely regulates students' access to immersion programs. In this context, the parameters of interest, τ_5 and β_5, represent the precision-weighted unbiased effects of immersion enrollment in kindergarten on the outcomes of lottery compliers.

Results

To facilitate interpretation, we present our results in Figures 2 through 5, where the data points represent immersion-effect coefficients by grade level. We use solid data markers to represent coefficients that are statistically distinguishable from zero at the 5-percent level, and hollow markers to indicate those that are not. For readers who wish to see the coefficients and their standard errors in tabular form, they are reported in the technical appendix available online.

<Figure 2 about here>

Full sample. Figure 2 presents full-sample, ITT, and IV estimates for reading (left panel) and for math and science (right panel). The full-sample estimates (solid line) pertain to all pre-k and kindergarten entrants to the district during the 2004-05 through 2010-11. Even though these estimates are not based on a randomized sample, they shed light on the causal immersion effect in a couple of ways. First, if selection bias favors immersion students, such that the families in the district who enroll in immersion programs are more motivated or well-informed than other such families, then the full-sample estimates represent a plausible upper bound on the causal effect of immersion education. Moreover, though the full-sample estimates are compromised from the perspective of internal validity, they have advantages from an external validity perspective, because they include students at one immersion school that does not participate in the lottery, as well as applicants to immersion lottery categories that were undersubscribed or that were too low in priority to have available slots. The full sample analysis, in other words, is more inclusive, but also more vulnerable to selection bias.

Examining the full-sample estimates in Figure 2, we see large, positive, and statistically significant estimates in reading, mathematics, and science (solid triangles) at each observed grade level. In reading, estimates range from nearly a tenth of a standard deviation in grade 3 to about a fifth of a standard deviation in grade 8. In mathematics, immersion students outperform their peers by 12% to 31% of a standard deviation, depending on grade level, and in science, they outperform by 14% to 27% of a standard deviation. The question is whether these are substantiated in the more-rigorous ITT analysis.

Intent-to-treat. Turning to the ITT estimates in the lottery sample, which represent the effects of random assignment to an immersion program before kindergarten, we find test score coefficients that are smaller in magnitude than the full-sample estimates and that are statistically distinguishable from zero in only a few cases, suggesting upward bias in the full-sample estimates. In reading, find evidence of positive effects that increase over time. In grade 5, lottery winners outperform their counterparts by 13% of a standard deviation, and they do so by 22% of a standard deviation in grade 8—both of which are statistically significant at the 5% level, as well as substantively meaningful. The fifth grade effect translates to about 7 months of student learning in the fifth grade sample, and the eighth grade effect translates to about 9 months, or nearly a full academic year of learning in English language arts.

We find less evidence of immersion effects in mathematics or science. Though the ITT estimates are generally positive, they are noisy and not distinguishable from zero except in grade 4 mathematics, where the positive estimate, 10 percent of a standard deviation, is marginally significant at the 10% level.

Instrumental variables. By scaling the ITT estimates to reflect treatment status compliance rates, our instrumental variables analysis provides a causal estimate of the treatment effect for compliers. The direction and statistical significance of the IV estimates reflect those of the corresponding ITT estimates, but the magnitude of the IV estimates is greater because they partial out individuals who do not adhere to their randomly assigned status. Though mathematics effects are still non-significant, and we lack sufficient data points for IV estimation of science effects, the estimates for reading and exit from EL status are substantial, with significant or marginally significant estimates from nearly a fifth of a standard deviation in grade 3 to half a standard deviation in grade 8. Because IV estimates have less precision and stability than ITT estimates, we focus conservatively on ITT estimates in our discussion of disaggregated subgroup effects in the next section.

Differential Effects by Program Type and Native Language

In response to research question 2, Figure 3 presents ITT estimates for the randomized sample, disaggregated by whether the applicant's first-choice program is a one-way or two-way immersion program (top row) and by whether it is a Spanish program or program in Mandarin, Japanese, or Russian (bottom row). For each outcome variable, the dotted line represents the main effect for the category coded as default (two-way, or Spanish), whereas the solid line represents the estimate for the interaction category (one-way, or other languages). The 95%

confidence interval in each panel pertains to the interaction category; if the default category falls within it, then there is no significant difference between estimates for the two categories.

<Figure 3 about here>

In practice, the two-way and Spanish indicators are nearly collinear. All but one of the two-way programs in the sample were Spanish programs during the study years (the other was Russian), and all but one of the one-way programs focused on Mandarin or Japanese. However, comparing the estimates for two-way versus one-way against the estimates for Spanish versus other languages provides some indication of whether any differential program effects are associated with the program's language or its instructional model. In reading, we find almost no estimated differences between two-way and one-way programs, and a slightly larger difference between Spanish and other-language programs, though none of these differences are statistically significant at the 5% level. In mathematics, we find the reverse, with mostly non-significant but often positive differential effects for non-Spanish languages.

<Figure 4 about here>

Addressing research question 3, Figure 4 disaggregates the ITT effects by whether the student's native or home language is English (top row) and by whether the student's native or home language matches the partner language (bottom row) of their first-choice program. Examining effects for native English speakers versus native speakers of other languages, we find statistically significant interactions only in eighth grade mathematics, where ITT immersion effects for native English speakers are about two-fifths of a standard deviation higher (0.2 vs. -0.2) than for native speakers of a language other than English. This would be a finding of some concern, except that the randomized sample of non-native English speakers observable to grade 8 is quite small, so it may be unstable even though the difference is statistically significant.

Finally, we estimate ITT effects for students who are native speakers of the partner language versus those who are not. Because native speakers of the partner language have lottery preferences in some schools, the randomized analytic sample for this group is small (184 students), but the estimates are instructive nevertheless. The reading estimates for native speakers of the partner language suggest that they benefit from immersion to the same extent, if not modestly (and non-significantly) more than other immersion students, though they do show a negative trend in math relative to other immersion students beyond fourth-or-sixth grade.

EL Classification Rates

Regarding the probability of EL classification in each year, controlling for EL status at baseline, our full-sample estimates in Figure 5 (left panel) suggest that immersion students are roughly two percentage points more likely to remain classified as ELs in grades 1 through 4, after which their probabilities are mostly indistinguishable from those of non-immersion students. In the ITT analysis, we find some evidence that, controlling for baseline EL status, students randomly assigned to immersion have similar rates of EL classification as those randomly assigned to non-immersion programs until grades 6 and 7, at which point their estimated probabilities of remaining in EL are respectively 3 percentage-points and 4 percentage-points lower than those of their non-immersion peers.

<Figure 5 about here>

Even with a small number of randomized native speakers of the partner languages to which they applied, we find large and significant interaction effects in terms of persistence in EL status in nearly every grade. Through grade 3, native speakers of the partner language remain more likely to be classified as ELs in a given year, but by fifth and sixth grades, their probabilities are 6 and 14 points lower, respectively, than those of native speakers of the partner language who did not win immersion slots. Their exit rates are markedly faster than for lottery winners who are EL and are non-native speakers of the partner language (about 93 students, most of whom are Vietnamese speakers who applied to Spanish programs or speakers of non-Mandarin Chinese dialects who applied to Mandarin programs).

Discussion and Conclusion

Our study contributes to the immersion literature in several key ways. First, it provides longitudinal, causal estimates of immersion education on both native English speakers and native speakers of other languages, finding similar effects for both groups. Specifically, we find that students randomly assigned to immersion outperform their peers on state accountability tests in reading by about seven months of learning in grade 5, and nine months of learning in grade 8. Examining mathematics and science scores, we find no statistically significant immersion benefit, but also no detriment. This is important given that students study mathematics and science at least partially in the partner language through grade 5. Moreover, our study includes a fairly wide array of immersion programs, allowing us to generalize beyond any single school or program, and to disaggregate estimates for two-way versus one-way models and Spanish programs versus programs in Mandarin, Japanese, and Russian. The fact that we find a *slightly* larger Spanish-program advantage than two-way program advantage suggests that impacts may vary more by language than by two-way versus one-way models, though this distinction is quite speculative.

What is clear is that among students randomly assigned to immersion, those whose native language matches the partner language show a 6-percentage-point reduction in the probability of being classified as an EL as of about fifth grade, and a 14-point reduction in sixth grade. This finding corroborates other research showing an immersion advantage in EL reclassification beyond the early grades.

Of course, the limitations of this research are important to bear in mind. First, though our ITT estimates are aggregated across numerous immersion schools and programs in Portland, they are still generalizable only to families who apply to an immersion lottery. It is possible that if we were to randomly assign students whose families had shown no interest in dual-language learning, results could differ from those reported here.

In addition, the mechanism by which immersion programs drive achievement are not entirely clear, and our research design cannot fully disentangle the effects of dual-language instruction itself from other possible mechanisms, such as differences in peer composition or teacher quality. In fact, one rationale for placing EL students in two-way immersion programs rather than transitional bilingual classes is that two-way immersion integrates them with native English speakers while also supporting their native language development (Collier & Thomas, 2004). In Portland, students who win immersion slots may change not only their classroom placement but the school they attend, and it is possible that features of immersion schools differ in key ways (such as culture of learning or parent involvement) that classroom-level teacher and peer attributes do not capture. It is also possible that simply moving to a classroom in which most peers are lottery applicants yields a different level of peer motivation than one would find in control-group classes.

Because the policy implications of this work depend to some extent on the mechanisms, our online appendix Table A5 includes an exploratory instrumental variables analysis in which we estimate the effect of lottery winning on the peer, class size, and teacher characteristics of our ITT sample in 2012-13, as well as the extent to which these lottery-driven environmental effects predict reading scores. As expected, we find modest differences for lottery winners and their counterparts in the share of class peers who are English learners, special education eligible, Hispanic, black, and white, and we find that their teachers are slightly less experienced and less likely to be highly qualified under No Child Left Behind. Still, we find no evidence that these small differences drive reading outcomes. Nevertheless, peer and teacher attributes that matter most may not be captured in administrative data. Our study is designed to test the causal effect of access to immersion in Portland, and we acknowledge that the treatment yields access not only to instruction in two languages, but also to teachers and peers who have been drawn to that instructional model.[8] If dual-language immersion were scaled very widely—say, to all schools in a city—then this pattern would no longer hold. Moreover, rapid scaling without provisions to ensure quality might attenuate the treatment effect even if dual-language instruction is the critical mechanism.

The lesson for policymakers pursuing path-breaking 21st century reform is that language immersion may benefit students' English reading skills from mid-elementary school and

[8] Similar challenges in distinguishing mechanisms affect most causal studies of school choice programs, including studies of random assignment to charter or private schools (Abdulkadiroglu et al., 2009; Angrist, Bettinger, Bloom, King, & Kremer, 2002; Hoxby & Murarka, 2009; Krueger & Zhu, 2004).

enhance English learning for ELs. Though effects in mathematics and science are less evident, a program that yields improved reading in English, improved long-term exit rates from EL status, no apparent detriment to mathematics and science skills, and promotes proficiency in two languages seems difficult to criticize. Of course, as with any promising reform, efforts to scale beyond the level adopted by Portland would entail many logistical and staffing challenges, and the promise of immersion may be squandered if efforts are not put in place to ensure program quality. Moreover, promoting equitable access to these programs seems critical, not only to protect the integrity of two-way models, but also to ensure that academic benefits are fairly distributed within a community. If some degree of multilingualism can be achieved while enhancing students' reading skills in English, then it is conceivable that expanding access to language immersion from early childhood could become the next frontier in the struggle for educational opportunity in America.

References

Abdulkadiroglu, A., Angrist, J. D., Cohodes, S., Dynarski, S., Fullerton, J., Kane, T. J., & Pathak, P. (2009). *Informing the debate: Comparing Boston's charter, pilot, and traditional schools.* Technical report, The Boston Foundation, Boston, MA.

Alloway, T. P. (2007). *Automated working memory assessment (AWMA)*. Retrieved from http://www.psychcorp.co.uk/product.aspx?n=1343&s=1492&cat=1356&skey=3909&mode=Product%20home%20and%20price%20list

Anghel, B., Cabrales, A., & Carro, J. M. (2012). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning.

Angrist, J. D., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review, 92*(5), 1535-1558.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Avant Assessment. (2015). *STAMP 4S benchmark and rubric guide*. Retrieved from Eugene, OR: http://avantassessment.com/stamp4s.html

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.

Barnett, W. S., Yarosz, D., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly, 22*(3), 277-293.

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

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 859-873

Caldas, S. J., & Boudreaux, N. (1999). Poverty, race, and foreign language immersion: Predictors of math and English language arts performance. *Learning Languages, 5*(1), 4-15.

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/

Coderre, E. L., & van Heuven, W. J. B. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology, 5*, 1070.

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.

D'Amico, A., & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences, 15*(3), 189-202.

Esposito, A. G., & Baker-Ward, L. (2013). Dual-language education for low-income children: preliminary evidence and benefits for executive function. *Bilingual Research Journal, 36*(3), 295-310.

Francis, D. J., Lesaux, N., & August, D. (2006). Language of instruction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-langauge learners*: Report of the National Literacy Panel on language-minority children and youth, 365-413. Mahway, NJ: Lawrence Erlbaum Associates.

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.

Greene, J. P. (1997). A meta-analysis of the Rossell and Baker review of bilingual education research. *Bilingual Research Journal, 21*(2-3), 103-122.

Hoff, E., Core, C., Place, S., Rumiche, R., Senor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of child language, 39*(1), 1-27.

Hoxby, C. M., & Murarka, S. (2009). Charter schools in New York City: Who enrolls and how they affect their students' achievement. NBER Working Paper No. 14852. *National Bureau of Economic Research.*

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*(2), 467-475.

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*.

Krueger, A. B., & Zhu, P. (2004). Another look at the New York City School voucher experiment. *American Behavioral Scientist, 47*(5), 658-698.

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.

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.

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

Maxwell, L. A. (2012, March 28). 'Dual' classes see growth in popularity. *Education Week, 31*(1), 16-17.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947-959.

New York City Department of Education. (2015). Chancellor Fariña to launch 40 dual language programs in September. Retrieved from http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2014-2015/Chancellor+Fari%C3%B1a+to+Launch+40+Dual+Language+Programs+in+September.htm

North Carolina Department of Education. (2014). North Carolina dual language/immersion programs: 2014-2015. Retrieved from https://drive.google.com/file/d/0B0K1-ZK0_vg5cktjbjBiYXNva28/view

Paap, K. R., Darrow, J., Dalibar, C., & Johnson, H. A. (2014). Effects of script similarity on bilingual advantages in executive control are likely to be negligible or null. *Frontiers in Psychology, 5*, 1539.

Padilla, A. M., Fan, L., Xu, Z., & Silva, D. (2013). A Mandarin/English two-way immersion program: Language proficiency and academic achievement. *Foreign Language Annals, 46*(4), 661-679.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). London: Sage Publications.

Rossell, C. H., & Baker, K. (1996). The effectiveness of bilingual education. *Research in the Teaching of English, 30*, 7-74.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701.

Schneider, A. (2013). 29 new dual language programs to open. Retrieved from http://insideschools.org/blog/item/1000679-29-new-dual-language-programs-to-open-in-fall

Schultz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine, 8(1),18. Double check journal.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247-284.

Slavin, R. E., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2010). Reading and language outcomes of a five-year randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis, 33*(1), 47-58. Retrieved from http://epa.sagepub.com/content/33/1/47.short

Thomas, W. P., & Collier, V. P. (2011). *English learners in North Carolina dual language programs: Year 3 of this study: School year 2009-10*. Raleigh, NC: North Carolina Department of Public Instruction.

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

U.S. Department of Education. (2014). *The condition of education (Indicator: English language learners)*. Retrieved from http://nces.ed.gov/programs/coe/indicator_cgf.asp

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.

Utah State Office of Education. (2014). Utah dual language immersion: Providing a world of opportunities for students. Retrieved from http://www.schools.utah.gov/CURR/dualimmersion/Home/UtahDLISchools.aspx

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,* 0162373715573310.

Watanabe, T. (2011, May 8). Dual-language immersion programs growing in popularity. *Los Angeles Times*. Retrieved from http://articles.latimes.com/2011/may/08/local/la-me-bilingual-20110508

What Works Clearinghouse. (2014). *Procedures and standards handbook (version 3.0)*. Washington, DC: Institute of Education Sciences, U. S. Department of Education.

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*(3), 269-317.

Table 1. Summary of Portland Public Schools immersion programs in the study

Program Type	Native Language of Students	% of Instruction in Partner Language	Languages	Schools (Elem, Middle, High)	Students in 2012-13 (and % of total)
90/10 Two-Way	≈ ½ English ≈ ½ Partner Language	90% in Grade K 80% in Grade 1 70% in Grade 2 60% in Grade 3 50% in Grades 4- 5 2 periods in MS 1-2 periods in HS	Spanish	7 ES 3 MS 2 HS	1,644 (42.6%)
90/10 Two-Way			Russian	1 ES	193 (5.0%)
50/50 One-Way	Mostly English *(no native speaker set-aside slots)*	50% in Gr. K-5 2 periods in MS 1 period in HS	Spanish	1 ES 1 MS 1 HS	614 (16.0%)
			Japanese	1 ES 1 MS 1 HS	920 (23.8%)
			Mandarin	1 ES 1 MS 1 HS	489 (12.7%)

Table 2. Descriptive statistics for applicants to binding lottery strata who are observed in the analysis, and for all kindergarten entrants to the district in the same cohort (proportions are within column)

	Binding Lottery Applicants Only					All Kindergarten Entrants to PPS					
Variable	**All**	**Won Slot**	**Not Placed**	**Differ-ence (unadj)**	**p Diff (strata-adj)**	**All**	**DLI in K**	**Non-DLI in K**	**Differ-ence**	**p Diff**	***Pooled SD***
N	1,625	752	873			27,741	2,500	25,241			
Proportion		0.463	0.537				0.090	0.910			
Female	0.529	0.508	0.546	-0.038	0.15	0.498	0.543	0.493	0.050	0.00	*0.500*
Asian	0.144	0.178	0.115	0.064	0.61	0.098	0.134	0.094	0.039	0.00	*0.297*
Black	0.056	0.052	0.060	-0.008	0.77	0.133	0.044	0.142	-0.098	0.00	*0.340*
Hispanic	0.170	0.177	0.164	0.013	0.65	0.157	0.296	0.143	0.153	0.00	*0.364*
White	0.540	0.517	0.559	-0.042	0.25	0.548	0.451	0.558	-0.107	0.00	*0.498*
Other Race	0.068	0.063	0.073	-0.011	0.01	0.042	0.060	0.040	0.020	0.00	*0.201*
FARMS	0.260	0.273	0.250	0.023	0.63	0.248	0.288	0.244	0.044	0.00	*0.432*
Sp. Needs in K	0.041	0.052	0.032	0.020	0.29	0.086	0.057	0.089	-0.032	0.00	*0.281*
Gifted in K	0.040	0.044	0.037	0.007	0.63	0.030	0.033	0.029	0.004	0.25	*0.169*
EL in K	0.127	0.153	0.105	0.048	0.91	0.161	0.241	0.153	0.088	0.00	*0.368*
First Lang Not Eng.	0.178	0.206	0.153	0.053	0.58	0.171	0.291	0.159	0.131	0.00	*0.377*
First Lang Partner	0.063	0.092	0.038	0.054	0.01	0.020	0.218	-	0.218	0.00	*0.139*
Ns By Grade											
Grade K	1,625	752	873			27,741	2,500	25,241			
Grade 1	1,625	752	873			25,189	2,476	22,713			
Grade 2	1,625	752	873			23,620	2,437	21,183			
Grade 3	1,589	729	860			21,810	2,286	19,524			
Grade 4	1,254	570	684			17,776	1,861	15,915			
Grade 5	983	428	555			13,837	1,429	12,408			
Grade 6	690	289	401			10,176	1,015	9,161			
Grade 7	517	196	321			7,192	663	6,529			
Grade 8	343	123	220			4,562	424	4,138			
Grade 9	179	56	123			1,977	192	1,785			

Notes: For the binding lottery subgroup, p-values reflect balance within randomization strata. Ns by grade in the analytic sample reflect not only attrition but the fact that cohorts are observed for different lengths of time.

Figure 1. CONSORT Sample Attrition Diagram

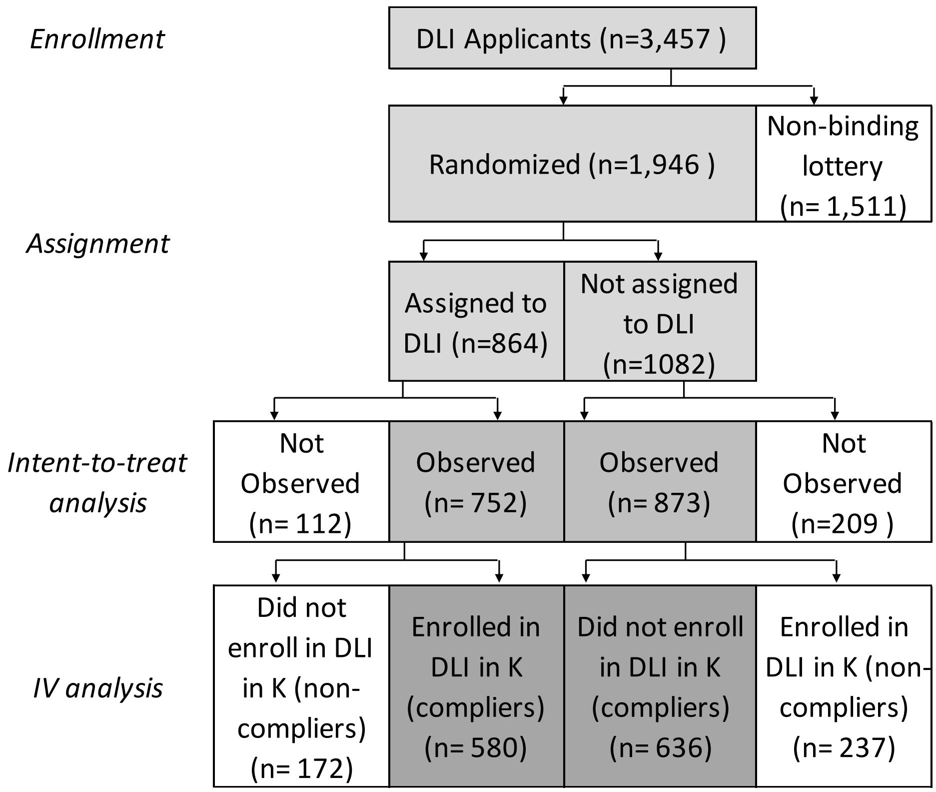
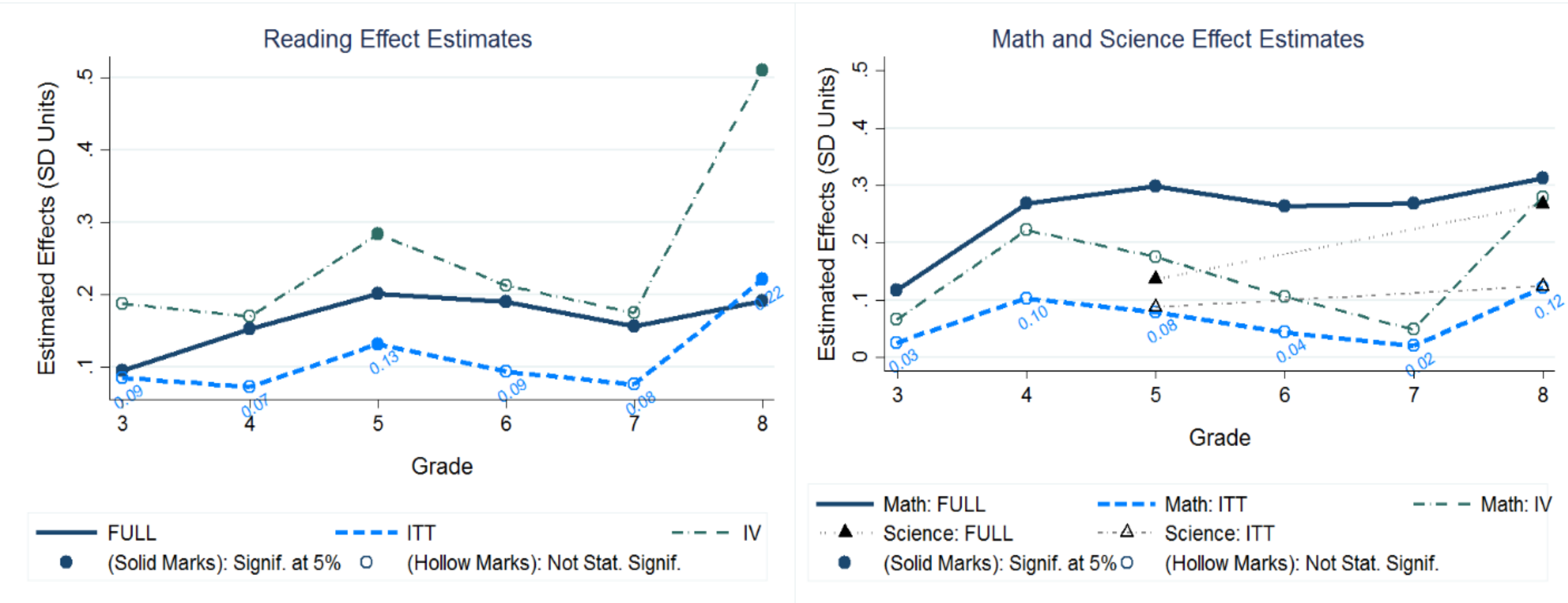
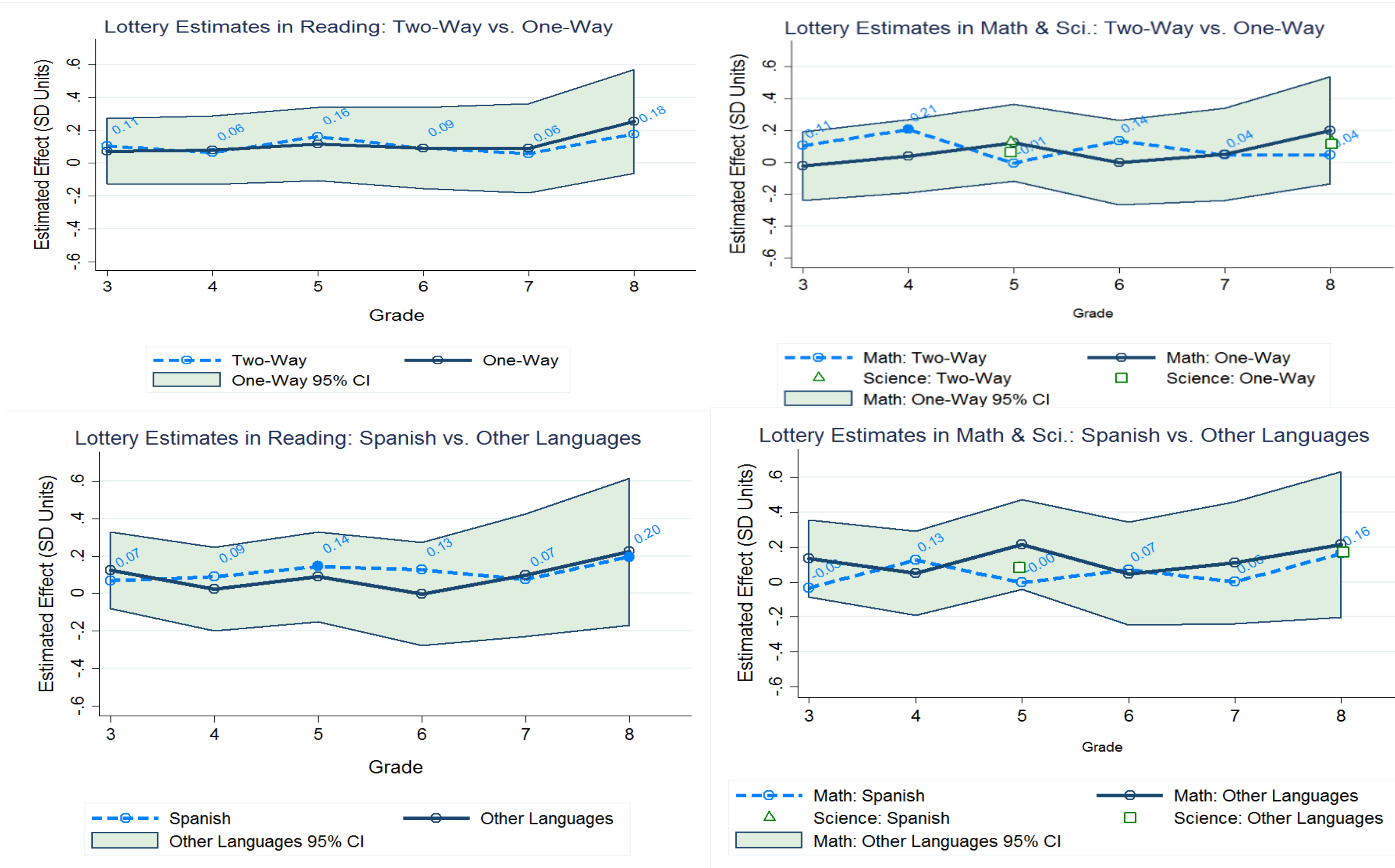


Figure 2. Estimated full-sample, intent-to-treat, and instrumental variable immersion effects in reading, mathematics, and science



Note for Figures 2-4: n=1,451 students and 4,608 observations in reading; n=1,447 students and 4,632 observations in math; n=822 students and 1,059 observations in science.

Figure 3. Estimated intent-to-treat immersion effects in two-way versus one-way and Spanish versus other language programs



Note for Figures 3 and 4: The 95% confidence interval (CI) pertains to the program-type interaction effect, represented by the solid line. When the dotted-line main effect falls within the solid-line CI, this indicates no statistically significant differences between the two program types. A solid marker on a dotted (main-effect) line indicates that the main effect is statistically different from 0.

Figure 4. Estimated intent-to-treat immersion effects for native English speakers and native speakers of other languages (top row), and for students whose native language does and does not match the partner language (bottom row)

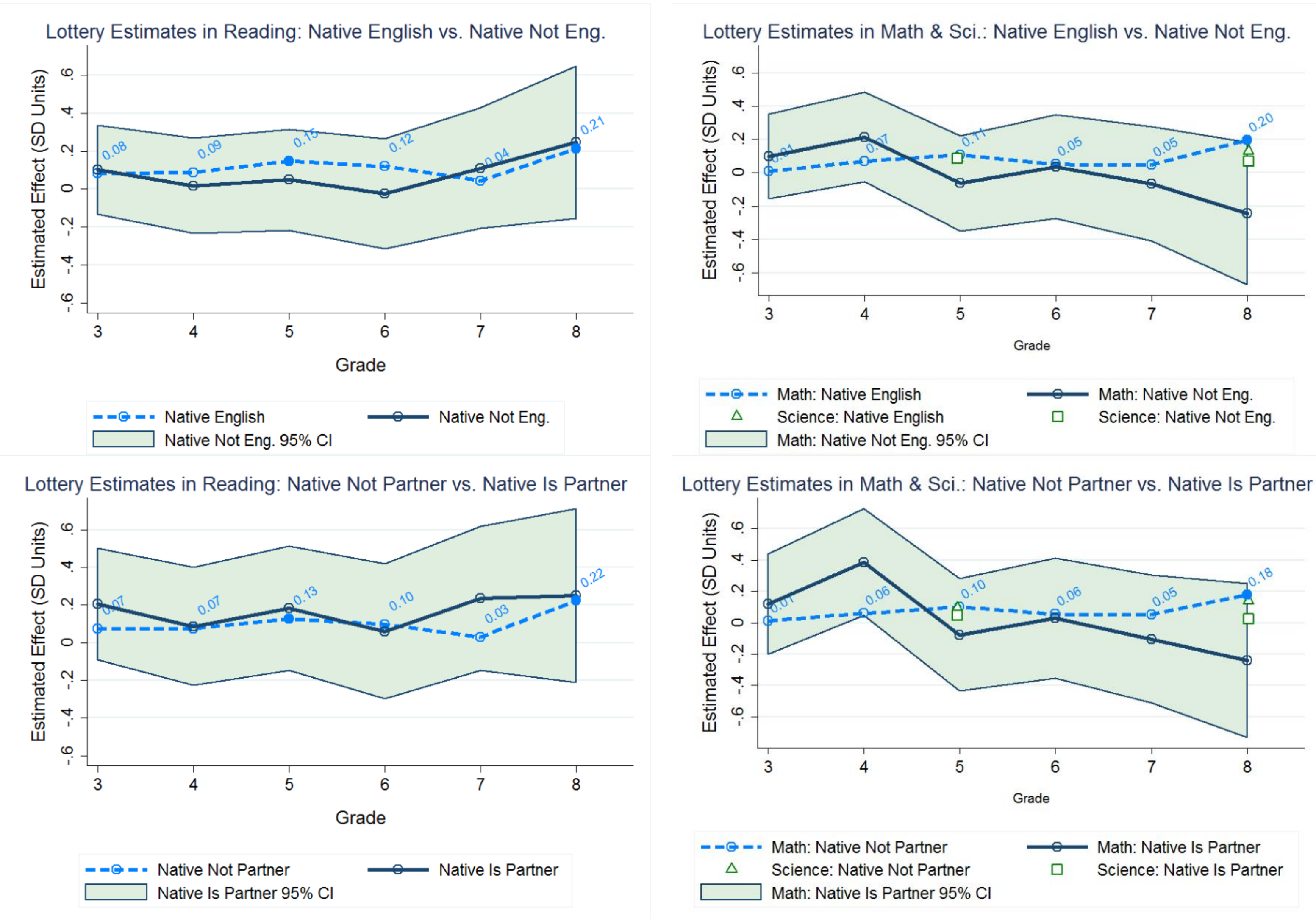
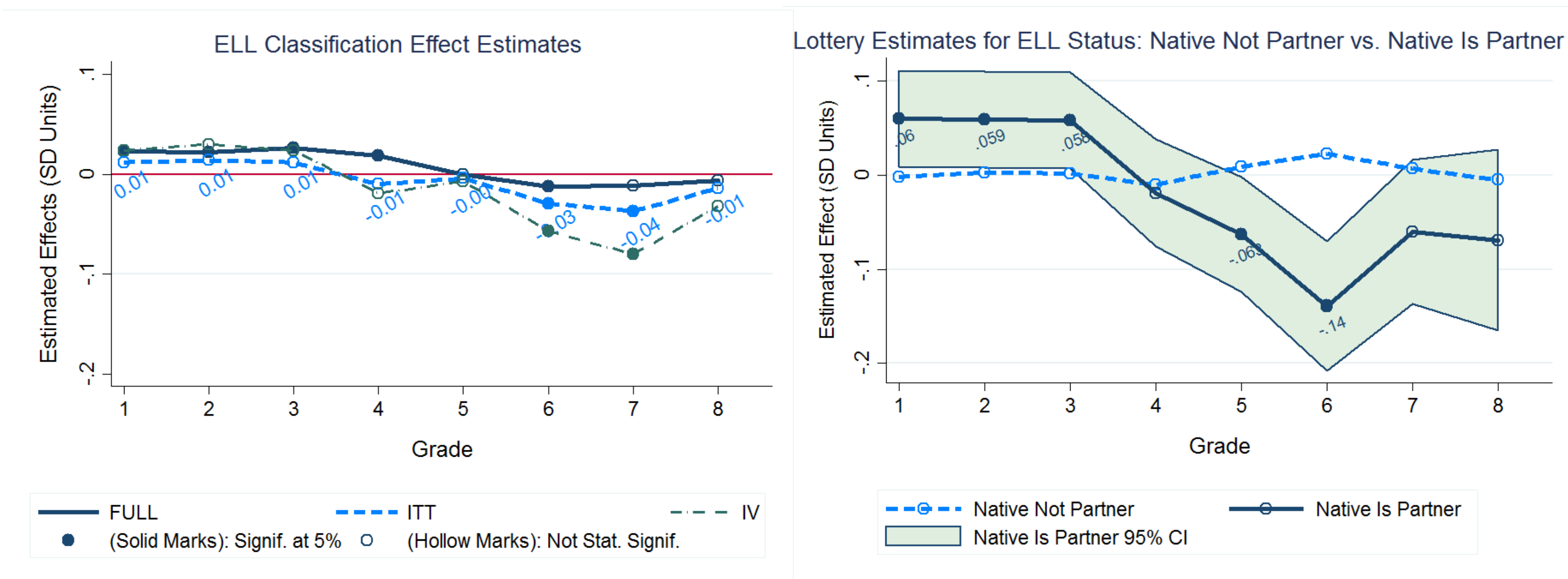


Figure 5. Estimated effects of immersion on probability of EL classification in each grade beyond kindergarten



Note for Figure 5: The full-sample model (left panel) includes 25,189 students and 126,139 observations. ITT (and IV) models in both panels include 1,625 students (184 with native languages matching the partner language) and 8,805 student-by-time observations.

Appendix

Figure A1. Immersion Program Logic Model

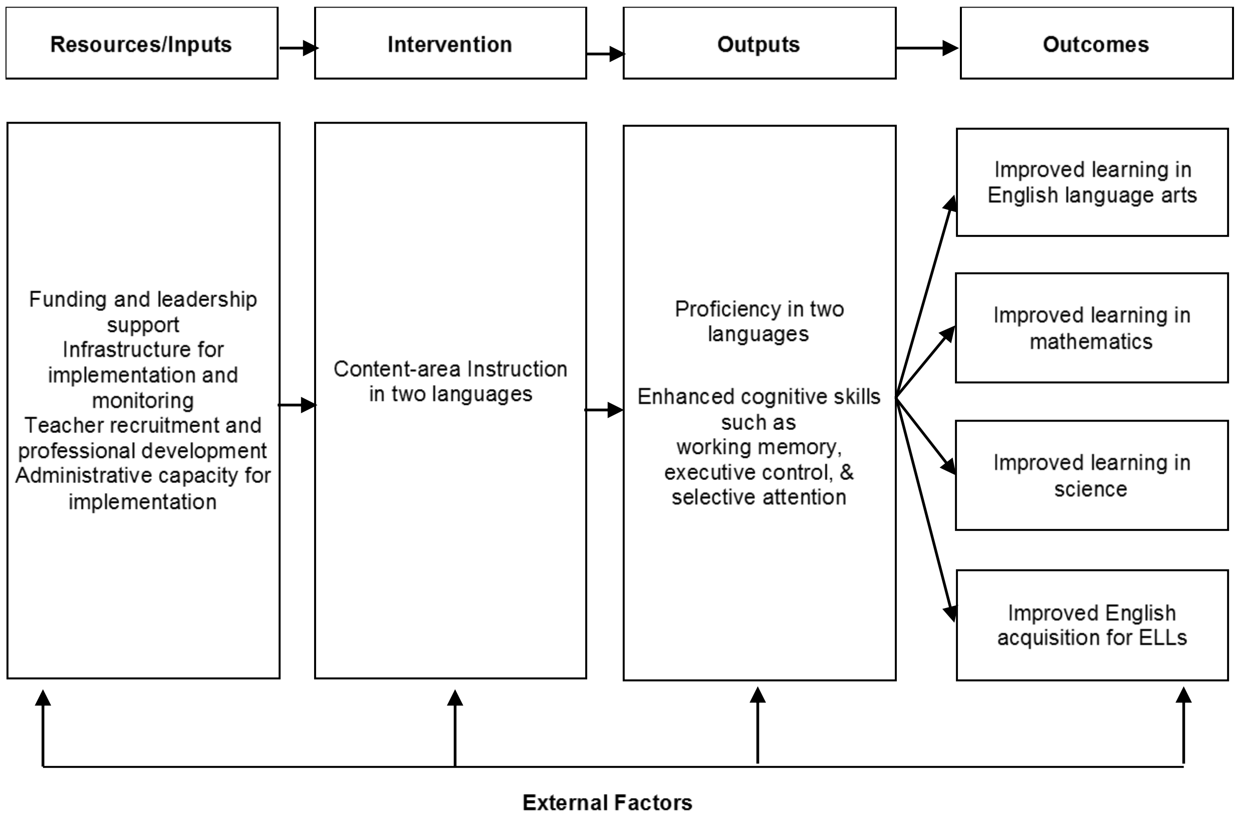


Table A1. Coefficients (and standard errors) for immersion effects on the dependent variables of interest, using full-sample, intent-to-treat, and instrumental variable models

	Reading			**Mathematics**			**Science**		**EL Classification**		
Grade	*Full Samp*	*ITT*	*IV*	*Full Samp*	*ITT*	*IV*	*Full Samp*	*ITT*	*Full Samp*	*ITT*	*IV*
1									0.023***	0.012	0.024
									(0.004)	(0.009)	(0.019)
2									0.022***	0.014	0.030
									(0.004)	(0.009)	(0.019)
3	0.095***	0.085~	0.188~	0.117***	0.026	0.066			0.026***	0.012	0.024
	(0.020)	(0.049)	(0.103)	(0.021)	(0.053)	(0.108)			(0.004)	(0.009)	(0.019)
4	0.153***	0.073	0.17	0.269***	0.103~	0.222~			0.019***	-0.010	-0.019
	(0.022)	(0.052)	(0.113)	(0.022)	(0.056)	(0.118)			(0.004)	(0.010)	(0.021)
5	0.202***	0.132*	0.284*	0.298***	0.079	0.176	0.136***	0.087	0	-0.004	-0.007
	(0.023)	(0.055)	(0.118)	(0.024)	(0.060)	(0.123)	(0.026)	(0.065)	(0.005)	(0.011)	(0.023)
6	0.190***	0.094	0.213	0.263***	0.044	0.106			-0.012*	-0.029*	-0.057*
	(0.025)	(0.062)	(0.134)	(0.026)	(0.066)	(0.140)			(0.006)	(0.013)	(0.027)
7	0.157***	0.076	0.175	0.268***	0.02	0.049			-0.011	-0.037*	-0.080*
	(0.029)	(0.068)	(0.155)	(0.030)	(0.073)	(0.161)			(0.007)	(0.015)	(0.033)
8	0.192***	0.221**	0.510**	0.313***	0.122	0.28	0.267***	0.124	-0.006	-0.014	-0.032
	(0.034)	(0.079)	(0.185)	(0.035)	(0.085)	(0.190)	(0.039)	(0.097)	(0.009)	(0.018)	(0.041)
Observations	70,586	4,608	4,608	70,730	4,632	4,632	16,518	1,059	126,139	8,805	8,805
Students	21,057	1,451	1,451	21,034	1,447	1,447	12,622	822	25,189	1,625	1,625

~p<.10 * p<.05 **p<.01 ***p<.001

Note: The sample size does not permit IV estimation for science scores.

Table A2. Main effect and interaction coefficients and net treatment effect estimates for two-way versus one-way programs

Grade	Reading			Mathematics			Science			EL Classification		
	ITT (for Two-way)	*ITT* One-way*	*Net ITT Effect for One-way*	*ITT (for Two-way)*	*ITT* One-way*	*Net ITT Effect for One-way*	*ITT (for Two-way)*	*ITT* One-way*	*Net ITT Effect for One-way*	*ITT (for Two-way)*	*ITT* One-way*	*Net ITT Effect for One-way*
1										0.029~	-0.029	0
										(0.015)	(0.019)	
2										0.02	-0.011	0.009
										(0.015)	(0.019)	
3	0.106	-0.032	0.074	0.105	-0.128	-0.023				0.017	-0.011	0.006
	(0.080)	(0.101)		(0.087)	(0.109)					(0.015)	(0.019)	
4	0.064	0.016	0.08	0.206*	-0.167	0.039				-0.014	0.001	-0.013
	(0.085)	(0.107)		(0.091)	(0.116)					(0.016)	(0.021)	
5	0.163~	-0.046	0.117	-0.005	0.127	0.122	0.129	-0.066	0.063	-0.023	0.028	0.005
	(0.091)	(0.114)		(0.097)	(0.123)		(0.106)	(0.133)		(0.018)	(0.023)	
6	0.092	0	0.092	0.137	-0.139	-0.002				-0.079***	0.104***	0.025
	(0.098)	(0.126)		(0.105)	(0.135)					(0.020)	(0.027)	
7	0.059	0.031	0.09	0.045	0.006	0.051				-0.053*	0.056~	0.003
	(0.105)	(0.139)		(0.112)	(0.148)					(0.023)	(0.030)	
8	0.178	0.077	0.255	0.045	0.156	0.201	0.134	-0.014	0.12	-0.014	0.01	-0.004
	(0.122)	(0.161)		(0.130)	(0.171)		(0.148)	(0.195)		(0.028)	(0.037)	
Obs		4,608			4,632			1,059			8,805	
Students		1,451			1,447			822			1,625	

~p<.10 * p<.05 **p<.01 ***p<.001

Note: "Net effects" column estimates are the sums of the immersion and immersion-by-category coefficients, representing the total estimated ITT effect for the category in the interaction term. We do not show significance stars for the net effects, because they are calculated from the default and interaction coefficient columns. If the interaction coefficients (middle column of each set) are not statistically significant, then the estimated treatment effects for the category in the interaction term cannot be said to differ from the main effect in the left column of the outcome set.

Table A3. Main effect and interaction coefficients and net treatment effect estimates for Spanish versus other immersion programs

Grade	Reading			Mathematics			Science			EL Classification		
	ITT (for Spanish)	*ITT*Other Languages*	*Net ITT Effect for Other Languages*	*ITT (for Spanish)*	*ITT*Other Languages*	*Net ITT Effect for Other Languages*	*ITT (for Spanish)*	*ITT*Other Languages*	*Net ITT Effect for Other Languages*	*ITT (for Spanish)*	*ITT*Other Languages*	*Net ITT Effect for Other Languages*
1										0.023*	-0.047*	-0.024
										(0.011)	(0.019)	
2										0.017	-0.021	-0.004
										(0.011)	(0.019)	
3	0.069	0.055	0.124	-0.032	0.167	0.135				0.015	-0.012	0.003
	(0.059)	(0.104)		(0.064)	(0.113)					(0.011)	(0.020)	
4	0.09	-0.066	0.024	0.126~	-0.076	0.05				-0.007	0.008	0.001
	(0.063)	(0.114)		(0.068)	(0.123)					(0.012)	(0.023)	
5	0.145*	-0.055	0.09	-0.003	0.218~	0.215	0.079	0.003	0.082	-0.018	0.100***	0.082
	(0.066)	(0.122)		(0.071)	(0.131)		(0.076)	(0.143)		(0.013)	(0.026)	
6	0.127~	-0.13	-0.003	0.071	-0.023	0.048				-0.071***	0.237***	0.166
	(0.073)	(0.140)		(0.078)	(0.150)					(0.015)	(0.032)	
7	0.075	0.023	0.098	0.002	0.108	0.11				-0.053**	0.125**	0.072
	(0.078)	(0.167)		(0.083)	(0.178)					(0.016)	(0.041)	
8	0.196*	0.027	0.223	0.164~	0.052	0.216	0.165	0.011	0.176	-0.021	0.031	0.01
	(0.091)	(0.201)		(0.097)	(0.213)		(0.109)	(0.246)		(0.020)	(0.051)	
Obs		4,608			4,632			1,059			8,805	
Students		1,451			1,447			822			1,625	

~p<.10 * p<.05 **p<.01 ***p<.001

See Table A2 note for interpretive information.

Table A4. Main effect and interaction coefficients and net treatment effects for native English speakers versus native speakers of other languages

Grade	Reading			Mathematics			Science			EL Classification		
	ITT (Native English)	*ITT*Native Not English*	*Net ITT Effect for Native Not English*	*ITT (Native English)*	*ITT*Native Not English*	*Net ITT Effect for Native Not English*	*ITT (Native English)*	*ITT*Native Not English*	*Net ITT Effect for Native Not English*	*ITT (Native English)*	*ITT*Native Not English*	*Net ITT Effect for Native Not English*
1										0.005	-0.006	-0.001
										(0.009)	(0.021)	
2										0.005	0.020	0.025
										(0.009)	(0.021)	
3	0.082	0.02	0.102	0.01	0.09	0.1				0.005	0.019	0.024
	(0.053)	(0.120)		(0.058)	(0.130)					(0.009)	(0.021)	
4	0.088	-0.071	0.017	0.07	0.146	0.216				0.005	-0.065**	-0.060
	(0.057)	(0.128)		(0.062)	(0.138)					(0.010)	(0.023)	
5	0.148*	-0.099	0.049	0.109~	-0.171	-0.062	0.087	0.001	0.088	0.006	-0.002	0.004
	(0.061)	(0.136)		(0.065)	(0.146)		(0.070)	(0.157)		(0.011)	(0.024)	
6	0.120~	-0.144	-0.024	0.051	-0.014	0.037				0.006	-0.072*	-0.066
	(0.069)	(0.148)		(0.074)	(0.159)					(0.013)	(0.028)	
7	0.044	0.067	0.111	0.049	-0.114	-0.065				0.005	-0.059~	-0.054
	(0.076)	(0.163)		(0.081)	(0.175)					(0.015)	(0.032)	
8	0.214*	0.033	0.247	0.196*	-0.439*	-0.243	0.135	-0.062	0.073	0.006	-0.058	-0.052
	(0.087)	(0.205)		(0.093)	(0.219)		(0.105)	(0.244)		(0.017)	(0.042)	
Obs		4,608			4,632			1,059			8,805	
Students		1,451			1,447			822			1,625	

~p<.10 * p<.05 **p<.01 ***p<.001

See Table A2 note for interpretive information.

Table A5. Main effect coefficients, interaction coefficients, and net treatment effects for students whose native language differs from the partner language of the first-choice immersion program, versus those whose native language matches it

	Reading			Mathematics			Science			EL Classification		
Grade	*ITT (Native Lang Not Partner)*	*ITT*Native Lang Partner*	*Net ITT Effect for Native Lang Partner*	*ITT (Native Lang Not Partner)*	*ITT*Native Lang Partner*	*Net ITT Effect for Native Lang Partner*	*ITT (Native Lang Not Partner)*	*ITT*Native Lang Partner*	*Net ITT Effect for Native Lang Partner*	*ITT (Native Lang Not Partner)*	*ITT*Native Lang Partner*	*Net ITT Effect for Native Lang Partner*
1										-0.002	0.063*	0.060
										(0.009)	(0.026)	
2										0.003	0.057*	0.059
										(0.009)	(0.026)	
3	0.073	0.132	0.205	0.013	0.106	0.119				0.002	0.057*	0.058
	(0.051)	(0.151)		(0.056)	(0.163)					(0.009)	(0.026)	
4	0.074	0.012	0.086	0.061	0.327~	0.388				-0.011	-0.007	-0.019
	(0.055)	(0.160)		(0.059)	(0.173)					(0.010)	(0.029)	
5	0.126*	0.056	0.182	0.101	-0.177	-0.076	0.094	-0.044	0.05	0.009	-0.071*	-0.063
	(0.059)	(0.169)		(0.063)	(0.182)		(0.068)	(0.199)		(0.011)	(0.031)	
6	0.097	-0.037	0.06	0.055	-0.025	0.03				0.023~	-0.162***	-0.140
	(0.066)	(0.182)		(0.071)	(0.196)					(0.013)	(0.035)	
7	0.03	0.204	0.234	0.053	-0.156	-0.103				0.008	-0.067~	-0.060
	(0.073)	(0.195)		(0.079)	(0.209)					(0.015)	(0.039)	
8	0.222**	0.027	0.249	0.179*	-0.419~	-0.24	0.143	-0.114	0.029	-0.005	-0.064	-0.069
	(0.085)	(0.235)		(0.091)	(0.251)		(0.104)	(0.278)		(0.018)	(0.049)	
Obs		4,608			4,632			1,059			8,805	
Students		1,451			1,447			822			1,625	

~p<.10 * p<.05 **p<.01 ***p<.001

See Table A2 note for interpretive information.

Table A6. Instrumental variables analysis of classroom characteristics that may mediate treatment effects in randomized sample

Panel A. First-stage estimated effect of winning immersion lottery on class characteristics in 2012-13

	First-stage outcomes	Coeff.	Std. Error	n
Proportion of students in class who are:	Subsidized-meal eligible	0.010	(0.013)	728
	English learner	0.018***	(0.005)	1,112
	Special education	-0.030***	(0.006)	1,112
	Talented & gifted	0.000	(0.009)	1,112
	Asian	-0.002	(0.006)	1,112
	Hispanic	0.065***	(0.009)	1,112
	Black	-0.010*	(0.005)	1,112
	White	-0.046***	(0.010)	1,112
	Teacher years of experience	-1.263*	(0.521)	1,076
	Teacher has advanced degree	-0.009	(0.027)	1,071
	Teacher highly qualified under NCLB	-0.018~	(0.011)	1,033
	Students in classroom	-0.300	(0.295)	1,112

Panel B. Second-stage estimated effects of instrumented classroom attributes on reading scores

	Second-stage instrumented predictors	Coeff.	Std. Error	n
Proportion of students in class who are:	Subsidized-meal eligible	3.255	(3.428)	550
	English learner	3.096	(3.570)	847
	Special education	-1.946	(2.135)	847
	Talented & gifted	-18.624	(71.063)	847
	Asian	-10.335	(17.517)	847
	Hispanic	0.812	(0.899)	847
	Black	-8.341	(11.596)	847
	White	-1.340	(1.514)	847
	Teacher years of experience	-0.302	(0.683)	819
	Teacher has advanced degree	-5.504	(11.756)	814
	Teacher highly qualified under NCLB	-14.645	(20.319)	779
	Students in classroom	-0.168	(0.269)	847

~p<.10 * p<.05 **p<.01 ***p<.001

Note: Panel A represents first-stage estimates from instrumental variable models that include lottery strata fixed effects and individual covariates, as in Equation 4. Panel B represents the second-stage IV estimates, as specified in Equation 5.