American Educational Research Journal
April 2017, Vol. 54, No. 1S, pp. 282S–306S
DOI: 10.3102/0002831216634463

Effects of Dual-Language Immersion Programs on Student Achievement: Evidence From Lottery Data

Jennifer L. Steele
American University
Robert O. Slater
American Councils for International Education
Gema Zamarro
University of Arkansas
Trey Miller
Jennifer Li
Susan Burkhauser
RAND Corporation
Michael Bacon
Portland Public Schools

Using data from seven cohorts of language immersion lottery applicants in a large, urban school district, we estimate the causal effects of immersion programs 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% 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 3 to 4 percentage points lower than those of their counterparts. This effect is stronger for ELs whose native language matches the partner language.

Keywords: dual-language immersion, student achievement, English language learners, urban education, language education

Introduction

Dual-language immersion schools, which provide native English speakers and English learners (ELs) with general academic instruction in two languages from kindergarten onward, have shown recent and rapid proliferation in the United States. The Center for Applied Linguistics

(2011a, 2011b) estimates that the number of immersion schools in the United States grew from 278 to 448 between 1999 and 2011, but more recent extrapolations place the latest number between 1,000 and 2,000 (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 192, between the 2012–2013 and 2015–2016 school years (New York City Department of Education, 2015; Schneider, 2013). This proliferation is notable because in contrast to many other parts of the world, U.S. public schools have not traditionally exposed students to a second language in the early grades (Devlin, 2015). Even so, some evidence suggests that the popularity of dual-language immersion is growing internationally as well as in the United States (Tedick, Christian, & Fortune, 2011).

Domestically, this swift expansion of an approach that was recently considered boutique seems driven by several 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 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

Jennifer L. Steele is an associate professor of education at American University, 4400 Massachusetts Ave., NW, Washington, DC 20016-8030; e-mail: *steele@american.edu*. Her research addresses urban education policy, teacher quality, and transitions to college and career.

Robert O. Slater is a senior consultant for policy research at American Councils for International Education, where he studies programs aimed at fostering multilingualism and global citizenship.

Gema Zamarro is the 21st Century Endowed Chair in Teacher Quality at the Department of Education Reform at the University of Arkansas, where she studies applied econometrics in the areas of education, health, and labor.

Trey Miller is an economist at the RAND Corporation. His research focuses on developing, implementing, and evaluating policies and practices to improve student access and success in higher education while also reducing the cost of higher education to students and taxpayers.

Jennifer Li is a management scientist and applied linguist at the RAND Corporation, where she conducts research on language education, language policy, organization development, training, and workforce issues.

Susan Burkhauser is an assistant policy analyst at the RAND Corporation, where she studies school leadership and education policy.

Michael Bacon is the assistant director of dual language for the Portland Public Schools in Oregon, a role in which he develops and supports dual-language immersion programs across the district.

native English speakers who anticipate benefits of bilingualism within a globally competitive society (Maxwell, 2012). The expansion of these programs arrives at a time of rapid social and demographic change in the United States. Between 1980 and 2013, the share of young adults who spoke a language other than English at home more than doubled from 11% to 25% (U.S. Census Bureau, 2014). And recent projections by the Pew Research Center suggest that by 2065, first-generation immigrants and their immediate offspring will together constitute 36% of the U.S. population, versus 26% today (Cohn, 2015).

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, could not fully adjust for unobserved differences between immersion and non-immersion participants. Our study addresses this limitation by capitalizing on a lottery that randomly assigns students—both native English speakers and ELs—to language immersion in the Portland Public Schools (PPS) in Portland, Oregon. PPS is among the largest two public school districts in the Pacific Northwest, and our study represents the largest random-assignment study of dual-language immersion that we are aware of; it also allows us to track students across a diverse array of immersion schools for up to nine years. We find that students randomly assigned to immersion programs 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, and these estimates do not appear to vary by students' native language. Conditional on their EL status at school entry, lottery winners are 3 to 4 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 are indistinguishable from zero in most cases.

In subsequent sections, we discuss prior studies of dual-language immersion programs and explain how immersion is implemented in Portland. We then describe our sample, methods, and results. We conclude with implications for policy in the globalized 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 comprehending written material (Alloway, 2007; Gathercole, Alloway, Willis, & Adams, 2006). 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 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).

Research on academic impacts of dual-language immersion programs can be divided into studies that have focused primarily on native speakers of the cultural majority language (e.g., English in the United States) and those that have focused mainly on students who first arrive at school without fluency in the majority language (e.g., ELs in the U.S. context). The former category includes a few studies that are quite rigorous but small in scale, while the latter category features studies that, due to data availability, have been more vulnerable to selection bias. In the first category, 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 48 randomized participants observed 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 included only 150 students and was able to track students for only one year (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007). In a study of 124 mostly native English speakers in a Mandarin immersion program, Padilla, Fan, Xu, and Silva (2013) demonstrated that immersion students outperformed same-school peers on an English language arts examination in Grades 3 through 5, but though the immersion group was admitted by a randomized lottery, the same-school comparison group was not necessarily randomly assigned. Because all three studies focused on single schools, the extent to which their findings would generalize to larger-scale programs is also unclear. Other studies that have shown benefits of immersion programs for native English speakers in Canadian or U.S. contexts have generally not employed extensive controls for possible selection bias (Barik & Swain, 1978; Caldas & Boudreaux, 1999; Marian, Shook, & Schroeder, 2013; Turnbull, Hart, & Lapkin, 2003).

Meanwhile, 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. Note that for ELs, dual-language immersion serves as a possible alternative to monolingual English instruction and 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 alongside ELs and may therefore segregate ELs less than transitional or developmental bilingual programs. 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; Fortune & Tedick, 2008; Tedick et al., 2011).

It is plausible that dual dual-language programs may exert different learning effects for ELs than for native English speakers. Immersing ELs in their native language for at least part of the school day allows them to receive a substantial share of core academic content instruction in a language they understand, share a classroom with native English speakers, and begin school with a baseline advantage over their monolingual English-speaking peers in terms of knowledge of the partner language. The notion that ELs benefit from school-based instruction in their first language is bolstered by several meta-analyses that have focused not on dual-language immersion programs per se but on the effects of transitional bilingual education programs relative to English-only programs for ELs (Francis et al., 2006; Greene, 1997; Slavin & Cheung, 2005).

Though Valdés (1997) cautions that integrating native speakers of English with native speakers of the partner language may reinforce existing patterns of social inequality, studies that have specifically compared ELs attending dual-language immersion to those attending monolingual English or transitional bilingual programs have generally found outperformance among students in dual-language immersion (Collier & Thomas, 2004; Lindholm-Leary & Block, 2010; Marian et al., 2013; Thomas & Collier, 2015). Historically, these studies have not included many adjustments for baseline between-group differences, rendering them vulnerable to selection bias, but more recently, two studies have used large-scale administrative data with statistical adjustments to mitigate at least observable sources of bias. Specifically, Umansky and Reardon (2014) employed hazard analysis with extensive statistical controls, finding that Latino ELs placed in Spanish immersion classrooms were reclassified from English learner to English-proficient status more slowly in elementary school but at higher rates by high school.[1] Also, 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 the English language arts performance of EL students in all three of the bilingual programs, including dual-language immersion, grew as fast as or faster than their peers in monolingual English programs.

Taken together, the existing research on dual-language immersion education for ELs and native English speakers suggests that families who are able to enroll their children in dual-language immersion programs can expect to see equivalent performance 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 program on native English speakers as well as ELs in the United States and to do so longitudinally between kindergarten and (for the oldest two cohorts) eighth grade. Second, it examines effects at scale in a large urban district, focusing on 12 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 programs 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:

Research Question 1: What is the causal effect of random assignment to a dual-language immersion program 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?

Research Question 2: To what extent do immersion program effects differ for one-way versus two-way immersion programs and for programs in Spanish versus Mandarin, Japanese, and Russian?

Research Question 3: To what extent do immersion program effects depend on whether a student's first language is English and on whether the student's first language matches the partner language?

Our lottery-based design allows us to estimate causal effects based on students' random assignment to immersion programs, but because access to these programs may influence not only students' classroom language exposure but also the teachers and peers with whom students engage, we cannot definitively attribute all program effects strictly to the language of instruction. However, we do report on exploratory mediation analyses in the appendix in the online journal.

Intervention and Setting

Portland Public Schools began implementing dual-language immersion programs in 1986. During the 2012–2013 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. In that year, about 8% 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.

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 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, 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% to 100% of mathematics instruction, 56% to 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

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 (Elementary, Middle, High)	Students in 2012–2013 (and % of total)
90/10 two-way	≈ Half English ≈ Half 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 middle school 1–2 periods in high school	Spanish	7 elementary 3 middle school 2 high school	1,644 (42.6)
			Russian	1 elementary	193 (5.0)
50/50 one-way	Mostly English (no native speaker set-aside slots)	50 in Grades K–5 2 periods in middle school 1 period in high school	Spanish	1 elementary 1 middle school 1 high school	614 (16.0)
			Japanese	1 elementary 1 middle school 1 high school	920 (23.8)
			Mandarin	1 elementary 1 middle school 1 high school	489 (12.7)

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 2012–2013 academic year, we recorded similar distributions of on-task student behavior and instructional strategies across languages (including monolingual English classes), 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 some schools establish multiple preference categories, such as slots for native speakers of the partner language, students who live in the school's catchment neighborhood, and 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 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. 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 is 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.[2] Outcome data are measured through the 2013–2014 academic year, so the oldest cohort can be observed through ninth grade and the youngest through third grade. The lottery applicant sample includes 3,457 students, and we also have data on 24,841 other students who enrolled in the district as pre-kindergarteners or kindergarteners during the years in question.

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 (ODE), we were 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%.[3] Attrition is 13.0% for the treatment sample and 19.3% for the control sample, yielding differential attrition of 6.3 percentage points.[4] 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 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, 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. We use instrumental variables (IV) analyses (Angrist & Pischke, 2008) to recover the effect for those who comply with their random-assignment status.

Sample Characteristics

The left side of Table 2 presents descriptive statistics for the randomized (binding) analytic sample, and the right 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

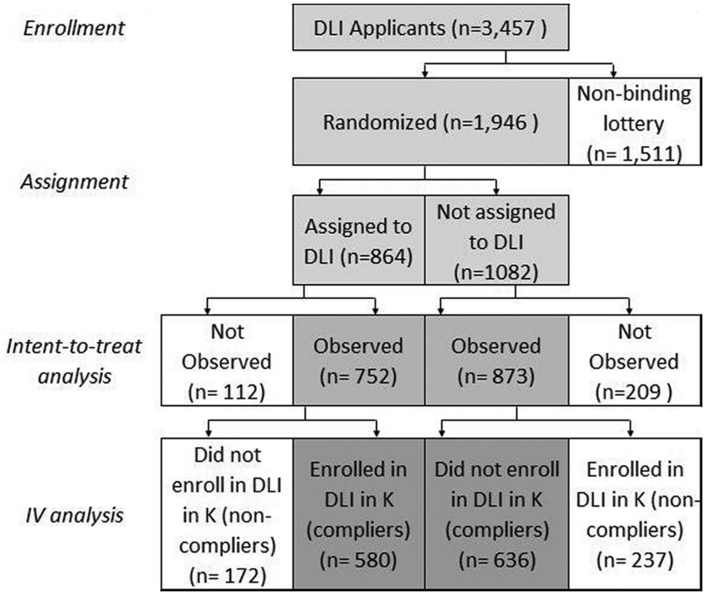

Figure 1. **CONSORT sample attrition diagram.**

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. 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 EL in each academic year after kindergarten, adjusting for his or

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 Portland Public Schools					
Variable	All	Won Slot	Not Placed	Difference (Unadjusted)	*p* Difference (Strata-Adjusted)	All	Dual-Language Immersion in Kindergarten	Non–Dual-Language Immersion in Kindergarten	Difference	*p* Difference	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	0.095	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	0.144	0.152	0.00	0.364
White	0.540	0.517	0.559	–0.042	0.25	0.547	0.451	0.557	–0.106	0.00	0.498
Other race	0.068	0.063	0.073	–0.011	0.01	0.043	0.060	0.041	0.019	0.00	0.202
Subsidized meal-eligible	0.260	0.273	0.250	0.023	0.63	0.248	0.288	0.244	0.045	0.00	0.432
Special needs in kindergarten	0.041	0.052	0.032	0.020	0.29	0.086	0.057	0.089	–0.032	0.00	0.281
Gifted in kindergarten	0.040	0.044	0.037	0.007	0.63	0.029	0.033	0.029	0.004	0.24	0.169
English learner in kindergarten	0.127	0.15.3	0.105	0.048	0.91	0.161	0.241	0.153	0.088	0.00	0.367
First language not English	0.180	0.206	0.157	0.049	0.42	0.173	0.292	0.161	0.131	0.00	0.378
First language partner	0.113	0.138	0.092	0.047	0.92	0.023	0.250	—	—	—	0.149
*N*s by grade											
Kindergarten	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			

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

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 or she no longer qualifies for services based on ELPA scores.[5]

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 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:

$$y_{it} = a_1 + \tau_1 DLI_i^{kg} + \theta_1 \boldsymbol{G_{it}} + \boldsymbol{\beta}_1 \left(\boldsymbol{DLI_i^{kg} G_{it}} \right) + \boldsymbol{\delta}_1 \boldsymbol{X}_i + u_{1i} + \varepsilon_{1it}, \quad (1)$$

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 θ_1. The predictors of interest are the observed value of immersion enrollment in kindergarten, DLI_i^{kg}, and its interaction with grade level, $\boldsymbol{DLI_i}^{kg} \boldsymbol{G_{it}}$. 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.[6]

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 wining 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:

$$y_{it} = a_2 + \tau_2 z_i + \boldsymbol{\theta}_2 \boldsymbol{G_{it}} + \boldsymbol{\beta}_2 (\boldsymbol{z_i G_{it}}) + \boldsymbol{\delta}_2 \boldsymbol{X_i} + \boldsymbol{\gamma}_2 \boldsymbol{L_i} + u_{2i} + \varepsilon_{2it}, \quad (2)$$

where the terms are as described previously, 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 vs. other languages) and 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 ($\boldsymbol{G_{it}}$):

$$\begin{aligned} y_{it} = {} & a_3 + \tau_3 z_i + \boldsymbol{\theta}_3 \boldsymbol{G_{it}} + \boldsymbol{\beta}_3 (\boldsymbol{z_i G_{it}}) + \upsilon_3 c_{it} + k_3 (z_i c_{it}) + \boldsymbol{\phi}_3 (\boldsymbol{G_{it} c_{it}}) + \\ & \boldsymbol{\eta}_3 (\boldsymbol{z_i G_{it} c_{it}}) + \boldsymbol{\delta}_3 \boldsymbol{X_i} + \boldsymbol{\gamma}_3 \boldsymbol{L_i} + u_{3i} + \varepsilon_{3it}. \end{aligned} \quad (3)$$

The key parameters of interest are the coefficients on the treatment assignment by category interaction, k_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 dual-language immersion (DLI) enrollment in kindergarten, specifying a two-stage least squares regression model as follows:

$$DLI_i^{kg} = a_4 + \tau_4 z_i + \boldsymbol{\theta}_4 \boldsymbol{G_{it}} + \boldsymbol{\beta}_4 (\boldsymbol{z_i G_{it}}) + \boldsymbol{\delta}_4 \boldsymbol{X_i} + \boldsymbol{\gamma}_4 \boldsymbol{L_i} + u_{4i} + \varepsilon_{4it} \quad (4)$$

$$y_{it} = a_5 + \tau_5 \widehat{DLI}_i^{kg} + \boldsymbol{\theta}_5 \boldsymbol{G_{it}} + \boldsymbol{\beta}_5 \left(\widehat{\boldsymbol{DLI}}_{\boldsymbol{i}}^{\boldsymbol{kg}} * \boldsymbol{G_{it}} \right) + + \boldsymbol{\delta}_5 \boldsymbol{X_{ij}} + \boldsymbol{\gamma}_5 \boldsymbol{L_j} + u_{5i} + \varepsilon_{5it}. \qquad (5)$$

In the first stage (Equation 4), the randomly assigned lottery admission status, z_i and its interaction with grade level, $\boldsymbol{z_i G_{it}}$, serve as instruments for kindergarten enrollment in an immersion program in the district, DLI_i^{kg}, and its interaction with grade level, $\widehat{\boldsymbol{DLI}}_{\boldsymbol{i}}^{\boldsymbol{kg}} * \boldsymbol{G_{it}}$. In the second stage (Equation 5), the estimated values of $\widehat{DLI}_i^{kg}$and $\widehat{\boldsymbol{DLI}}_{\boldsymbol{i}}^{\boldsymbol{kg}} * \boldsymbol{G_{it}}$ 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% 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 in the online journal.

Full Sample

Figure 2 presents full-sample, intent-to-treat (ITT), and IV estimates for reading (left panel) and math and science (right panel). The full-sample estimates (represented by a solid line) pertain to all pre–k and kindergarten entrants to the district during the 2004–2005 through 2010–2011 academic years. 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

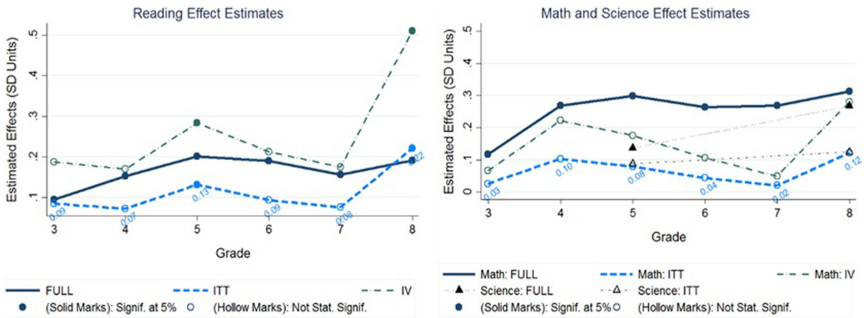

Figure 2. **Estimated full-sample, intent-to-treat, and instrumental variable immersion effects in reading, mathematics, and science.**

Note. 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.

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.

Examining the full-sample estimates in Figure 2, we see large, positive, and statistically significant estimates in reading, mathematics, and science at each observed grade level. In reading, advantages of immersion program entry in kindergarten range from nearly a tenth of a standard deviation in Grade 3 to about a fifth of a standard deviation by 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. Given that the full-sample estimates are somewhat vulnerable to selection on unobserved student and family characteristics, the question of interest is whether these observed effects are substantiated by 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, we 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% 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 assume the program has zero effect on individuals who do not adhere to their randomly assigned status. Though mathematics effects are still nonsignificant 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 net estimate for the interaction category (one-way or other languages). The 95% confidence interval in each panel pertains to the estimated effect for the interaction category; if the line representing the default category falls within that interval, this means there is no significant difference between estimates for the two categories.

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

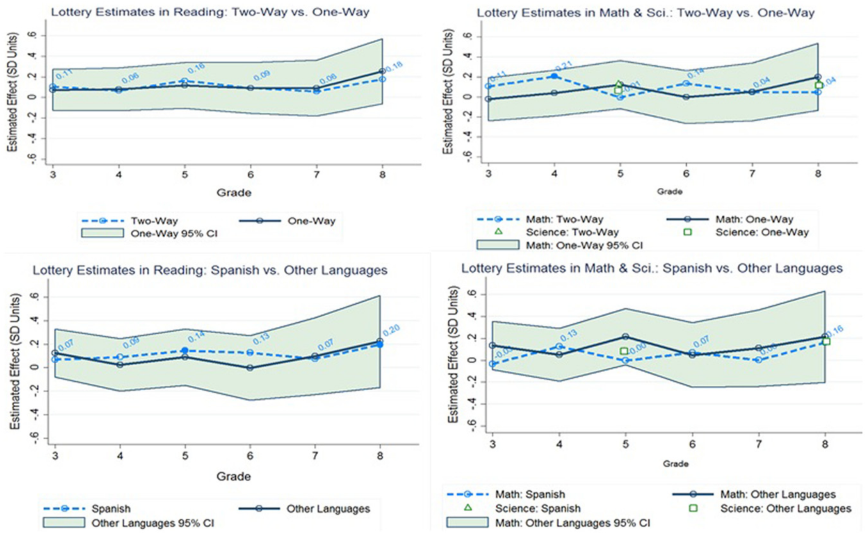

Figure 3. **Estimated intent-to-treat immersion effects in two-way versus one-way and Spanish versus other language programs.**

Note. 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. 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.

model. In reading, we find almost no estimated differences between two-way and one-way programs and a slightly larger difference favoring Spanish over other-language programs, though none of these differences are statistically significant at the 5% level. In mathematics, we find the reverse, with nonsignificant but often positive differential effects favoring non-Spanish languages.

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 than for native speakers of a language other than English. This would be a finding of concern except that the randomized

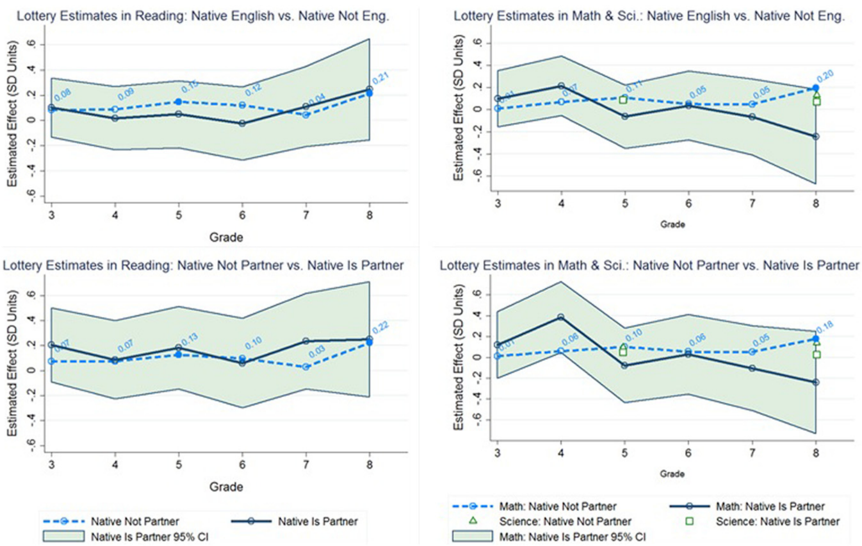

Figure 4. **Estimated intent-to-treat immersion effects for native English speakers and native speakers of other languages (top row) and for students whose native language does and does not match the partner language (bottom row).**

Note. 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. 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.

sample of non-native English speakers observable to Grade 8 is quite small, making the estimate very imprecise.

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 nonsignificantly) more than other immersion students. In contrast, their mathematics performance relative to other immersion students (non-native speakers of the partner language) shows a modest negative differential beyond about fourth grade, but the differences are not statistically significant.

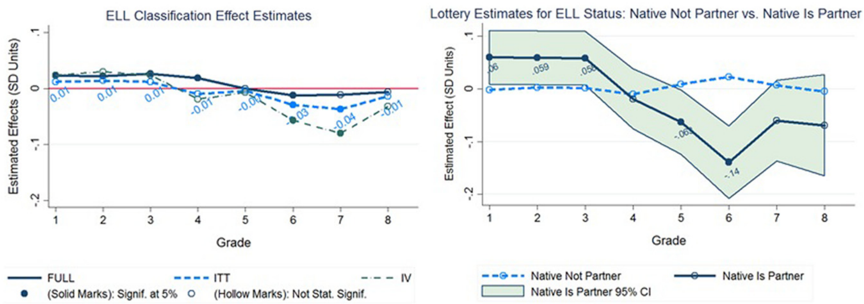

Figure 5. **Estimated effects of immersion on probability of English learner classification in each grade beyond kindergarten.**

Note. The full-sample model (left panel) includes 25,189 students and 126,139 observations. Intent-to-treat (and instrumental variables) models in both panels include 1,625 students (184 with native languages matching the partner language) and 8,805 student by time observations.

EL Reclassification Over Time

We turn now to the ITT effects of immersion programs on the probability that a student is classified as EL in each year, conditional on his or her EL status in kindergarten. We define EL status in kindergarten so that any student who is ever identified as EL is classified as such from kindergarten onward, until the student is reclassified as English proficient, ages out of the sample, or exits public schools in Oregon. Controlling for EL status at baseline, our *full-sample* (non-randomized) estimates in the left panel of Figure 5 suggest that among students who enter kindergarten as ELs, those who begin in immersion programs are roughly 2 percentage points more likely than their non-immersion peers to remain classified as ELs in Grades 1 through 4, after which their probabilities are mostly indistinguishable from those of their non-immersion EL counterparts. However, in the *ITT analysis*, which controls for baseline EL status, we find that 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 being EL are, respectively, 3 percentage points and 4 percentage points lower than those of their non-immersion peers.

Even with a small number of native speakers who were actually randomized to the partner languages to which they applied, we find modest evidence that a student's continued EL status depends not only on randomization to immersion but on whether the partner language matches the student's native language. Through Grade 3, native speakers of the partner language who are randomized to immersion remain more likely than their non-immersion counterparts 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 (see Figure 5, right panel). Notably, the randomized sample of EL students whose native language *does not* match the partner language is limited to only 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. Because some students age out of the sample each year, the estimates over time for this subgroup are quite imprecise.

Discussion and Conclusion

Our study contributes to the immersion literature in several key ways. First, it provides longitudinal, causal estimates of immersion programs 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 receive 25% to 100% of their mathematics and science instruction in the partner language through Grade 5. The fact that we find a *slightly* larger Spanish program advantage than two-way program advantage in reading suggests that impacts may vary more by partner language than by two-way versus one-way approaches, 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, the results might differ.

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; Fortune & Tedick, 2008). 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 (e.g., academic culture 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 appendix Table A6 in the online journal 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 2012–2013 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 the Elementary and Secondary Education Act (formerly No Child Left Behind, reauthorized in 2015 as the Every Student Succeeds Act). We find no evidence that these differences drive the estimated reading effects. Nevertheless, our study is designed to test the causal effect of *access to immersion* in Portland, which may yield access not only to instruction in two languages but also to teachers and peers who have been drawn to that instructional model.[7] If dual-language immersion program were scaled very widely—say, to all schools in a city—this would no longer be true. Moreover, rapid scaling without provisions to ensure quality might attenuate the treatment effect even if instruction in two languages 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 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, and no apparent detriment to mathematics and science skills—all while promoting 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 schools can prepare multilingual citizens 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 21st-century America.

Notes

This study was supported by Grant No. R305E120003 from the U.S. Department of Education's Institute of Education Sciences (IES). Particular thanks go to Debbie Armendariz, Joseph Suggs, Karin Brown, and Jennifer Miller in Portland Public Schools; Jonathan Wiens at Oregon Department of Education; and Allen Ruby at IES.

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

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

[3]To capture academic outcomes for individuals who enroll in Oregon public schools outside of Portland, we were able to match Portland Public Schools to Oregon Department of Education data. This augmented the analytic sample by 11% and improved grade-specific samples by 7% to 24%.

[4]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.

[5]In the data, reclassification is highly consistent with English Language Proficiency Assessment (ELPA) proficiency, suggesting strong adherence to the policy.

[6]We use a linear probability model for EL status, but logit models yielded similar estimates.

[7]Similar challenges in distinguishing mechanisms affect most random-assignment studies of school choice programs (e.g., Krueger & Zhu, 2004).

References

Alloway, T. P. (2007). *Automated working memory assessment (AWMA)*. Oxford, UK: Pearson.

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*. Eugene, OR: Author.

Barik, H. C., & Swain, M. (1978). Evaluation 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*, 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*, 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.

Cohn, D. V. (2015). *Future immigration will change the face of America by 2065*. Washington, DC: Pew Research Center.

Collier, V. P., & Thomas, W. P. (2004). *The astounding effectiveness of dual language education for all*. Fairfax, VA: George Mason University.

Devlin, K. (2015). *Learning a foreign language a "must" in Europe, not so in America*. Washington, DC: Pew Research Center.

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.

Fortune, T. W., & Tedick, D. J. (2008). One-way, two-way, and Indigenous immersion: A call for cross-fertilization. In T. W. Fortune & D. J. Tedick (Eds.), *Pathways to multilingualism: Evolving perspectives on immersion education* (pp. 3–21). Bristol, UK: Multilingual Matters.

Francis, D. J., Lesaux, N., & August, D. (2006). Language of instruction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 365–413). Mahwah, 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.

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

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.

New York City Department of Education. (2015). *Dual language and transitional bilingual education programs SY 2015–16*. Retrieved from http://schools.nyc.gov/NR/rdonlyres/49375968-8BB1-4617-A287-7A18C795B1FF/0/BilingualProgramsListSY201516.pdf

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.

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. *Trials*, *11*, 32.

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.

Tedick, D. J., Christian, D., & Fortune, T. W. (2011). The future of immersion education: An invitation to "dwell in possibility." In D. J. Tedick, D. Christian, & T. W. Fortune (Eds.), *Immersion education practices, policies, possibilities* (pp. 1–11). Bristol, UK: Multilingual Matters.

Thomas, W. P., & Collier, V. P. (2015). *English learners in North Carolina dual language programs: Year 3 of this study: School year 2009–10*. Fairfax, VA: George Mason University.

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

U.S. Census Bureau. (2014). *Young adults then and now: American Community Survey, 2009–2013*. Retrieved from http://www.census.gov/censusexplorer/censusexplorer-youngadults.html

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

Valdés, G. (1997). Dual-language immersion programs: A cautionary note concerning the education of language-minority students. *Harvard Educational Review*, *67*(3), 391–429.

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.

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.

Manuscript received June 16, 2015
Final revision received December 22, 2015
Accepted January 16, 2016