Dual-Language Immersion Education at Scale:

An Analysis of Program Costs, Mechanisms, and Moderators

Jennifer L. Steele *(Corresponding Author)*
American University
4400 Massachusetts Ave., NW, Washington, DC 20016-8030
(202) 885-3762 (phone), (202) 885-1187 (fax)
steele@american.edu

Robert O. Slater
American Councils for International Education
1828 L St, NW, Washington, DC 20036
(202) 833-7522
rslater@americancouncils.org

Jennifer Li
RAND Corporation
1776 Main St., Santa Monica, CA 90401
(310) 393-0411
jennifer@rand.org

Gema Zamarro
University of Arkansas
219B Graduate Education Building, Fayetteville, AR 72701
(479) 575-7024 (phone), (479) 575-3196 (fax)
gzamarro@uark.edu

Trey Miller
American Institutes for Research
4700 Mueller Blvd.
Austin, Texas 78723
(512) 476-6861
tmiller@air.org

Michael Bacon
Portland Public Schools
501 N. Dixon St., Portland, OR 97227
(503) 916-3151
mbacon@pps.net

Dual-Language Immersion Education at Scale:

An Analysis of Program Costs, Mechanisms, and Moderators

Abstract: Using input and outcome data from a randomized study of dual-language immersion programs in an urban district, we examine the mediating relationships of dosage, expenditures, and classroom characteristics to students' academic performance, and the moderating role of students' race/ethnicity. Differential costs of immersion were concentrated at the district level and were modest, at about 2% to 4% of per-pupil spending annually. We estimate that an additional $100 spent per immersion student in a given year was associated with an additional 8% of a standard deviation in language arts performance in English, which was just over a third of the causal point-in-time enrollment effect of 22% of a standard deviation. We find no generalizable evidence of differential effects by race/ethnicity.

Keywords: economics of education, educational policy, finance, language comprehension/development, bilingual/bicultural, urban education

Methodology Keywords: econometric analysis, experimental design

Acknowledgements: This study 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 from research collaboration with Deborah Armendariz, Director of Dual Language Immersion in the Portland Public Schools, and with Susan Burkhauser at RAND. It would not have been possible without data assistance from Joseph Suggs, Karin Brown, Jennifer Miller, Judy Brennan, and Janet Ruddell in the Portland Public Schools and Jonathan Wiens in the Oregon Department of Education. The authors are grateful for research feedback from Umut Özek, David Knight, and Teppei Yamamoto, from IES program officers Allen Ruby and Molly Faulkner-Bond, and from three anonymous reviewers. The opinions expressed here are those of the authors and do not represent the views of IES, the U.S. Department of Education, or the authors' respective institutions.

Data Statement: Because the study uses de-identified student-level administrative data and confidential principal and central office interview data, we regret that we cannot make our data publicly available. However, our interview protocols are included in an appendix file.

Dual-Language Immersion Education at Scale:
An Analysis of Program Costs, Mechanisms, and Moderators

Introduction

In evaluations of individuals' responses to behavioral interventions, randomized studies are considered the gold standard. Because they remove, in expectation, both unobserved and observed differences between treatment and control groups, randomized studies can eliminate concerns about selection bias that otherwise thwart studies of interventions—medical, behavioral, educational, etc.—to which people have some control over their own exposure. Of course, many social science interventions are multifaceted. Though random assignment to multifaceted interventions may allow researchers to estimate their causal effects, additional work is needed to illuminate the mechanisms by which such effects may be produced. Understanding the effects of randomization on access to intervention components or contexts can shed light for policymakers on how programs are being implemented in the real world, and on the resources involved. Ideally, evaluations of randomized policies or interventions should go beyond the question of whether they work into an exploration of why, and at what cost.

This article considers these questions in the context of a district-wide school choice option focused on delivering a bilingual educational experience. Specifically, *dual-language immersion (DLI) programs* are programs in which students receive general academic instruction in two languages from early grades onward. In our context, they include both *two-way programs*, in which at least a third of students in a classroom are native speakers of each of the two classroom languages (in the U.S., typically English and a "partner" language), and *one-way programs*, in which most students in the classroom are native speakers of English who are immersed in a non-English "partner" language. A key objective of both types of programs is to

produce students who are bilingual and biliterate, regardless of their first language (Fortune, 2012).

Research on DLI Education

Once considered niche, DLI programs are becoming more widespread. In just a few years, the number of programs operated under the DLI rubric nationwide has likely reached at least 2,000, as California, Delaware, Louisiana, Georgia, Minnesota, Ohio, Oregon, Texas and Utah have implemented programs.[1] In large cities like Chicago, Los Angeles and New York, DLI programs are steadily growing in number. The New York City Department of Education, for example, more than doubled the number of DLI 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). With more than 150,000 students classified as English Learners (ELs) in NYC (and 10% of those enrolled in immersion programs), and as many as 185 different languages spoken in students' homes, these programs are a cornerstone of the plan to deliver access to quality education for all NYC students.

The swift expansion of DLI programs may be driven by a number of complementary factors, including economic, cognitive, and academic rationales. The *economic rationale* focuses on preparing students for an increasingly global economy by equipping them with multilingual skills. Rigorous estimates of the earnings returns to bilingualism in North America range from 2% to 3% for non-English languages in the United States (Saiz & Zoido, 2005), to 4% to 6% for

[1] There are unfortunately no reliable statistics on the number and growth of DLI across the U.S. But there is incontrovertible evidence of a steady growth (e.g., North Carolina Department of Education, 2014). A 2011 *Los Angeles Times* article (Watanabe, 2011) estimated the number at more than 1,000 programs, with 224 programs (extrapolating beyond information from the Center for Applied Linguistics, 2011a, 2011b). Another article in *Education Week* (Maxwell, 2012) placed the number at over 2,000.

French in Anglophone Canada, to 7% to 8% for English in Francophone Quebec (Christofides & Swidinsky, 2010). In the U.S., the demand for workers with bilingual skills in health care, business, diplomacy, and national security arenas appears to be growing (Committee for Economic Development, 2006; Geisler, 2012; Zwerdling, 2012). Santibañez and Zárate (2014) found that Latino high school students who frequently speak Spanish are more likely, on average, to enroll in a four-year college than those who do not use Spanish frequently or are not proficient in Spanish.

The *cognitive rationale* for DLI is grounded in laboratory evidence that bilinguals outperform monolinguals on numerous verbal and non-verbal tasks, including working memory and executive function tasks, where the latter include attention control and task switching (Bialystok & Craik, 2010; Luk & Bialystok, 2014). Though these advantages tend to be small in magnitude, they could conceivably make modest contributions to students' ability to control their attention and focus, and to succeed on academic tasks, particularly those that require strong working memories (Gathercole, Alloway, Willis, & Adams, 2006). There is also evidence that familiarity with bilingual environments improves young children's social perceptive-taking skills (Fan, Liberman, Keysar, & Kinzler, 2016; Greenberg, Bellana, & Bialystok, 2013), which may help them work more effectively with teachers and peers. In addition, bilingualism appears to improve students' metalinguistic skills, such as lexical and semantic awareness (awareness of word meanings and sentence structures), and their ability to acquire additional languages (Cenoz, 2003; Keshavarz & Astaneh, 2004; Klein, 1995). This suggests that bilingual students may have advantages in understanding language structures and components in ways that make them better readers and writers in both languages.

The *academic rationale* follows logically from the cognitive rationale, based on the notion that instruction in two languages from early grades produces higher academic achievement in core academic content (e.g., language arts, mathematics, and science) tested in English. In the absence of evidence about the cognitive rationale, this notion might seem counterintuitive. Conventional wisdom suggest that there may be efficiency losses when the language of instruction and of academic testing are not well-aligned. On the other hand, the academic rationale has empirical backing. Studies of French immersion programs serving native English speakers in Canada have shown that immersion students perform as well as or better than their peers in English-tested content by about fifth grade (Barik & Swain, 1978; Caldas & Boudreaux, 1999; Marian, Shook, & Schroeder, 2013; Turnbull, Hart, & Lapkin, 2003), and some of these studies have used reasonably convincing longitudinal designs with matching on baseline attributes (Lambert, Tucker, & d'Anglejan, 1973). More recently, a study of one-way and two-way immersion programs in Utah used propensity score matching to find greater math gains from third to fourth grade for immersion students as compared to non-immersion students, but students were matched on post-treatment performance, meaning that cumulative immersion effects between first and third grade were not captured (Watzinger-Tharp, Swenson, & Mayne, 2016). In another recent study, Bibler (2017) used randomized lottery data from North Carolina to estimate causal effects of cumulative DLI dosage; he found benefits for native English speakers of 0.09 of a standard deviation per year in math, and 0.05 of a standard deviation per year in English language arts (ELA), with benefits of 0.06 of a standard deviation per year in math and ELA for ELs.

In the United States, most existing studies of academic outcomes among DLI students have focused on ELs enrolled in two-way programs.. Numerous studies of ELs exposed to two-

way immersion have shown them eventually outperforming peers exposed to monolingual English programs or to traditional bilingual education programs on language arts and math tests in English (Collier & Thomas, 2004; Lindholm-Leary & Block, 2010; Marian et al., 2013; Thomas & Collier, 2015). Although such studies have not historically used designs that adjust for students' selection into the programs, a few recent studies have taken steps to mitigate selection bias.

For instance, Umansky and Reardon (2014) employed a longitudinal analysis with extensive statistical controls, finding that Latino ELs placed in Spanish immersion classrooms were reclassified from EL to English-proficient status more slowly in elementary school but at higher rates by high school. And Valentino and Reardon (2015), using baseline controls and adjusting for parental preferences, found that ELs placed in DLI in kindergarten initially performed lower in ELA than those placed in short-term or long-term bilingual education or in monolingual English classrooms, but they showed much stronger growth in ELA performance between second and seventh grade than those placed in the other program types. The authors found lower growth over time in math for students in long-term bilingual education, but no statistically significant differences in math growth between the other three program types.

These studies implicitly raise the question of how immersion promotes positive outcomes. For instance, in the Valentino and Reardon (2015) study, if DLI yields steeper growth in ELA than other bilingual education programs, this may imply that the peer group or program attributes may be partly responsible, since the native and classroom languages are aligned in all except the monolingual English programs. Yet we have found very little research that empirically illuminates pathways through which immersion may influence achievement. Descriptive studies that have compared unadjusted achievement of elementary students

immersed in the partner language for 50% versus 90% of the school day (50:50 versus 90:10 programs) have found similar outcomes in terms of English and math tested in English (Collier & Thomas, 2004), with 90:10 programs showing similar or better performance in terms of partner-language proficiency (Christian, Montone, Lindham, & Carranza, 1997; Lindholm-Leary, 2001). In addition, Lindholm-Leary (2005) synthesized best-practices research to offer advice to immersion program directors on such features as curriculum alignment and school leader support, but the analysis reviewed best-practice literature and did not include an empirical comparison of programs with different attributes. Our study contributes to what is known about immersion mechanisms through descriptive analyses of classroom features that have potential to mediate program effects. Put another way, it comments on how well various immersion program attributes empirically account for the covariance between an immersion lottery win and subsequent academic performance.

If DLI is to be understood as a scalable reform model that promotes educational equity, then it is also important to understand whether benefits hold across racial/ethnic groups. In fact, dual-language education researchers have documented the need for research to examine differential demographic effects (Parkes, Ruth, Anberg-Espinoza, & de Jong, 2009). However, while numerous studies have focused on immersion's effects on ELs, and others have documented effects on native English speakers, we know of only two studies that have examined differential effects by race/ethnicity. Focusing on achievement trends for ELs between second and seventh grade in a large, urban district, Valentino and Reardon (2015) found that the long-term benefits of DLI relative to English-only education in ELA and math were statistically significant for Latino ELs but not for Chinese ELs, though modest long-term ELA benefits were observable for Chinese ELs in the sample. Their study used extensive statistical controls but

focused only on ELs and thus did not address racial/ethnic differences for native English speakers. Thomas and Collier (2014) examined test scores in six North Carolina districts for students who were and were not enrolled in two-way DLI programs, disaggregating outcomes for native English speakers by students' race/ethnicity. African American native English speakers in two-way immersion outperformed their non-immersion peers by 0.36 to 1.12 standard deviations in ELA and by 0.54 to 1.17 standard deviations in math; the corresponding relative performance among white native English speakers were -0.11 to 0.55 in ELA and 0.10 to 0.68 in math. The limitation of these estimates is that they are observational and unadjusted for baseline differences, making them vulnerable to selection bias, and they are cross-sectional, making it difficult to precisely estimate effects over time. Still, because they suggest that African American students whose families choose immersion outperform their same-race peers to an even greater extent than white students whose families do the same, they suggest a need for further study.

The current paper considers both the plausible mechanisms and differential effects of immersion using data from a lottery-based, randomized study of DLI education conducted in Portland, Oregon. Steele, Slater, Zamarro, Miller, Li, Burkhauser, & Bacon (2017) leveraged random-assignment lottery data for immersion program applicants in the Portland Public Schools to examine the effects of DLI on students' ELA, math, and science achievement through eighth grade. Using data from seven cohorts of students who were randomly assigned to DLI or business-as-usual before kindergarten in 2004-05 through 2010-11, the study found positive intent-to-treat effects in ELA of about 7 months of learning in fifth grade (0.13 SD) and 9 months of learning in eighth grade (0.22 SD). Using a quadratic grade specification instead of differential effects by grade, the estimated overall effect of winning an immersion lottery was a statistically significant 9% of a standard deviation in ELA ($p<.05$). Moreover, these effects were

statistically similar regardless of the classroom language (Spanish versus Chinese, Japanese, and Russian); students' native language (English versus the classroom partner language); and program type (one-way versus two-way). In addition, ELs randomly assigned to immersion had higher rates of English proficiency by sixth grade. The study found no statistically significant positive effects in math or science, but also no detrimental effects.

The present analysis builds on this work by documenting DLI inputs in Portland and the extent to which these inputs appear to mediate the effects of immersion on achievement. With regard to inputs, we first examine the costs and cost sources of Portland's DLI model relative to the default monolingual English curriculum in the district. We then examine the extent to which spending differentials and other inputs, including peer-group characteristics and teacher characteristics, covary with estimated lottery effects on student learning. Specifically, if classroom environments change as a function of winning or losing a lottery slot, then it is possible that the benefits of immersion are, in fact, benefits of being randomized to a different type of classroom environment. We use instrumental variables (IV) analysis to investigate the causal mediation effects of immersion enrollment and cumulative dosage on achievement and to explore the extent to which lottery-mediated changes in per-pupil expenditures and classroom characteristics correspond to changes in student achievement. Finally, to consider the relevance of the immersion model for students of different racial/ethnic backgrounds, we examine the role of students' racial and ethnic backgrounds in moderating the causal effects of immersion on student achievement.

Research on Immersion Costs

By some arguments, DLI education should be cost-neutral. Because core academic instruction occurs in two languages, second-language instruction does not necessarily require additional classes or staff. In this model, a second language is not an elective or an enrichment class, like music, art, or physical education; it is simply a medium of instruction. The notion is that if students receive core instruction in two languages from young ages, they will acquire both somewhat naturally, without sacrificing instructional time in core content (e.g., mathematics, science, social studies, and language arts) and without requiring additional staffing or classroom space for elective classes. Though the case for cost neutrality can be made in theory, in practice, additional resources may be needed to support the model, including resources spent on recruiting and hiring bilingual teachers, translating or purchasing curricula and assessments, translating parental communications (especially in two-way programs, where parents may be native speakers of either instructional language), and helping teachers learn to teach in a DLI program—whether it be a two-teacher model in which students switch between teachers of the two languages (and teachers must coordinate their curriculum), or a one-teacher model, in which a bilingual educator teaches in both classroom languages. There may also be logistical costs involved in allowing families to opt into or out of dual-language programs and in managing demand for slots, and there may be costs associated with transportation to these programs. Another set of possible costs would be associated with the launch of new programs—these include analyzing families' demand for DLI, deciding where to locate these programs, recruiting the initial set of teachers, and purchasing curriculum in the partner language (Lara-Alecio, Galloway, Mason, Irby, & Brown, 2004). Because two-way immersion programs often serve

ELs alongside native English speakers, it is also important to consider the extent to which DLI supplants rather than complements the instructional support ELs would otherwise receive.

To understand how program costs mediate student achievement effects, we must first estimate those costs and their sources. We know of only two studies that have systematically attempted to measure the costs of DLI programs over and above other per-pupil expenditures. Parrish (1994) studied the costs of two-way immersion program classrooms relative to non-immersion (English-only) classrooms as part of a 15-school cost study of programs serving ELs in 11 California districts. Though the schools chosen were described as "exemplars" of their models (p. 261), site-selection and data collection methods were not directly stated. Focusing only on classroom-level resources, he initially found lower classroom costs of $74 per student (about 3.9%) for two-way immersion programs relative to English-only programs in the same school, which cost $1873 per pupil on average. Since average class sizes were the same, the cost difference was attributable to slightly lower instructional aide staffing in immersion relative to non-immersion classrooms, at 9.0 versus 12.9 hours per week. He also found higher parent volunteerism in two-way immersion relative to non-immersion classes (7.6 versus 2.1 hours per week), but appears not to have assigned a dollar value to this input. In addition to these costs, which were calculated based on average classroom sizes in each program type, Parrish documented program resource and administration costs based on number of EL students in each site. If we adjust his per-EL pupil figures for the ratio of students he reported to be ELs in two-way immersion programs (0.34), the resource and administrative costs represent an additional $298 per immersion pupil, including 57% administrative costs, 22% direct instructional costs, and 21% teacher training and in-service costs. Adjusting for the classroom costs, which were

lower in immersion sites, this would bring the total additional cost of immersion program to $224 per pupil, or an additional 12% at the time of the study.

In a later study, Lara-Alecio et al. (2004) surveyed directors of the 166 DLI programs operating in Texas, receiving a 50% response rate. Survey items focused on schools' reports about dollars spent on each of 12 input categories for ELs, over and above what the district spent on transitional bilingual education programs, in which ELs typically persisted for 2 to 4 years before transitioning to English-only classrooms. The authors estimated additional immersion cost, on average, of $525 per pupil per year. Though they did not contextualize this in terms of baseline costs, the U.S. Census Bureau (2017) reports that average per-pupil spending in Texas was $7267 in 2004-05, the nearest year with data available. This implies that the differential for two-way immersion relative to transitional bilingual education was about 7% of annual per pupil spending in the state. The authors did find economies of scale, such that in the largest programs (those with more than 240 students), the annual differential costs were only $290 per pupil per year. In those programs, the largest cost categories included 23% toward curriculum development; 21% toward managerial costs at the central office level; and 19% toward staff development. The districts also reported spending about 9% on additional staffing including teachers, paraprofessionals, and tutors.

Because the two studies focus on different comparisons programs, they are not themselves perfectly comparable, but they suggest that supplemental immersion costs range from roughly 7% to 12%, with key cost categories falling into administration, additional instructional staffing, professional development, and curriculum development. Both studies acknowledged that they were unable to tie program costs to effectiveness. Though focused on a single urban district, our study aims to build on these prior studies by using variation in immersion program

expenditures over time to assess the mediating effect of such expenditures on students' language arts achievement in English.

Organization of the Paper

Having described the extant literature on DLI effects and differential costs, we set forth our research questions in the next section. We then describe our data sources, followed by a section describing our empirical strategy. Our results focus first on our conclusions about relative immersion costs. We then examine the roles of immersion enrollment, dosage, relative expenditures and classroom characteristics in mediating these programs' effects on language arts and mathematics achievement tested in English. Next, we report on the extent to which race/ethnicity appears to moderate the effects of immersion programs. We conclude with a discussion of our findings and their implications for immersion policy.

Research Questions

Our paper contributes to the literature on the relative costs of DLI programs; the role of enrollment and cumulative dosage, per-pupil expenditure and classroom characteristics in mediating the effects of immersion education; and the role of students' race/ethnicity in moderating immersion effects as implemented in Portland.

Our overarching research questions are as follows:

1. What are the additional, annual costs of DLI education relative to monolingual education in the district, and what are the sources of these costs?

2. To what extent do program enrollment and dosage, program expenditures, and classroom characteristics (teachers, peers, and class size) appear to mediate the relationship between immersion access and student achievement?
3. To what extent do DLI effects on ELA and math achievement differ by the race/ethnicity of randomized program applicants?

Research Setting and Intervention

When our research began in 2012-13, the Portland Public Schools offered immersion programs in 11 elementary schools, 4 middle schools, and 5 high schools, with instruction in Spanish, Mandarin, Japanese, and Russian. All but one of these schools offered both immersion and non-immersion programs within the same school. In 2012-13, about 8% of Portland's students, or 3,860 individuals, were enrolled in immersion programs.

Two-way programs. During the school years in our analysis, the Russian program and all but one of the Spanish programs followed a *two-way* instructional 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 two-way programs in Portland followed a 90/10 instructional model that was specified by the central office, meaning that in kindergarten, 90% of the school day was conducted in the partner language, and 10% in English. The partner-language proportion then declined by 10 percentage points per grade. In grades K-3, students received 75% to 100% of mathematics instruction, 56% to 100% of language arts instruction, and 100% of science and social studies instruction in the partner language. In grades 4 and 5, they received 25% of mathematics, 58% of language arts, and 100% of science and social studies instruction in the partner language. Middle school students took 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 were conducted in English. High school immersion students typically took only one class per day—an advanced language class—in the partner language.

One-way programs. 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. In Portland's one-way programs, instruction of core content (mathematics, language arts, science, and social studies) followed a 50/50 instructional model in each elementary grade. Each day, as specified by the central office, half of the instruction in each core subject occurred in the partner language, and half occurred in English. In middle and high school, however, one-way and two-way programs were designed to operated 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.

Lottery-based random assignment. Students received 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 could apply for lottery-based admission to a DLI program of their choice or to one of several other choice programs offered by the district. Each immersion elementary school established the number of lottery slots available in a given year, with some offering particular slots for students living within and outside of the catchment neighborhoods, and others offering all slots district-wide. Schools with two-way programs designated about half of lottery slots for native speakers of the partner language to promote a balance of native and non-native partner language speakers in the program. Schools with one-

way programs did not have partner-language native-speaker set-aside slots during the years of our analysis.

Though families could list three school preferences in rank order, in practice, their first choice was determinative, as all slots filled in the first lottery round. Lottery slots in a given school and preference category were filled first by students who had siblings at the school, then by other applicants who resided with the school district, and then by applicants from outside the district, thus creating an even larger number of random-assignment strata. Only categories in which the number of available slots was positive and the number of applicants exceeded the number of slots yielded true randomization, and we consider only applicants in binding lottery strata as part of our randomized sample.[2] Lottery applicants who did not win immersion slots were assigned to the regular instructional programs in their default neighborhood schools.[3]

Data Sources

Input and Expenditure Data

To address our first question, about additional, annual costs of DLI relative to monolingual education in the district, we conducted 90-minute interviews with 14 of the 19 immersion school principals at the time. This yielded a principal interview response rate of 74%. The response sample included 7 elementary schools (out of 10); 3 middle schools (out of 4) and 4 high schools (out of 5). In terms of program type, 6 of a possible 8 were principals of one-way programs, and 8 of a possible 11 were principals of two-way programs.

[2] In practice, nearly all truly randomized strata were within district, and none were sibling strata.

[3] A CONSORT table describing binding and non-binding lottery applications, results, and compliance can be found in Figure 1 of Steele et al. (2017). We are able to follow randomized students who leave the district as long as they stay in public schools in Oregon. In Steele (2017), our resulting attrition levels appear to fall within the liberal attrition threshold set by the What Works Clearinghouse (2014) for meeting standards without reservations in all grades, and within the conservative attrition threshold in for a few later-grade comparisons.

Following the ingredient approach to cost analysis (Levin & McEwan, 2001), in which differential inputs are documented and costs are subsequently assigned, we asked principals about the time they devoted to various facets of their schools' immersion and non-immersion programs. We also asked about within-school differences, if any, in terms of teacher engagement, parent engagement, field trips, technology, grants, and other resources. A copy of the principal interview protocol is included in the appendix.

In the summer of 2014, we also conducted 120-minute interviews with the director and assistant director of the Dual Language Department for the district with regard to their resources, budgets, and time use during the 2013-14 school year. These interviews concerned central office staffing for immersion programs, including salary categories and responsibilities, funding sources for the DLI program, development and purchasing of curriculum and assessments in the partner languages, professional development offerings and costs for immersion administrators and teachers, parent outreach efforts and cost, costs associated with planning new programs, and miscellaneous costs. A copy of the central office interview protocol is also included in the appendix.

The interviews revealed that the district had increased its direct investments in immersion over time, so we needed to understand not only the available inputs in 2013-14, but also what those inputs had been for the students enrolled in immersion programs during each year of our student achievement study, where the oldest students had begun kindergarten in 2004-05. To do so, we also collected data from annual district budget reports, which were publicly available for the 2006-07 through the 2013-14 academic years. In particular, the published budget reports illuminated the size and source of grant dollars the immersion program had received in each year since 2006-07. Since budget reports were not available prior to 2006-07, we imputed data for

2004-05 and 2005-06 by replicating amounts in the 2006-07 year. To scale the cost input data in per-pupil terms, we used historical immersion program enrollment counts provided by the district for 2005-06 through 2013-14. Data for 2004-05 and for 2009-10 and 2010-11 were linearly interpolated based on actual figures from the other years. All dollar values and salaries from prior years were rescaled as 2013-14 constant dollars.

Since we were also interested in the role of classroom characteristics as mediators of students' immersion lottery status, we used classroom-characteristics data provided by the district for the 2012-13 school year. This dataset provides point-in-time descriptive information about teachers and peers in the homeroom, mathematics, and language arts classes of all lottery applicants in the analytic sample, including the number of students in the class, the proportion in each racial/ethnic group, the proportion receiving special education services, the proportion designated as gifted, the proportion designated as ELs, and the proportion eligible for subsidized meals. It also indicates highest education level and the years of experience of the classroom teacher, as well as an indicator of whether he/she was defined as "highly qualified" (i.e., licensed in his/her subject) under No Child Left Behind.

Student Achievement Data

In the student achievement analysis, we examine seven cohorts of students who entered pre-K or kindergarten in the academic years 2004–05 through 2010–11, and we track them through the 2013-14 academic year, so that all cohorts are observed through grade 3, and the two oldest cohorts are observed through grade 8. The data include 3,457 lottery applicants, 1,946 of whom were truly randomized in an oversubscribed lottery stratum (based on school, year, and preference category), and 1,625 of whom could be included in the student outcomes analyses

because their outcomes were observed. Of these, 752 were assigned immersion slots, and 873 were not. As shown in the left-hand panel of Table 1, which presents descriptive statistics for the randomized sample, the groups were balanced within randomization strata on all observable characteristics. About 18% of the randomized sample spoke a language other than English at home, and about 13% of the randomized sample were classified as ELs.[4]

[Insert Table 1 about here]

We measure student achievement using students' scores on the state accountability test. During the period of the study, this was the Oregon Assessment of Knowledge and Skills (OAKS), and we focus here on language arts and mathematics scores, which were collected in the spring of grades 3 through 8.

Empirical Strategy

Research question 1 concerns the additional per-pupil expenditures on immersion relative to expenditures on monolingual education in the district. Whereas prior cost analyses have focused on differential costs exclusively for ELs, we focus on differential costs of the model, which functions homogenously for non-ELs and ELs. We adopted this approach because at the inception of the study, the English as a Second Language Department, which focuses on EL support, and the Finance Department both clarified in interviews that supplemental resources provided to ELs in Portland were constant under both instructional models. Specifically, they noted that ELs received Sheltered English Instruction (SEI) as a default in monolingual English classrooms, meaning that instruction was in English, but that visuals and manipulatives were

[4] As a basis of comparison, descriptive statistics for all kindergarten entrants in the district in those years, not restricted to the randomized sample, can be found in Steele et al. (2017).

used to support English acquisition (Short, Echevarria, & Richards-Tutor, 2011).[5] In addition, ELs in the district qualify for periodic, supplemental instructional support using a classroom pull-out model. The English as a Second Language Department reported that these resources were provided equally in the district, whether the students were placed in English-only or DLI classrooms. Our team's subsequent interviews with immersion teachers for a separate component of the study corroborated the accuracy of this claim.

We calculated immersion cost data at the *central office level* using central office interview data, triangulating recollected information from central office staff with annual data from published district budget reports, so that costs and revenue sources could be tabulated by year. For instance, for some cost categories, staff recollection of specific expenditures were not available or were explained in terms of the grants that paid for them. In these cases, historical budget data about revenue sources available to the Dual Language Department by year helped to provide a more complete cost history.

Our analysis excludes certain categories of central office expenditures that supported the immersion programs alongside other instructional and school choice programs in the district. Through structured interviews with personnel from several departments during the first two years of the study, we determined that the workload of these departments would have been relatively constant with or without DLI programs. These departments included Human Resources, which did describe supporting the dual-language program's teacher recruitment efforts, but said they did not spend proportionally more time on immersion than they would have on other teacher openings, as well as the Enrollment and Transfer Department, whose job, they explained, was to facilitate lotteries for a variety of school choice programs across the district, regardless of their

[5] In fact, any teacher in the district who has students who are not fluent in the language of instruction is expected to use sheltered techniques, including DLI teachers.

focus, and the System Planning and Performance Department, which helped to site new immersion programs but only, they explained, as part of larger responsibilities for siting new schools and balancing enrollments. Adapting the work of these departments to the particular needs of DLI reportedly fell mainly to the Dual Language Department, whose annual costs therefore became our focus. We also exclude student transportation costs since these were controlled by individual immersion programs and were not tracked historically. We were also told that most schools did not provide supplemental immersion transportation during the study period, though central-office funding for immersion transportation has reportedly increased since the study ended.

We estimated *school-level costs* using interview data with principals of immersion schools. These, as noted above, include hidden or unbudgeted costs such as differential time use and differential levels of parents' volunteer labor. To analyze the principal interview data, we tabulated responses to each item for each school's immersion and non-immersion programs, calculating immersion-to-non-immersion input ratios for each school. For principals' time use, we asked about the hours they spent per week on particular tasks for their immersion and non-immersion programs, respectively, or (if they couldn't provide hours) about their total work time per week and the percent of time they spent on each task for their immersion and non-immersion programs. If they reported that their time use on a given task supported both programs universally, simultaneously, or proportionally, we recorded the total time spent on the task and allocated the hours proportionally by student enrollment. We then calculated total days of effort for each task-by-school-by-program type (immersion vs. non-immersion), and we calculated an immersion task proportion for each task and each school. To calculate an immersion effort ratio, we divided these immersion task proportions by the proportion of students in the school enrolled

in immersion during the interview year, 2013-14. An effort ratio of 1 indicates that a principal's effort on a given task is perfectly proportional to the share of immersion students in his/her school. A ratio greater than 1 suggests that the time on a given task is disproportionately focused on immersion; a ratio of less than 1 suggests that it is disproportionately focused on non-immersion programs. Tasks on which principals reported that their efforts were proportional or served the whole school equally were coded as proportional. Cross-school immersion effort ratios were calculated first within schools and then averaged across schools.

To address *research question 2*, which focuses on the role of immersion dosage, per-pupil expenditures, and classroom characteristics in mediating the effect of random assignment to immersion, we use an instrumental variables (IV) approach. Our approach is similar to the one used by Bifulco, Unterman, and Bloom (2014), who leveraged random-assignment lottery data to estimate the relative cost-effectiveness of attending a small high school of choice relative to a traditional public school. Using lottery-based access to small high schools as an instrument for per-graduate expenditures, they found that winning access to a small high school yielded lower per-graduate costs despite slightly higher levels of per-pupil spending at these schools.

In this analysis, we are able to estimate the causal effect of cumulative immersion dosage on the academic achievement of lottery status *compliers*. These are the students whose cumulative exposure to immersion was wholly determined by their lottery winning or losing status within the stratum to which they applied. This estimation yields a local average treatment effect (LATE) of dosage, which is interpreted as the average treatment effect on students whose dosage complies with their lottery status (Angrist & Pischke, 2008). Because dosage effects are predicted strictly based on students' lottery status and other pre-intervention covariates, they are

estimated net of endogenous student or family choices that may otherwise influence cumulative immersion exposure.

To draw causal inferences about dosage effects, we must satisfy three assumptions of instrumental variable analysis. First, the exogenous instrument must be strongly associated with the endogenous mediator—in this case, immersion dosage—it is predicting. The strength of this association can be established analytically; a common heuristic expectation is a first-stage F-statistic of 10 or greater (Stock, Wright, & Yogo, 2002). Second, the effect of the instrument must be monotonic, meaning unidirectional. In this case, policy restrictions on circumventing lottery results provide reasonable assurance that lottery winning can raise but not lower cumulative exposure to immersion. Third, instruments must satisfy the exclusion restriction, meaning that, conditional on pre-treatment covariates, the instrument and the outcome of interest must be related only via the mediator of interest, which in this case is cumulative immersion dosage (Angrist & Pischke, 2008; Imai, Keele, Tingley, & Yamamoto, 2011). This assumption of no third path is plausible because cumulative immersion exposure arguably captures all features of immersion programs to which lottery status regulates access. Though dosage varies over time, it can still be instrumented insofar as it meets the aforementioned assumptions. Angrist and Pischke (2008) clarify that when treatment intensity varies as it does in the case of dosage, the average causal response (i.e., instrumented effect) is a weighted average of the effects for compliers at each observed point in time (p. 98).

To estimate the LATEs for each additional year of immersion enrollment, we use a 2-stage least squares approach, as shown in equations 1 and 2:

$$DLI_{it}^{med} = a_1 + \tau_1 z_i + \theta_1 g_{it} + \beta_1 g_{it}^2 + \boldsymbol{\delta_1 X_i} + \boldsymbol{\gamma_1 L_i} + u_{1i} + \varepsilon_{1it} \quad (1)$$

$$y_{it} = a_2 + \tau_2 \widehat{DLI}_{it}^{med} + \theta_1 g_{it} + \beta_2 g_{it}^2 + \boldsymbol{\delta_2 X_{it}} + \boldsymbol{\gamma_2 L_i} + u_{2i} + \varepsilon_{2it} \quad (2)$$

In the first stage (equation 1), the randomly assigned lottery admission status for student *i* is given by z_i. It serves as an exogenous instrument for DLI_{it}^{med}, which here represents the cumulative years that student *i* as been enrolled in immersion as of time *t*. The model controls for a quadratic specification of time, captured by g_{it} and g_{it}^2, which respectively represent grade and its square. It also includes $\mathbf{L_i}$, a vector of time-invariant dichotomous cohort × school × randomization subgroup lottery indicators, and $\boldsymbol{X_{it}}$, a vector of time-varying student characteristics. The constant is a_1. Student-level and student-by-time random error terms are given by u_{1i} and ε_{1it}, respectively.[6]

In the second stage (equation 2), which is estimated simultaneously with equation 1 using *xtivreg* in Stata 14.1, the estimated value of dosage given by $\widehat{DLI}_{it}^{med}$ becomes the treatment variable predicting achievement in ELA or math for student *i* in time *t*. Student achievement is represented by y_{it}. The other terms are defined as in equation 1. The LATE, or causal effect of DLI_{it}^{med}, on achievement among compliers is given by parameter τ_2 (Angrist & Pischke, 2008). Because z_i is randomly assigned within strata denoted by $\mathbf{L_i}$, it can be assumed to be unrelated to y_{it} except through its effect on DLI program participation.

In fact, the LATE estimate for dosage aggregates the effects of a host of potential mediators that together constitute exposure to DLI in Portland. Program attributes that may individually mediate the relationship between lottery status and student achievement include core instructional features like the classroom language of instruction (which may affect metalinguistic

[6] The model differs slightly from the one used in Steele et al. (2017), which employed grade-level fixed effects and their interactions with the lottery winning indicator, z_i. Here, we use a quadratic time specification because the instrumental variables strategy has lower power than the intent-to-treat models emphasized in the prior paper, and because our parameters of interest—namely, the effects of time-varying attributes such as dosage and cumulative per-pupil spending—already capture differential outcomes for lottery winners as a function of time in immersion programs.

awareness, cognitive processing, etc.), as well as other classroom attributes such as class size, instructor preparation and motivation, curriculum quality, and the baseline skill and motivation of classroom peers. Some of these, such as instructor skill and curriculum quality, may be affected by per-pupil spending differentials, since the DLI department in Portland ensures the availability of aligned partner-language curricula.

We would like to be able to decompose the LATE of immersion dosage into what Imai et al. (2011) term the *average causal mediation effects* (ACMEs) of observed attributes such as per-pupil spending, class size, teacher experience, and peer demographics, and into the *average direct effect* (ADE), which represents unobserved component mediators of interest, such as the cognitive effects of bilingual instruction. Though we do estimate *average mediation effects* of these observed attributes, we cannot go as far as saying that the effects of these attributes are causally estimated, because these attributes may be correlated with one another in ways that are not fully observed (Imai et al., 2011). For example, it is possible that classrooms serving more affluent students also have teachers who are more experienced or motivated, particularly if such schools provide easier working conditions (see, for instance, Lankford, Loeb, & Wyckoff, 2002; Steele, Pepper, Springer, & Lockwood, 2015). In addition, it is possible that variation in per-pupil spending over time is associated with other programmatic differences (school leadership or teacher staffing changes) that are not directly driven by changes in immersion program spending.

Subject to that limitation, we estimate the relationships of observed, plausible component mediators to student achievement. Adapting equations 1 and 2, we allow DLI_{it}^{med} to represent potential mediators such as per-pupil expenditure differentials at the central office level by observed dosage over time, as well as observed classroom and teacher characteristics that were available for a single year, 2013-14. We recognize that if these attributes are correlated with

unobserved program characteristics that also mediate immersion effects, then our estimates may either understate or overstate their causal effects on achievement (Schochet, Puma, & Deke, 2014). But the analysis remains potentially illuminating since there is little extant research even documenting the observed relationships among immersion programs, immersion attributes, and student learning.

Question 3 concerns the role of students' race/ethnicity in moderating the causal effect of DLI enrollment on student outcomes. To address it, we adapt equations 1 and 2 with DLI_{it}^{med} now defined as enrollment in an immersion program at time *t*. This parameter of interest, τ_2, in this case represents the average observation-weighted causal effect across grades. We interact the lottery-winning indicator, z_i, with each of the race/ethnicity indicator variables, using these interactions to instrument the interaction effects between DLI_{it}^{med} and the race/ethnicity indicators. This allows us to estimate the differential causal effects of immersion enrollment on students of different racial/ethnic backgrounds, adjusting for all aforementioned covariates including lottery stratum. As we test for differential effects by race/ethnicity, we cannot fully disentangle differential subgroup effects from the programs to which each subgroup applied in the immersion lotteries. For instance, students with Asian backgrounds disproportionately applied to the Chinese (70%) or Japanese (14%) immersion programs, whereas Spanish programs were more frequently chosen by lottery applicants identified as Black (93%), Hispanic (94%), or White (73%). So the moderating effects must be interpreted with these programmatic differences in mind, meaning that they may be attributable to the particular immersion--and non-immersion-- programs that disproportionately serve each subgroup.

Results

Additional Per-Pupil Expenditures Relative to Monolingual Education

Based on principal and central office interviews, we concluded that differential expenditures occurred primarily at the *central office level*. These were concentrated mainly on staffing of the Dual Language Department and in external grants to the district that were procured by central office dual-language program staff. Historically, staffing levels have been small, though they have grown in recent years. Staffing increases have outpaced increases in immersion enrollments due to targeted increases in support by the district, so per-pupil expenditures have risen over time. Drawing on annual budget data as described in the data section above, and triangulating with contextual information from district interviews, Figure 1 shows per-pupil expenditures for all immersion-enrolled students in 2004-05 through 2013-14, disaggregated into two sources: the general fund paid by the district, and external grants. The district general fund expenditures remained relatively flat during the observed period, with an average expenditure of $78 per immersion pupil, ranging from $46 to $124 across observed years. External grants for immersion during this period consisted mainly of the federal Language Flagship grants of $400-500K per year, and the federal Foreign Language Assistance Program (FLAP) grants, which were similar in size to the Flagship grants but were discontinued by the federal government in 2012. The total per-pupil expenditure line in Figure 1 represents the sum of the two expenditure sources. It ranged between $163 and $481 in the observed years, with a mean of $300 per pupil. To put these figures in context, the average level of per-pupil spending across the district during this period was approximately $10,800, ranging from $9,306 in 2004-05 to $11,318 in 2013-14. Thus, immersion spending per-pupil ranged from about 1.8% to 4.2% of per-pupil spending in the district in any given year.

[Insert Figure 1 about here]

The Language Flagship grants were awarded to support the district's Mandarin immersion program, which, during the study period, was housed in only one elementary, one middle, and one high school in the district. However, the grant allowed the district to use the funds to provide capacity for dual language programs at the central office level, which is where the spending was concentrated, with only about a 25% pass-through to the schools. Because the Language Flagship funds have gone primarily toward central office support for dual language programs, we treat them as part of the per-pupil expenditures on DLI in general and not specifically for students in Mandarin programs.

In interviews, the district reported that it had used the FLAP grants largely to support the development and administration of standardized tests for assessing students' proficiency in the partner languages. Capacity for this testing has gradually grown over time, such that researchers were able to track partner-language trajectories in Burkhauser, Steele, Li, Slater, Bacon, & Miller(2016). FLAP funds also reportedly went toward curriculum translations in the less commonly taught languages, especially Russian.

Central office staffing time and effort during the study period were reportedly spent on providing professional development for teachers to facilitate consistent instructional practices, and on supporting the Human Resources department in targeted recruitment and hiring of licensed teachers who were qualified to teach DLI. They reported that immersion teachers and principals received the same amount of professional development as other teachers and principals in the district, but that required professional development was often customized for immersion teachers and principals by the Dual Language Department. In cases where teachers volunteered for supplemental immersion professional development, they reportedly received

contractual hourly stipends paid for by the Dual Language Department; such costs were described as small and are included in Figure 1. Central office staff reported that a small fraction of their time also went toward working with the Curriculum Department to ensure that compatible curriculum were available in the partner languages in the needed grades and subject areas, but the costs of the curricula themselves, given that they were non-duplicative, were borne by the Curriculum Department.

In terms of differential resources at the *school level*, we found immersion and non-immersion class sizes to be similar within and between immersion schools. Meanwhile, the district reported that they did not purchase duplicate curricula for DLI students. Instead, the curriculum was generally provided in a single language for any given content area and grade. In one-way programs, when a particular subject was taught in both languages in the same grade, the curricula were described as print-on-demand, thus avoiding duplicated purchases.

We also asked principals about the percent of time they devoted to particular tasks for the immersion and non-immersion programs in their respective schools. We then adjusted their reports according to the sizes of their immersion and non-immersion populations. This yielded task-specific adjusted proportions of effort for each principal's immersion and non-immersion program. Tasks that principals said applied equally to both groups were coded as such. We divided the immersion proportion by the non-immersion proportion for each task and school to yield a set of *immersion effort ratios*. The cross-school averages of these ratios by task are presented in Figure 2.

[Insert Figure 2 about here]

Figure 2 illustrates that on average across tasks, principals' self-reported efforts were highly proportional. The top bar of the graph shows the mean cross-task ratio, which is one. As

one might expect, the least proportionate category was "Other tasks specific to running a DLI school." The tasks principals mentioned in this category included organizing special celebrations and events, as well as communicating with local advocacy groups.

The task with the second-highest reported ratio was outreach to prospective parents, whose ratio of 1.9 suggested that principals spent almost twice as much time per pupil on this task for their immersion program as for their non-immersion program. Also disproportionately high were outreach to current parents, with a ratio of 1.47, and teacher observation, with a ratio of 1.25. Principals described the latter as modestly more challenging in immersion programs since they did not always speak the partner language and thus had to observe with a focus on general pedagogy and student engagement or find a co-observer fluent in the partner language. Tasks with ratios less than one were tasks in which principals reported spending a relatively large amount of time on their non-immersion programs. These included teacher hiring (0.96) and credential waivers for teachers (0.37). These ratios were surprising given anecdotes we had heard across the district about the difficulty of hiring licensed bilingual teachers. But principals noted that they looked to the DLI department in the district central office for recruitment, hiring, and development of immersion teachers.

In addition, we asked principals about other school-based resources, such as teachers' work time, parents' volunteer time, number and duration of field trips, and grant dollars coming directly to the school (excluding grants that passed through the district, like Flagship and FLAP). Principals' responses with regard to school-level resources for their immersion and non-immersion programs are summarized in Table 2. Principals estimated that their immersion teachers worked more hours beyond the contract day than their non-immersion teachers, but because they admitted to having incomplete data on teachers' work hours, we view this mean

difference as a measure of principals' perceptions rather than as a direct measure of teachers' labor. The number of field trips they reported per year were proportional or, in the case of travel field trips, favored non-immersion programs on average. On the other hand, 54% of principals did report that immersion classrooms had additional technology, such as smartboards or laptops. They reported that these had been purchased with one-time grants or fundraising on the part of teachers, and their ages were not known.

[Insert Table 2 about here]

With regard to direct grant dollars per student, we calculated this in two ways. First, we calculated the total grant dollars principals reported for immersion programs versus the total they reported per student school-wide for the 2013-14 school year, divided by the number of immersion students in the first case, and by the total school enrollment in the second case. In other words, the non-DLI statistics are actually whole-school statistics with respect to the grant dollar line items. The *inclusive* measure of grant dollars per student includes pass-through from district grants like Language Flagship, and it includes parent contributions earmarked for fieldtrips. Excluding these two categories (since we include *all* grants to the district in the central office cost analysis, and since we include field trips as separate line items in Table 2), the net grant dollar per student for immersion students is relatively low, at about $12 per student per year, as compared to an average of $114 per student per year in schoolwide grant funds. That $12 can still be understood as additional funds for immersion students, since grant dollars to the school were rarely earmarked specifically for non-immersion students, but the ratio of immersion-specific dollars to schoolwide dollars was very small—only 2%—when calculated across school-by-year observations.

In summary, our analysis suggests that expenditure differentials associated with DLI in Portland between 2004-05 and 2013-14 were relatively modest, at roughly 1.7% to 4.2% of per-pupil spending in the district, and were supported mainly by external grant dollars, with quantities that varied notably from year to year. Central office interview data with various departments suggested that differential effort to support immersion was concentrated within the Dual Language Department. Cost interviews with immersion school principals about the distribution of both effort and resources revealed proportional effort and suggested that any within-school expenditure differentials were idiosyncratic and minute. In a district that carefully adheres to proportional allocation of resources, this finding of proportional school-level expenditures is consistent with expectations. However, it suggests that the implementation of a program such as Portland's depends on a centralized and active DLI department to reinforce consistent and effective implementation across schools.

First-Stage Effects of Lottery Status on Enrollment, Dosage, and Expenditures

We turn now to question 2, in which we examine potential mediators between immersion lottery status and student achievement in ELA and math, focusing in Table 3 on those for which we have annual data. Table 3 is divided into a top and bottom panel—A and B, respectively—with first-stage estimates in Panel A and second-stage estimates in Panel B. In both panels, estimates for ELA are shown on the left, and estimates for math on the right. The first two mediators of interest are point-in-time enrollment and dosage; the second two are point-in-time per-pupil spending differentials and cumulative per-pupil spending differentials at the central office level, based on immersion dosage as influenced by lottery-winning status. Per-pupil expenditures are defined at the district level, since that is where the differential is concentrated.

[Insert Table 3 about here]

Conditional on lottery strata fixed effects and demographic controls, we find that winning an immersion lottery increased the probability of enrolling in immersion in any given year by 41 percentage points, as shown in Panel A. Though the effect may seem modest, initial compliance with lottery-assigned status was 77% among lottery winners and 73% among lottery losers (see Steele et al., 2017, Figure 1), and the F-statistic of 51 in the ELA analysis and 50 in math greatly exceed the heuristic threshold of 10 for instrument strength (Stock & Yogo, 2005).[7] Imperfect compliance among lottery winners may be due to ambivalence among some applicants, as lottery applications require only completion of a short form online or in person. Imperfect compliance among lottery losers is primarily due to entry from wait lists due to winner noncompliance. We must interpret entry from wait lists as noncompliance because it depends on endogenous choices among winners and among the sequence of wait-listed lottery losers.

Panel A also shows that lottery winning increased students' cumulative years of immersion program enrollment, averaged across observed grade levels, by about 2.2 years. It increased the students' received per-pupil spending by $114 per year, on average. As noted above, the average per-pupil expenditure per year in Portland during the study period was approximately $10,800 per year, so this represents a spending premium of about 1%, concentrated mainly on the provision of logistical support and professional development for DLI teachers. Note that this first-stage differential spending estimate as a function of winning an immersion lottery is smaller than the 1.7% to 4.2% annual immersion spending differentials we discussed above with respect to Figure 1 due to imperfect compliance with lottery status. Winning the immersion lottery raised the amount of cumulative per-pupil spending students

[7] The first stage estimate for enrollment in kindergarten was similarly 0.456 (not shown), with an F statistic of 118.45.

received by $653, on average across the years in which they were observed. The first-stage estimates for math are nearly identical to those in ELA because almost all students were tested in both subjects in any given grade.

Causal Effects of Immersion Enrollment and Dosage on Student Achievement

Panel B of Table 3 illustrates the observed relationships of the first-stage input differentials to student achievement in ELA and math. As described above, the point-in-time immersion enrollment effect and the cumulative enrollment effects are causally identified holistic effects, as they satisfy the exclusion restriction by which lottery status affects student achievement.

We find that the causal effect of immersion enrollment in any given year was 0.22 of a standard deviation in ELAand was statistically significant at the five-percent level. This represents a substantial benefit for those whose enrollment in any given year depended on their winning an immersion slot in a binding lottery. The corresponding LATE estimates for math, at 0.125 and 0.139 (columns 6 and 7 respectively), did not reach statistical significance, so we consider these to be non-effects in terms of generalizing beyond the study sample.[8]

Our estimates suggest that in ELA, the cumulative immersion dosage effect was also positive. The LATE estimate of the cumulative effect of each additional year of immersion exposure was 0.04. This means that each additional year of immersion enrollment raised reading achievement an additional 4% of a standard deviation, on average. In math, the cumulative

[8] Corresponding effects from kindergarten enrollment rather than point-in-time enrollment were 0.203 ($p<.05$) for reading, and a nonsignificant 0.125 for math. The attenuation relative to point-in-time enrollment is to be expected, since enrollment in kindergarten captures immersion exposure less precisely than point-in-time enrollment.

dosage effect was 0.026 of a standard deviation, but the standard error was large, and the estimate did not approach statistical significance.

Relationships of Per-Pupil Expenditure Differences to Student Achievement

Table 3 also presents descriptive estimates of the role of differential per-pupil expenditures in mediating the relationship between lottery status and student achievement. As noted above, we limit our analysis to per-pupil expenditure differentials at the district level, since this is where we found differential spending to be concentrated. We find that an additional $100 dollars in per-pupil spending for lottery compliers in a given year was associated with an additional 8% of a standard deviation in students' ELA achievement in any given year (τ=0.08, $p<0.05$), whereas the corresponding estimate for math (τ=0.05 of a standard deviation) did not approach statistical significance. The ELA estimate represents a substantial relationship, given than an additional $100 per year constitutes a spending premium of less than 1%.

Finally, we consider the relationship of cumulative expenditure over time to student achievement. Here, we find that an additional $100 spent on immersion students cumulatively (i.e., above and beyond additional dollars in prior grades) was positively associated with an additional 1.4% of a standard deviation of ELA achievement per year ($p<.05$). The corresponding estimate in math was just under 1% of a standard deviation but did not approach statistical significance.

To put the descriptive per-pupil spending effects into context, it is useful to consider them in comparison to the corresponding, causally identified dosage effects. Per-pupil spending differences are a function not only of students' immersion enrollment dosage but also of the particular year in which they applied to an immersion lottery, since expenditures varied over

time. In ELA, our descriptive effect estimate for an additional $100 of per-pupil spending in a single year (0.081) was 36%—or just over a third—the size of the holistic point-in-time enrollment effect (0.224). Similarly, our descriptive effect estimate for an additional $100 of per-pupil spending cumulatively (0.014) was one-third the LATE estimate for an additional year enrolled in immersion (0.042).

These findings suggest that about one third of the causal dosage effect in ELA is associated with per-pupil expenditure differentials, though we cannot strictly say that the dosage effects would have been lower by a third had there been no spending differential. It remains possible that other program features changed over time in ways that were correlated with, but not explicitly driven by, these spending differentials.

Relationships of Classroom and Teacher Characteristics to Student Achievement

To further address research question 2, we consider the descriptive point-in-time relationships of teacher and classroom characteristics to students' ELA performance as a function of winning an immersion lottery. In Table 4, we focus on the teacher and classroom characteristics for which we only have single-year data from 2012-13. Panel A of Table 4 presents first-stage estimates from separate regressions of each classroom and teacher characteristic variable from 2012-13 on the lottery winning indicator. As shown in Panel A, we find that winning the immersion lottery did increase the share of students' classmates in 2012-13 who were ELs and who were Hispanic. It also reduced the share who qualified for special education services and the share who were white. It negatively predicted the share eligible for subsidized meals, but that relationship was only marginally significant ($p<0.1$).

[Insert Table 4 about here]

Though lottery winning did drive these classroom attributes in the first stage, we find no evidence that these differences were associated with the academic benefits of immersion on ELA achievement. In Panel B of Table 4, we find that the instrumented effects on ELA of the share of peers in class who qualified for subsidized meals or who were English language learners were positive but very noisy. Estimates for the instrumented effects of other classroom and teacher attributes are negative but noisy. None of the instrumented effects approaches statistical significance. In other words, the small shifts in classroom and teacher attributes that appear to result from winning an immersion lottery have virtually no explanatory power in accounting for the positive, holistic ELA effects we identified.

In a separate analysis available from the authors upon request, we examine the same teacher and peer characteristics as mediators of point-in-time math achievement. We find results very similar to the ELA estimates, with no evidence that teacher or peer characteristics mediate the effect of DLI programs on students' math achievement. Given that we find no statistically significant intent-to-treat (ITT) effects of immersion programs on math achievement, our finding of no mediating effect for teacher or peer characteristics is not surprising.

Our finding that classroom characteristics were not strongly associated with the effects of DLI in ELA—at least within our point-in-time data—suggests that other mechanisms may be in play. These include, plausibly, the language of instruction. Of course, given that lottery winners enroll in classrooms of mostly other lottery applicants, whereas their counterparts do not, it remains possible that unobserved differences in peer characteristics—attributes such as families' educational priorities—contribute to the effects. It is also important to note that, because the classroom attribute data come from 2012-13 only, the point-in-time estimates have less power than the longitudinal estimates in Table 3. Nevertheless, the magnitudes are inconsistent with a

scenario in which immersion effects were strongly driven by sorting toward a more advantaged peer set.

Differential Effects of Immersion by Students' Race/Ethnicity

Question 3 asks whether causally identified immersion effects appear to differ by students' race/ethnicity. In Table 5 we report the LATE estimates of DLI enrollment on students' achievement in ELA and math in a given year. We do not show first-stage estimates because each enrollment-by-subgroup interaction has its own first stage, but all interactions are strongly instrumented by statistically significant lottery status-by-subgroup interactions.

[Insert Table 5 about here]

Though the main effects in ELA are statistically significant as anticipated, we find no statistically significant differential effects in terms of either ELA or math. Of course, our power to disaggregate effects by subgroup is more restricted than in the overall analysis. Though the subgroup effects are causally identified, the power constraints make it difficult to extrapolate them beyond the sample. Still, our results are consistent with the disaggregated but not causally identified subgroup effects reported by Thomas and Collier (2014), in that we find larger within-sample benefits for Black than for White students. In terms of LATE estimates, we find that both ELA and math effects of immersion are much more positive for Black students than for White students, with an estimated point-in-time achievement differential of 0.72 of a standard deviation in ELA and 0.74 of a standard deviation in math, though the standard errors of both estimates are large, and neither approaches statistical significance. (Corresponding ITT differentials, which are not tabulated, are a non-significant 0.04 and 0.09, respectively.) Estimated differential effects for Asian students are negative in ELA (-0.52 on the interaction, yielding a negative net effect of -

0.28 of a standard deviation) and positive in math (0.32 of a standard deviation), but again, neither differential approaches statistical significance. This finding is perhaps somewhat consistent with Valentino and Reardon's (2015) estimation of more-positive DLI effects for Latino than for Chinese ELs, though the subgroup analysis here is not limited to ELs. Also, the differential effect estimates should be cautiously interpreted. Our causally identified effects for racial/ethnic subgroups capture the differential experiences of lottery winners versus losers in the same racial/ethnic group and lottery application stratum. Because randomized students from Asian backgrounds disproportionately applied to the Chinese and Japanese programs in Portland, whereas Black, Hispanic, and White students disproportionately applied to the district's Spanish programs, the subgroup effects for students of different backgrounds will, of course, reflect the effectiveness of the immersion programs to which those students applied as compared to the schools they would have otherwise attended. For this reason, our estimates shed light on causal effects for each subgroup in Portland, but subgroup effects in other contexts would depend in part on the relative effectiveness of immersion and non-immersion programs available to students of each subgroup. The Portland-based estimates show us what is happening, in other words, in a large urban district with a well-established, large-scale system of immersion programs, but effects will always depend on how well such programs are implemented and on the relative quality of the next best alternative available to students.

Discussion and Conclusion

Our analysis offers a substantial contribution to the thin research base on the costs of DLI programs. Whereas prior cost studies have focused on the cost of DLI education as compared with alternative services for ELs, Portland did not treat two-way immersion as a type of EL

service and instead offered consistent services for ELs regardless of whether they were placed in immersion or non-immersion settings. Thus, our cost analysis focused on the differential costs of immersion programs relative to monolingual instruction.

Through central office and principal interview data, we sought to understand the sources of differential inputs at the district and school levels. We supplemented these data with information about class sizes in immersion and non-immersion classrooms in the district, and with historical budget data about revenue sources for the district's immersion programs. We found that resources within schools had been distributed for parity across programs. Where discrepancies existed, they were situated at the central office level, where administrators strategically managed the programs, ensuring that teachers had adequate curriculum and that both teachers and principals were well-supported in carrying out the tenets of the DLI model.

The central office per-pupil immersion differentials ranged from $163 to $482 during the study period, relative to an average per-pupil expenditure in the district of about $10,900 a year, on average. We found additional per-pupil expenditures of $114 per year for immersion lottery winners relative to immersion lottery losers in any given year based on actual dosages received as a function of winning the immersion lottery. This represents about 1% of per-pupil spending in the district during the study period. This is a smaller per-pupil spending differential than prior studies have found, with differentials ranging from 7% (Lara-Alecio et al., 2004) to 12% per year (Parrish, 1994). However, prior studies examined differentials for ELs, which were precluded by the consistent provision of EL services regardless of immersion placement in Portland. Moreover, if EL-focused resources at the central office level were excluded from Parrish's (1994) analysis, then per-pupil expenditures for two-way immersion would actually have been about 3.9% lower than for non-immersion classes due to slightly lower use of instructional aides.

This is the first study we are aware of to estimate the role of per-pupil cost differentials as a potential mediator of student achievement. We estimate that an additional $100 of per-pupil spending on immersion was associated with an additional 8% of a standard deviation of ELA achievement, on average across grades, for students who enrolled in immersion as a function of winning the immersion lottery. This is just over one third the size of the causally identified LATE for immersion enrollment in a given year, meaning that just over a third of the enrollment effect corresponds to variation in per-pupil spending over time. There was no statistically significant corresponding effect for math. On average, lottery winners received an additional 2.18 years in immersion as a function of their lottery-winning status, and received an additional $654 in per-pupil spending from kindergarten entry during the period under observation (third grade for the youngest cohort, and eighth grade for the oldest two cohorts). This yielded a cumulative ELA achievement effect of 4% of a standard deviation per enrolled year, and 1.4% of a standard deviation for each $100 spent cumulatively. These are substantively meaningful effects. By comparison, several studies have found that a difference of one-standard deviation in teacher effectiveness in a given year—a very large, hard-to-scale difference—yields only about a tenth of a standard deviation in students' academic achievement in that year (Kane & Staiger, 2005; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). And the Tennessee STAR experiment of class size effects in the early grades found that reducing kindergarten through third grade class sizes from about 25 to 15 students yielded first-grade ITT student achievement gains of about a quarter of a standard deviation in math and ELA (Mosteller, 1995), at an estimated cost of about $13,100 per student, scaled in 2004 dollars (Levin, Belfield, Muennig, & Rouse, 2007). This suggests that each $100 would produce just under two thousandths of a standard deviation (0.002 SD) in student achievement in the ITT analysis. These comparisons suggest that the DLI effects

we observed in ELA were strikingly cost-effective, assuming that other contexts could replicate Portland's emphasis on centralized, district-wide support for its diverse school programs. And recent lottery-based research suggests that our estimates of immersion effectiveness at scale could even be conservative. Our estimate of a cumulative LATE in reading of 4% of a standard deviation per year is similar to Bibler's (2017) cumulative LATE estimates in ELA of between 5% and 6% of a standard deviation per year, while our non-significant math estimate of just under 3% of a standard deviation is actually much smaller than his math estimates of 6% to 9% of a standard deviation per year.

Though we do not find statistically significant effects in math, DLI in Portland may have yielded benefits beyond test scores on English-administered accountability tests. For instance, immersion students in the study reached intermediate mid-to-high levels of proficiency in a second language by eighth grade, as compared to the early novice levels reached by eighth graders who had taken a foreign language elective class in Spanish (Burkhauser et al., 2016)—an advantage that may have economic as well as cognitive and language-acquisition benefits (Saiz & Zoido, 2005; Luk & Bialystok, 2014; Cenoz, 2003).

Importantly, we find no evidence that observable teacher or peer characteristics mediated the immersion effects in ELA. Our analysis does not rule out the possibility that unobserved classroom or teacher characteristics may have played a role, but it clarifies that observable differences in classroom environments were not driving the ELA effect estimates.

Finally, because policymakers interested in closing historical opportunity and achievement gaps may be concerned about differential effects of DLI by students' race/ethnicity, we have also examined the extent to which race/ethnicity moderates the causal effect of DLI enrollment. We do not find statistically significant differences in causal immersion effects by

subgroup, though admittedly, our power to detect differences at statistically significant levels is constrained. From a descriptive standpoint within the sample, we find more-positive causal immersion effects in ELA for Black students than for White and Hispanic students, whose effects were nearly identical, and we find negative ELA effects (and positive math effects) for Asian students, but again, none of these subgroup differences are generalizable beyond the sample. Also, it is not clear what drives these differences. The differential effects by race/ethnicity in the sample may be driven by unobserved attributes of the immersion programs to which the students applied relative unobserved attributes of their default catchment schools. Still, the finding is policy-relevant within Portland, given that Black students were underrepresented in Portland's immersion programs during the period of the study. The district has since taken steps to expand access to immersion in historically African American neighborhoods of the city.

Our work suggests that positive immersion effects in ELA can be achieved at scale with modest investments at the central office level, concentrated on supporting high-quality dual-language instruction through professional development and curriculum support. The field would continue to benefit from work that carefully examines immersion cost differentials and associated effects on student achievement and attainment in various settings. There is a particular need for longer-term work that examines students' graduation rates, college-going, and employment as a function of random assignment to DLI. Ideally, future research would also examine the causal effects of immersion on students' non-academic development, including their global awareness and civic participation. In the meantime, our study provides new evidence for the cost-effectiveness of immersion when implemented with central office support, and it attests to the broader potential of such programs for policy support and replication.

References

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

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.

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

Bibler, A. (2017). Dual language education and student achievement. (Working paper). Anchorage, AK: University of Alaska Anchorage.

Bifulco, R., Unterman, R., & Bloom, H. S. (2014). *The relative costs of New York City's new small public high schools of choice* (SSRN-id2574386). Retrieved from Syracuse, NY: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2574386

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

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

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

Center for Applied Linguistics (2011a). Directory of foreign language immersion programs in U.S. schools. Retrieved September 15, 2011, from http://www.cal.org/resources/immersion/

Center for Applied Linguistics (2011b). Directory of two-way bilingual immersion programs in the U.S. Retrieved September 15, 2011, from www.cal.org/twi/directory/

Christian, D., Montone, C. L., Lindholm, K. J., & Carranza, I. (1997). *Profiles in two-way immersion education*. Washington, DC: Center for Applied Linguistics and Delta Systems.

Christofides, L. N., & Swidinsky, R. (2010). *The economic returns to a second official language: English in Quebec and French in the Rest-of-Canada*. (2010-04). Institute for the Study of Labor.

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.

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

and national security. Washington, DC: Committee for Economic Development. https://www.ced.org/pdf/Education-for-Global-Leadership.pdf

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

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

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

Geisler, M. (2012, March 6). Larry Summers is wrong about languages. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/views/2012/03/06/geisler-essay-why-larry-summers-wrong-about-languages#ixzz23insfmo9

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

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review, 105*(4), 765-789.

Johnson, M., Lipscomb, S., Gill, B., Booker, K., & Bruch, J. (2012). Value-added models for the Pittsburgh Public Schools. Washington, DC: Mathematica Policy Research.

Kane, T. J., & Staiger, D. O. (2005). *Using imperfect information to identify effective teachers*. Cambridge, MA: National Bureau of Economic Research.

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

Klein, E., C. (1995). Second versus third language acquisition: Is there a difference? *Language Learning, 45*, 419-465.

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

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37-62.

Lara-Alecio, R., Galloway, M., Mason, B., Irby, B. J., & Brown, G. (2004). *Texas dual language program cost analysis*. College Station, TX: Texas A&M.

Levin, H., Belfield, C., Muennig, P., & Rouse, C. (2007). *The costs and benefits of an excellent education for all of America's children*. New York: Teacher's College, Columbia University.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd Edition). Thousand Oaks, CA: Sage Publications, Inc.

Lindholm-Leary, K. J. (2001). *Dual language education*. Avon, England: Multilingual Matters.

Lindholm-Leary, K. J. (2005). *Review of research and best practices on effective features of dual language education programs*. San Jose, CA: San Jose State University.

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.

Luk, G. & Bialystok, E. (2013) Bilingualism is not a categorical variable: interaction between language proficiency and usage. *Journal of Cognitive Psychology*. January, 605- 621.

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

Maxwell, L. A. (2012). 'Dual' classes see growth in popularity. *Education Week Spotlight*. 1-3.

Mosteller, F. (1995). The Tennessee Study of Class Size in the Early School Grades. *The Future of Children, 5*(2), 113-127.

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. Retrived January 28, 2015, from https://drive.google.com/file/d/0B0K1-ZK0_vg5cktjbjBiYXNva28/view

Parkes, J., Ruth, T., Anberg-Espinoza, M., & De Jong, E. (2009). *Urgent research questions and issues in dual language education: Dual language researcher convocation report*. Santa Fe, NM: Dual Language Education of New Mexico, University of New Mexico.

Parrish, T. B. (1994). A cost analysis of alternative instructional models for limited English proficient students in California. *Journal of Education Finance, 19*, 256-278.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.

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

Santibañez, L., & Zárate, M. E. (2014). Bilinguals in the U.S. and college enrollment. In R. M. Callahan & P. C. Gándara (Eds.), *The bilingual advantage: Language, literacy, and the U.S. labor market*. Tonawanda, NY: Multilingual Matters.

Schneider,A. (2013) 29 new dual language programs to open. *InsideSchools*. June 17. http://insideschools.org/blog/item/1000679-29-new-dual-language-programs-to-open-in-fall

Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods* (NCEE 2014-4017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Short, D. J., Echevarria, J., & Richards-Tutor, C. (2011). Research on academic literacy development in sheltered instruction classrooms. *Language Teaching Research, 15*(3), 363-380. doi:10.1177/1362168811401155

Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015). The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review, 48*, 86-101. doi:10.1016/j.econedurev.2015.05.009

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

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics, 20*, 518-529.

Stock J., & Yogo M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D.W.K., *Identification and inference for econometric models* (pp. 80-108). New York: Cambridge University Press.

Thomas, W. P., & Collier, V. P. (2014). *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. *The Alberta Journal of Educational Research, XLIX*(1), 6-23.

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.

U.S. Census Bureau. (2017). Table 11: States ranked according to per pupil elementary-secondary public school system finance amounts: 2004-05. *2005 Public Elementary-Secondary Education Finance Data.* Retrieved from https://www.census.gov/data/tables/2005/econ/school-finances/secondary-education-finance.html

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

Watzinger-Tharp, J., Swenson, K., & Mayne, Z. (2017). The academic achievement of Utah dual language immersion students. *International Journal of Bilingual Education and Bilingualism*.

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

Zwerdling, K. (2012). The value of bilingual/multilingual medical staff in the healthcare service industry. Retrieved from http://www.foreignstaffing.com/blog/the-value-of-bilingual-multilingual-medical-staff-in-the-healthcare-service/

Table 1. Descriptive statistics for applicants to binding lottery strata in the analysis

	Binding Lottery Applicants Only				
Variable	**All**	**Won Slot**	**Not Placed**	**Differ-ence (unadj)**	**p Diff (strata-adj)**
N	1,625	752	873		
Proportion		0.463	0.537		
Female	0.529	0.508	0.546	-0.038	0.15
Asian	0.144	0.178	0.115	0.064	0.61
Black	0.056	0.052	0.060	-0.008	0.77
Hispanic	0.170	0.177	0.164	0.013	0.65
White	0.540	0.517	0.559	-0.042	0.25
Other Race	0.068	0.063	0.073	-0.011	0.01
Subsidized meals	0.260	0.273	0.250	0.023	0.63
Sp. Needs in K	0.041	0.052	0.032	0.020	0.29
Gifted in K	0.040	0.044	0.037	0.007	0.63
EL in K	0.127	0.153	0.105	0.048	0.91
First Lang Not Eng.	0.180	0.206	0.157	0.049	0.42
First Lang Partner	0.113	0.138	0.092	0.047	0.92
Ns By Grade					
Grade K	1,625	752	873		
Grade 1	1,625	752	873		
Grade 2	1,625	752	873		
Grade 3	1,589	729	860		
Grade 4	1,254	570	684		
Grade 5	983	428	555		
Grade 6	690	289	401		
Grade 7	517	196	321		
Grade 8	343	123	220		
Grade 9	179	56	123		

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.

Table 2. Within-school resources for DLI and non-DLI programs based on principal interviews (n=14)

Other Principal Inputs	DLI Mean (and SD)	Non-DLI Mean (and SD)	Mean Difference
Teacher Hours Beyond Contract Per Week: Principal Estimates	10.33 (8.43)	6.32 (5.88)	4.01
Parent Hours Per Student Per Week	0.11 (0.10)	0.10 (0.14)	0.01
Number of One Day Field Trips Per Year	2.89 (1.24)	2.93 (1.27)	-0.04
Number of Travel Field Trips Per Year	0.46 (0.66)	1.36 (2.65)	-0.90
Principal Reports Extra Tech in DLI Classes	0.54		
Grant Dollars Per Student, Inclusive*	45.93 (117.39)	114.17 (101.39)	-68.24
Net Grant Dollars Per Student (Excludes Field-Trip Specific and Central Office Pass-Through)*	12.23 (20.48)	114.17 (101.39)	-101.94
School-by-Year Ratio of Net DLI Grants to Total Grants	0.02 (0.03)		

*For grants, non-DLI mean refers to grant dollars that are generally used school-wide, not just in non-DLI programs.

Table 3. First- and second-stage estimates from 2SLS regression models estimating instrumented effects of immersion exposure and differential cost on achievement

Panel A. First-stage estimates regressing mediators on lottery-winning indicator

ELA (n=1,415 students & 4,608 observations)	Coeff.	Std. Error	F model (87, 4520)	Math (n=1,447 students & 4,632 observations)	Coeff.	Std. Error	F model (87, 4544)
Entered Immersion in Kindergarten	0.456***	(0.014)	118.45***	Entered Immersion in Kindergarten	0.454***	(0.014)	119.58***
Enrolled in Immersion in Given Year	0.412**	(0.019)	51.04***	Enrolled in Immersion in Given Year	0.411**	(0.019)	49.92***
Cumulative Years in Immersion	2.175***	(0.090)	76.78***	Cumulative Years in Immersion	2.173***	(0.090)	76.36***
Additional $100 Per-Pupil Dollars in Given Year	1.141***	(0.067)	34.50***	Additional $100 Per-Pupil Dollars in Given Year	1.138***	(0.068)	33.77***
Additional $100 Per-Pupil Dollars Cumulative	6.535***	(0.259)	86.22***	Additional $100 Per-Pupil Dollars Cumulative	6.533***	(0.259)	86.47***

Panel B. IV-estimated effects of immersion entry, dosage, and spending on ELA and math achievement in SD units

ELA	Coeff.	Std. Error	F model (87, 4520)	Math	Coeff.	Std. Error	F model (87, 4544)
Entered Immersion in Kindergarten	0.203*	(0.093)	10.20***	Entered Immersion in Kindergarten	0.125	(0.102)	8.82***
Enrolled in Immersion in Given Year	0.224*	(0.103)	10.21***	Enrolled in Immersion in Given Year	0.139	(0.112)	8.86***
Cumulative Years in Immersion	0.042*	(0.020)	10.17***	Cumulative Years in Immersion	0.026	(0.021)	9.02***
Additional $100 Per-Pupil Dollars in Given Year	0.081*	(0.037)	10.12***	Additional $100 Per-Pupil Dollars in Given Year	0.05	(0.041)	8.85***
Additional $100 Per-Pupil Dollars Cumulative	0.014*	(0.007)	10.17***	Additional $100 Per-Pupil Dollars Cumulative	0.009	(0.007)	9.06***

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Note: Models include controls for quadratic year specification, gender, race/ethnicity, special needs status in kindergarten, EL status in kindergarten, free/reduced-price lunch indicator in kindergarten, lottery-strata fixed effects, and random error terms at the student and observation levels.

Table 4. First- and second-stage estimates from 2SLS regression models estimating instrumented effects of classroom characteristics on achievement

Panel A. Effects of winning immersion lottery on classroom characteristics in 2012-13

First-stage estimates	Coeff.	Std. Error	n	DF (m)	DF (r)	F (instrum.)
Proportion of students in class who are:						
Subsidized-meal eligible	0.028~	(0.015)	550	71	478	11.626***
EL	0.019**	(0.006)	847	72	774	21.237***
Special education	-0.030***	(0.006)	847	72	774	3.408***
Talented & gifted	-0.003	(0.011)	847	72	774	5.102***
Asian	-0.006	(0.007)	847	72	774	27.28***
Hispanic	0.072***	(0.010)	847	72	774	18.860***
Black	-0.007	(0.006)	847	72	774	8.945***
White	-0.043***	(0.012)	847	72	774	17.966***
Teacher years of experience	-0.262	(0.545)	819	71	747	6.142***
Teacher has advanced degree	-0.016	(0.030)	814	71	742	6.134***
Teacher highly qualified (NCLB)	-0.008	(0.010)	779	71	707	3.136***
Students in classroom	-0.346	(0.374)	847	72	774	3.292***

Panel B. Effects of classroom attributes on ELA scores, instrumented by lottery assignment

Second-stage, instrumented predictors	Coeff.	Std. Error	n
Propotion of students in class who are:			
Subsidized-meal eligible	3.255	(3.428)	550
EL	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 (NCLB)	-14.645	(20.319)	779
Students in classroom	-0.168	(0.269)	847

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Notes: Panel A represents first-stage estimates from instrumental variable models that include lottery strata fixed effects and individual covariates, as in Equation 1. Panel B represents the second-stage IV estimates, as specified in Equation 2.

Table 5. Second-stage LATE estimates of immersion enrollment in a given year on ELA and math performance, disaggregated by students' race/ethnicity

	(1)	(2)
Instrumented predictors	**Reading**	**Math**
Enrolled in immersion	0.242*	0.121
	(0.109)	(0.119)
Enrolled in Immersion * Asian	-0.519	0.320
	(0.332)	(0.359)
Enrolled in Immersion * Black	0.724	0.739
	(0.892)	(0.988)
Enrolled in Immersion * Hispanic	0.0167	-0.229
	(0.311)	(0.338)
Enrolled in Immersion * Other	0.154	0.0937
	(0.338)	(0.366)
Observations	4,608	4,632
Students	1,451	1,447
Intraclass Correlation	0.631	0.650
F model (91, 4517_{rd}; 91, 4541_{m})	9.594	8.298

Note: Models include controls for quadratic year specification, gender, race/ethnicity, special needs status in kindergarten, ELstatus in kindergarten, free/reduced-price lunch indicator in kindergarten, lottery-strata fixed effects, and random error terms at the student and observation levels.

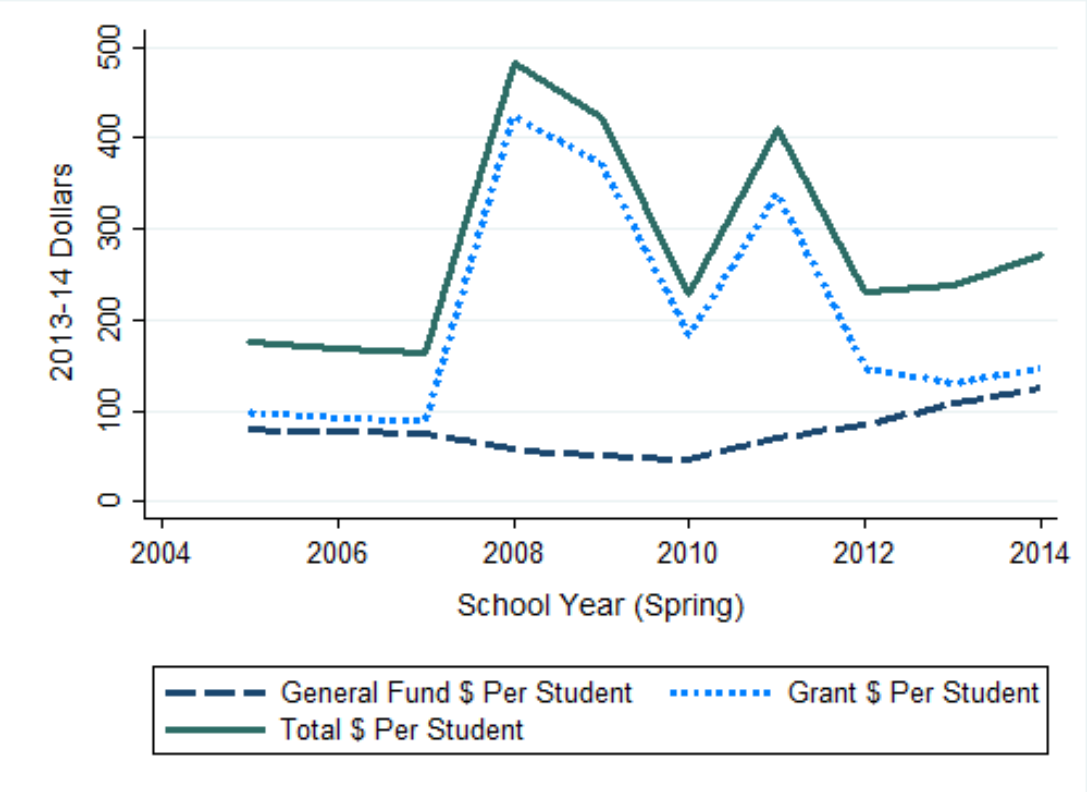

Figure 1. Centralized expenditures per student on immersion programs, by source and year, in 2014 dollars

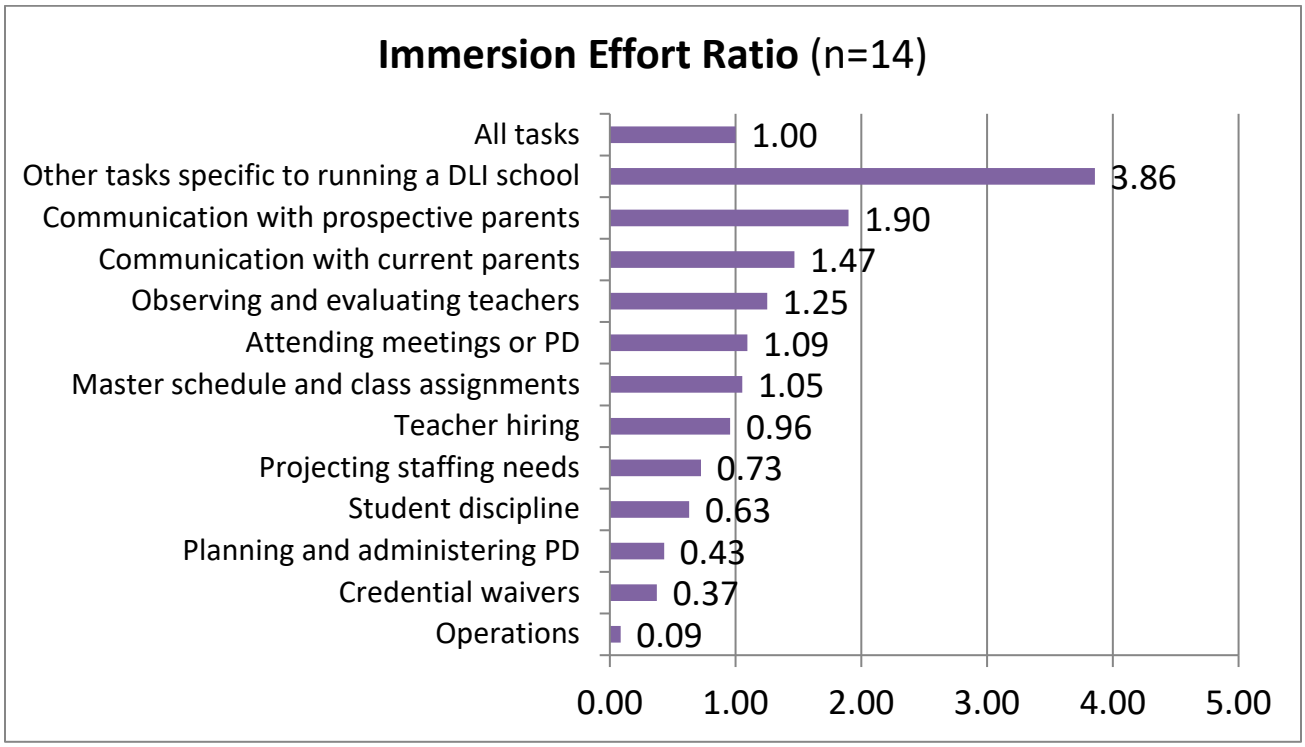

Figure 2. Ratio of principals' time devoted to a given task in immersion vs. non-immersion programs, adjusted for the share of the student body in immersion