

RESEARCH BRIEF

Student Performance Under Dual Language Immersion Scale-Up in Utah

Prepared by the Partnership to Study Dual Language Immersion in Utah
Jennifer L. Steele[†], Johanna Watzinger-Tharp[‡], Robert Slater[§], Gregg Roberts[§], Karl Bowman^{◊1}

With financial support from the U.S. Department of Education's Institute of Education Sciences and data support from the Utah State Board of Education and the Utah Dual Language Program Directors²

September 2019

Overview

In this research brief, we highlight findings from the federally funded Partnership to Study Dual Language Immersion in Utah regarding the academic performance of dual language immersion (DLI) students in Utah. DLI is an instructional model designed to cultivate bilingualism and biliteracy through immersive instruction in core academic content areas such as mathematics, science, and language arts. Students in DLI programs receive core-content instruction in two languages from the early grades onward. In Utah, DLI typically starts in first grade.

Our analysis examines the academic achievement of Utah students in both *oneway* and *twoway* DLI programs. In oneway programs, students in a classroom share a common native language; they receive academic instruction not only in that language but also in a classroom partner language they are all learning together. In twoway programs, students in a classroom hail from both native language groups, with a substantial share (in Utah, at least a third) being native speakers of the classroom partner language.

In this research brief, we first describe the achievement of Utah students enrolled in DLI in grades 3-6, demonstrating that they outperform their non-DLI peers in the same schools by up to 30 percent of a standard deviation. Next, we leverage the expansion of access to DLI over a ten-year period to isolate how much of this outperformance is attributable to the state's expansion of DLI. Here we find muted effects of DLI access, with modest positive effects of twoway programs in grade 3 and 6. We also report on a teacher survey we conducted in the Granite School District showing consistency in DLI implementation across schools.

Research Context and Objectives

In 2008, Utah placed itself on the vanguard of dual language immersion education in the U.S. with its passage of Senate Bill 41, which provided funding for public schools to open or expand

¹ American University[†], University of Utah[‡], American Councils for International Education[§], Utah State Board of Education[◊]

² We are particularly grateful to Kristin Campbell, Research Analyst at the Utah State Board of Education, for providing all student-by-year data from the state, and to the DLI directors in Alpine, Cache, Canyons, Davis, Granite, Tooele, and Wasatch School Districts for providing historical student-level DLI enrollment records from 2015-16 and before.

DLI programs across the state. After starting from just a handful of programs in 2009, by the 2018-19 academic year, Utah had become home to 224 DLI programs: 113 in Spanish, 65 Chinese, 30 French, 13 Portuguese, 2 German, and 1 Russian. In 2017-18, about 18% of schools in the state offered DLI programs, and about 5.2% of public school students (roughly 34,000 students) were enrolled in one of these programs. And in a state where about 9% of students were classified as English learners in 2017-18, current English learners accounted for about 12% of DLI enrollments in grades 1 through 9, meaning that they were enrolling in DLI at a higher rate than native English speakers, on average. English learners accounted for about 5% of students in oneway DLI programs and about 41% of students in twoway DLI programs.

For a state aiming to compete effectively in the global marketplace, the expansion was an ambitious but promising endeavor. Because second-language acquisition is accomplished more efficiently at early ages (Kuhl, 2010), exposing students to dual-language instruction from elementary school suggests a more efficient approach to public education. Moreover, numerous studies have demonstrated students' academic outperformance in dual-language programs (e.g., Lambert, Tucker, & d'Anglejan, 1973; Turnbull, Hart, & Lapkin, 2003), particularly but not exclusively for English learners (Collier & Thomas, 2004; Lindholm-Leary & Block, 2010; Thomas & Collier, 2014). Historically, many of these studies have not been able to disentangle the effects of who chooses immersion from the effects of the immersion program itself. This distinction may have limited importance from the perspective of parents, who are likely to wonder about how students in immersion programs actually perform relative to their peers. But for policymakers and scholars, the distinction is important, because it pertains to the effects of programmatic decisions under public control—namely, to what extent does access to dual language immersion *cause* students to perform differently from their peers? Recently, a few studies have tackled this question by leveraging the random assignment of students to immersion programs. These studies have shown modestly positive, plausibly causal effects on students' achievement as tested in English, particularly in language arts and possibly in mathematics (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007; Bibler, 2017; Steele et al., 2017; Umansky & Reardon, 2014; Valentino & Reardon, 2015). Given evidence that immersion students are also attaining proficiency in the partner language (Burkhauser et al., 2016; Watzinger-Tharp, Rubio, & Tharp, 2018) and at relatively low cost to their schools or districts (e.g., Barnett et al., 2007; Steele et al., 2018), the weight of recent evidence attests to DLI as a sound public education investment.

Building on existing research in DLI, Utah established a 50-50 instructional model, in which students receive half of their instruction in the partner language and half in English for grades kindergarten through 6, with partner language instruction in humanities in grades 7 through 8. Students demonstrating adequate proficiency are able to continue with advanced classes in the partner language in high school. To promote instructional consistency, the state adopted curriculum and instructional expectations for all schools taking part in the DLI expansion and created a professional development program to support DLI teachers across the state.

The rapid expansion of DLI in Utah over the past decade not only broadened Utah students' access to language learning opportunities, but also provided a vital research opportunity for illuminating how DLI programs influence student learning. To facilitate this research, the U.S. Department of Education's Institute of Education Sciences (IES) awarded the state a Research-

Practice Partnership grant designed to shed new light on student achievement under the state's DLI scale-up effort. Previous causal studies of DLI effects have relied primarily on immersion lotteries, where winning a DLI lottery affects not just access to dual-language instruction but to existing DLI schools and lottery-applicant peers. The rapid scale-up of DLI in Utah instead offers an opportunity to examine how the launch or expansion of dual-language immersion programs affects student achievement in a school. Insofar as students attend the schools to which they are residentially zoned, this design minimizes the threat of selection bias from families choosing immersion when it is offered and helps our research team isolate the causal effects of immersion access. Moreover, because this expansion occurred within schools over time, this design allows us to estimate an effect of DLI access that is separate from the stable effect of DLI schools and of DLI-choosing peers, providing a clearer estimate of the causal effect of DLI access when scaled across a state.

Because this design parcels out effects not only of individual-level selection, but also of schools and peers, we would expect estimates of the causal effects of DLI access to be smaller than those found in other studies, assuming that at least a fraction of prior estimates was attributable to the stable effects of established DLI schools (leadership, teachers, parental involvement, etc.) or peers. Utah's scale-up plan therefore allows for the first study we know of to examine the causal effects of DLI expansion as existing schools transition from monolingual to dual language, especially when implemented on a large scale. In this way, Utah's natural experiment contributes not just to students' language acquisition opportunities, but to a broader understanding of the relationship of DLI to student achievement.

Finding 1: Descriptively, DLI Students Outperform their Peers

Examining the achievement of DLI students, the first question we address is how well DLI students in Utah are performing on state accountability tests in language arts, mathematics, and science relative to their peers in the same schools and cohorts who are not enrolled in DLI. To make the two groups as comparable as possible, we control statistically for a wide array of students' characteristics as measured in the first year in which they enrolled in a Utah public school. These include students' gender, race/ethnicity, baseline free/reduced-price lunch eligibility, baseline special education status, home language status at baseline (English or other), and migrant status. The models also control for the average income and education levels of students' residential zip code, and for the characteristics of students' same-grade peers in their schools each year, including the fraction who were minority students, who qualified for subsidized meals, who were English learners at baseline, and who qualified for special education services at baseline. We also control for dichotomous indicators of students' cohort and school, meaning that comparisons are within cohort and within school.

Adjusting for all of these factors, we examine how DLI students have performed relative to their peers in grades 3 through 6 in each of the three subjects measured by accountability tests. We present separate estimates for students in oneway and twoway DLI programs (bearing in mind that all twoway programs are Spanish programs) and for students who began school as native English speakers versus those who began school as English learners.

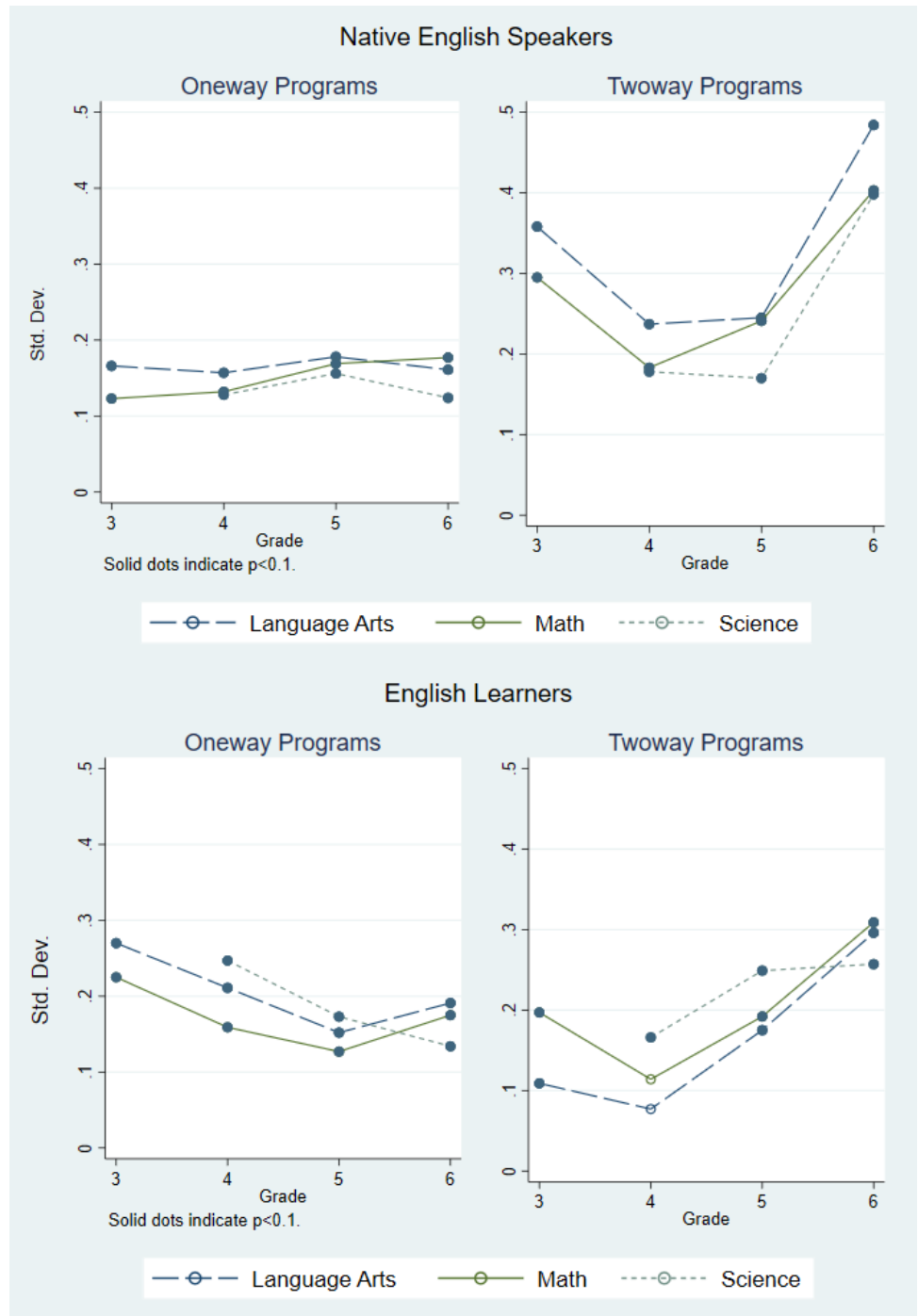
Students' DLI enrollment status is available from all 22 DLI Utah school districts in the 2016-17 and 2017-18 academic years, when these data were collected systematically by the state. Prior to

those years, it is available only for seven districts that provided the project with historical DLI student enrollment data, so only those seven districts contribute to the estimates in 2009-10 through 2015-16. Because the descriptive estimates do not differ notably in the years before and after statewide data availability, we present estimates here based on the most comprehensive DLI enrollment data available. Analyses include about 502,000 unique students, roughly 433,000 native English speakers and 69,000 English learners, enrolled in grades 3 through 6 in Utah between the academic years 2009-10 and 2017-18. Of those, about 26,000 were enrolled in DLI—approximately 21,000 native English speakers and 5,000 English learners.

Estimates for the descriptive analyses are shown in Figure 1. The lines in these graphs represent the relative

average performance of DLI students as compared to observably similar non-DLI students in the same schools and cohorts. Solid point estimates are statistically significant at the 10% level, meaning they are precise enough that we can conclude they are different from 0 in the population. Hollow point estimates are not statistically significant, meaning they cannot be generalized beyond the students in the analysis.

Figure 1. Descriptive performance of DLI relative to similar non-DLI peers



Differences are expressed in standard deviations of student achievement on the state accountability test, where a difference of 0 would mean that DLI and non-DLI students performed identically on average, and positive differences would mean that DLI students outperformed their non-DLI peers.³ In educational interventions, differences of only 0.1 to 0.2 of a standard deviation are often considered substantively meaningful, as educational treatment effects larger than this are seen infrequently.

The differences we observe in Figure 1 hover between 0.12 and 0.18 of a standard deviation for native English speakers in oneway programs. A difference of 0.18 of a standard deviation means that for a non-DLI student scoring at the 50th percentile, an observably similar DLI peer would be expected to score at the 57th percentile. Estimates for native English speakers in twoway programs are higher, reaching as high as 0.48 of a standard deviation in sixth grade. For English learners in both oneway and twoway programs, estimates are also positive and statistically significant, ranging between 0.11 and 0.25 of a standard deviation.

Table 1. Descriptive estimates aggregated across grades

Subject Area	n students-by-year	n schools	Estimate for DLI Enrollment in Oneway Programs	Differential (Interaction) Estimate for Enrollment in Twoway Programs	Net Estimate for Enrollment in Twoway Programs
Native English Speakers					
Language Arts	3,762,204	1016	0.177*** (0.016)	0.126* (0.062)	0.303***
Math	3,449,220	1013	0.160*** (0.020)	0.118~ (0.064)	0.278***
Science	3,071,524	1007	0.168*** (0.018)	0.035 (0.079)	0.203***
English Learners					
Language Arts	553,747	959	0.232*** (0.038)	-0.091 (0.061)	0.141***
Math	530,854	960	0.195*** (0.041)	-0.007 (0.067)	0.188***
Science	447,727	954	0.225*** (0.039)	-0.019 (0.098)	0.206***

~p<0.1, *p<0.05, **p<0.01, ***p<0.001
Standard errors are shown in parentheses.

Table 1 summarizes overall estimates for oneway and twoway programs without disaggregating by grade. Native English speakers enrolled in DLI outperformed their peers by about 16% to 18% of a standard deviation in oneway programs and by up to 30% of a standard deviation in twoway programs. English learners in DLI outperformed their peers by 20% to 23% of a standard deviation in oneway programs and by 14% to 21% of a standard deviation in twoway programs.

The findings in Figure 1 and Table 1 clarify

³ The state accountability test was the Utah Criterion-Referenced Test (CRT) through 2012-13, and the Utah Student Assessment of Growth and Excellence (SAGE) from 2013-14 onward. To achieve a common scale, scores are standardized to mean 0 and standard deviation 1 within each subject, grade, and year.

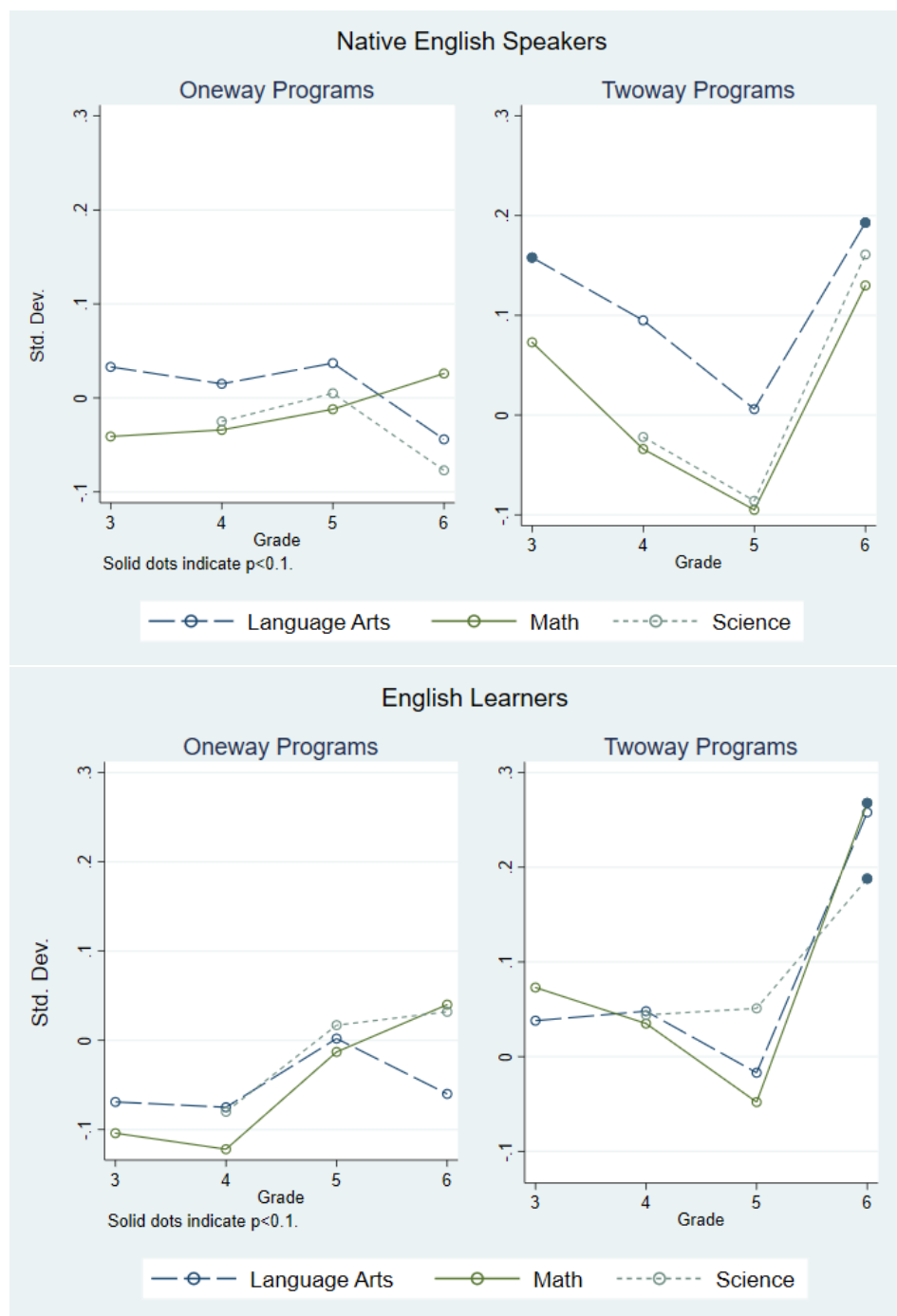
that DLI students were indeed outperforming their same-school peers who were not enrolled in DLI, and they were doing so in all three core content areas by educationally meaningful and statistically significant margins.

Finding 2: Expansion of DLI Access Has Had Little Causal Effect on Student Achievement

The descriptive estimates of DLI students' relative performance are important, but they tell only part of the story. The second part lies in disentangling the causal effects of DLI access on student achievement from the unobserved characteristics of students and families who choose immersion when it is offered. By examining intent-to-treat effects of DLI access when students were first graders on their subsequent achievement (regardless of whether they themselves enrolled in DLI), we find little clear effect of DLI access on achievement.

Fortunately, the rapid scale-up of DLI within Utah, facilitated by the state's concerted investment, provides a natural experiment that allows our

Figure 2. Estimated intent-to-treat effects of DLI access on student achievement



research team to disentangle the effects of access from the unobserved attributes of the families who chose immersion. To obtain causal estimates of the effect of DLI access on student performance—that is, the effect of a school’s offer of DLI on the average achievement of its students—we examine what happens to students’ achievement over time as a function of the availability of DLI slots in their initial school in their first grade year. This availability variable, defined as DLI slots per first grader, becomes our causal “intent-to-treat” variable, since it does not depend on a student’s choosing or not choosing DLI when it is offered. We use this intent-to-treat variable to estimate the causal effect of DLI access on students’ core content achievement in grades 3 through 6 using statistical models that are otherwise the same as those in the descriptive analysis.

Table 2. Intent-to-treat estimates aggregated across grades

Subject Area	n students-by-year	n schools	Effect of Slots Per First Grader in Oneway Programs	Differential Effect of Slots in Twoway Programs	Net Effect of Slots in Twoway Programs
Native English Speakers					
Language Arts	3,762,204	1016	0.03 (0.026)	0.063 (0.055)	0.093
Math	3,449,220	1013	0.001 (0.031)	0.043 (0.055)	0.044
Science	3,071,524	1007	-0.001 (0.034)	-0.034 (0.070)	-0.035
English Learners					
Language Arts	553,747	959	-0.026 (0.053)	0.051 (0.073)	0.025
Math	530,854	960	-0.013 (0.073)	0.063 (0.096)	0.05
Science	447,727	954	0.002 (0.062)	0.021 (0.123)	0.023

~p<0.1, *p<0.05, **p<0.01, ***p<0.001
Standard errors are shown in parentheses.

Results by grade are shown in Figure 2, and cross-grade estimates appear in Table 2. The analytic sample includes about 502,137 public school students enrolled in grades 3 through 6 between 2011-12 and 2017-18.

As we would anticipate, the causal effects of DLI access on student achievement in a school are more muted than the descriptive relationships in Figure 1. As before, hollow point estimates are not statistically

distinguishable from 0. Estimates can be interpreted as the effect of increasing the share of DLI students from 0% to 100% in a student’s school in his or her first grade year. (This interpretation is due to the scaling of the slots variable. In practice, most DLI schools offer between 33% and 66% of their first-grade slots as DLI.)

We do not find statistically significant effects of oneway program access on student achievement in language arts, mathematics, or science for native English speakers (the top left panel of Figure 2) or for English learners (the bottom left panel). We do, however, find some evidence that providing access to twoway programs raised student achievement for native English speakers by about 0.16 to 0.19 of a standard deviation in language arts in grades 3 and 6, where 16% of a

standard deviation is equivalent to raising a student from the median to the 56th percentile. For English learners, we find positive and significant effects of DLI access in grade 6 in mathematics, at about 0.28 of a standard deviation—the equivalent of raising a student from the median to the 61st percentile—and of about 0.19 of a standard deviation in science, the equivalent of raising a student from the median to the 58th percentile. Estimates in other subjects and grades are not statistically distinguishable from 0.

In Table 2, we aggregate the intent-to-treat estimates across grade levels. We find that none of the aggregated estimates reaches statistical significance at the 10% level.

In additional analyses (not shown), we do find modest variability in intent-to-treat estimates by program language, though estimates vary markedly among tested subjects and grades and are difficult to generalize. We also find variability in estimates between districts, but these also vary from grade to grade and subject to subject. This variability is not well explained by the baseline achievement levels or demographic attributes of the districts or by how long the programs have been in operation.

Finding 3: Teacher Surveys from Sample District Suggest Consistent Implementation

In addition to DLI enrollment and causal effects of DLI on academic achievement, the IES grant supported the pilot-testing of a DLI implementation analysis. After field-testing questions and conducting a pilot survey of 16 teachers selected by DLI administrators in several districts, in the spring of 2019 we conducted a district-wide survey in Granite School District with 62 elementary DLI teachers (grades 1-6) in Chinese, French, and Spanish, including teachers in both oneway and twoway programs.

Survey questions addressed five implementation domains: target language instruction and target language use, core instructional strategies, target language materials, teacher collaboration, EL enrollment and support, and program support.

Teachers' self-reported descriptions of their implementation (core-subject time allocation, partner language use, partner teacher collaboration, and ELL enrollment) suggested fidelity to the Utah DLI model. Teachers reported strong support from their principals for 100% target language use and DLI access for all. Overall, implementation in the district appeared to be consistent across languages and program types.

Conclusion

The rapid expansion of dual language immersion in Utah not only increased students' access to language acquisition opportunities in their public schools but also provided a natural experiment through which researchers and educators could better understand the effects of dual language immersion when scaled across a state. Consistent with other studies, this study finds that students enrolled in dual language immersion programs in Utah have markedly outperformed their same-school counterparts on standardized tests in language arts, mathematics, and science, all of which are tested in English. This is true for native English speakers, with cross-grade descriptive estimates of up to 30% of a standard deviation, and it is true for English learners, whose cross-grade descriptive estimates reach 23% of a standard deviation.

Though we find limited evidence that the expanded access to DLI in public schools has caused higher student achievement, except possibly in twoway programs in grades 3 and 6, we also find no evidence that schools' provision of DLI access or the delivery of instruction in a language other than English has harmed or slowed students' achievement. Given the disruption often caused by large-scale instructional reforms (Eastwood & Louis, 1992), and given that accountability tests are administered in English, this is a noteworthy finding.

Utah launched its DLI initiative with the goal of preparing a globally competitive bilingual and biliterate workforce. A recent analysis of partner language assessments (Watzinger-Tharp et al., 2018) found that DLI students are meeting or exceeding partner language benchmarks at nearly all grade levels and across linguistic skills. Furthermore, descriptive evidence from the state suggests that many in the early cohorts are earning Advanced Placement (AP) credit in their partner languages. If Utah students are achieving these markers of progress toward bilingualism and biliteracy without a detriment to their core content achievement, then the state can be considered en route to achieving its goals for DLI.

References

- Barnett, W. S., Yarosz, D., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly, 22*, 277-293. doi: 10.1016/j.ecresq.2007.03.003
- Bibler, A. (2017). *Dual language education and student achievement*. Anchorage, AK: University of Alaska Anchorage. Retrieved from <https://drive.google.com/file/d/0B245E5sJ4KGHcUt2MDh6Nmt1STg/view>
- 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.
- 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.
- Eastwood, K. W., & Louis, K. S. (1992). Restructuring that lasts: Managing the performance dip. *Journal of School Leadership, 2*(2), 212-224. doi: 10.1177/105268469200200206
- Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713-727. doi: 10.1016/j.neuron.2010.08.038
- Lambert, W. E., Tucker, G. R., & d'Anglejan, A. (1973). Cognitive and attitudinal consequences of bilingual schooling: The St. Lambert Project through grade five. *Journal of Educational Psychology, 65*(2), 141-159.
- Lindholm-Leary, K. J., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism, 13*(1), 43-60. doi: 10.1080/13670050902777546
- Steele, J. L., Slater, R. O., Li, J., Zamarro, G., Miller, T., & Bacon, M. (2018). Dual-language immersion education at scale: An analysis of program costs, mechanisms, and moderators. *Educational Evaluation and Policy Analysis, 40*(3), 420-445. doi: <https://doi.org/10.3102/0162373718779457>
- Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal, 54*(1), 282S-306S. doi: 10.3102/0002831216634463

- 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. *Alberta Journal of Educational Research, 49*(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. doi: 10.3102/0002831214545110
- Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English Learners: Variations by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis, 37*(4), 612-637. doi: 10.3102/0162373715573310
- Watzinger-Tharp, J., Rubio, F., & Tharp, D. S. (2018). Linguistic performance of dual language immersion students. *Foreign Language Annals, 51*(3), 575-595. doi: 10.1111/flan.12354