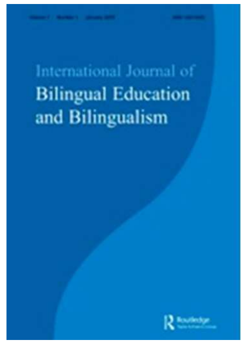


Academic Achievement of Utah Students in Dual Language Immersion

Journal:	*International Journal of Bilingual Education and Bilingualism*
Manuscript ID:	BEB-1465
Manuscript Type:	Paper
Keywords:	Dual Language Immersion, Academic Achievement, Two-Way Immersion, Content And Language Integrated Learning (CLIL), ELL, Bilingual Education

Abstract:

This article reports on a study that investigated achievement in math of third and forth grade dual language immersion (DLI) students in the State of Utah, building on research that has demonstrated the academic achievement of students who receive content instruction in a target language. The study included one-way programs in three languages (Chinese, French and Spanish) and two-way Spanish-English programs; and it relied on propensity matching to mitigate possible effects of school and student differences. In our third grade study, we compared students' math scores in relation to their ELA achievement to control for pre-existing differences between DLI and non-DLI students. DLI students who attained the same levels in ELA, and who received math instruction in the target language, performed at the same level as their non-DLI peers. For the forth grade study, we compared DLI students to a propensity-matched non-DLI group, and found that DLI students grew more in math than their counterparts not in DLI. These compelling results from Utah's natural experiment indicate that further research is warranted.

Academic Achievement of Utah Students in Dual Language Immersion

This article reports on a study that investigated achievement in math of third and forth grade dual language immersion (DLI) students in the State of Utah, building on research that has demonstrated the academic achievement of students who receive content instruction in a target language. The study included one-way programs in three languages (Chinese, French and Spanish) and two-way Spanish-English programs; and it relied on propensity matching to mitigate possible effects of school and student differences. In our third grade study, we compared students' math scores in relation to their ELA achievement to control for pre-existing differences between DLI and non-DLI students. DLI students who attained the same levels in ELA, and who received math instruction in the target language, performed at the same level as their non-DLI peers. For the forth grade study, we compared DLI students to a propensity-matched non-DLI group, and found that DLI students grew more in math than their counterparts not in DLI. These compelling results from Utah's natural experiment indicate that further research is warranted.

Immersion Education

Language education scholars generally consider the creation of French immersion programs in the 1960s in Canada as the beginning of immersion education in North America.[1] One of the first, and perhaps the best-known French immersion school, was opened in Quebec in 1965 in response to demands by parents of English-speaking children to provide their students the opportunity to become bilingual in French and English. At St. Lambert Elementary in a suburb of Montreal, English-speaking children learned subject matter content almost exclusively in French in the early grades, and then shifted toward equally distributed instruction in French and English starting in second or third grade; thus, the program provided what became to be known as "total" and "partial" immersion in the target language. Around the same time, and in response to a similar grassroots effort by parents of English-speaking children, Coral Way Elementary in Miami-Dade County established its English-Spanish bilingual program for both native English-speaking and native Spanish-speaking students (Fortune and Tedick, 2008).

These two programs in Canada and the U.S. each represent a particular immersion education program type: One-way programs that predominantly enroll one language group, e.g., English native speakers in French or Spanish immersion; and two-way programs for a linguistically heterogeneous group, with students "moving in two distinct directions toward the native language of their linguistically different peers" (Fortune and Tedick, 2008: 6). To achieve a mutually beneficial learning environment, two-way programs require a balanced distribution of students' language backgrounds, usually operationalized as at least a one-third to two-thirds minimum ratio.

Educational policy in the U.S. did not promote the benefits of bilingual education for monolingual English speakers to the same extent as Canada (Cohen and Swain, 1976). While immersion programs flourished in Canada in the early 1970s, only a few U.S. programs followed the Canadian model. Cohen and Swain (1976) point to a K-4 Spanish program in California and an elementary French immersion program in Maryland as the only examples of bilingual education that targeted majority group English-speaking students (49). Moreover, for minority students, language-focused education was often limited to pull-out remedial English as a Second Language (ESL) lessons, which often stigmatized students not only as linguistically, but also as academically deficient. Two-way immersion programs (TWI), which integrate language education of minority children and monolingual English speakers, started receiving a great deal of attention in the 1980s, although bilingual programs established in the 1960s and 1970s shared some of the TWI characteristics (Christian, 1996).

Since these early beginnings a significant number of programs situated in the broad rubric of immersion education[2] have been established in the U.S. Starting with just a few in the 1970s and 1980s, immersion programs more than doubled in the 1990s from 119 in 1991 to 278 in 1999. By 2011, 448 immersion programs were recorded by the Center for Applied Linguistics (CAL) 2011, but it should be emphasized that these numbers are likely to be much higher, since CAL relies exclusively on self-reporting. Moreover, immersion education has recently experienced significant growth across the U.S. For example, North Carolina's dual language/immersion programs have grown from nine in 2005 to over 90 today. In Utah, DLI programs more than doubled from 58 to 138 between 2011 and 2015.

Immersion Education in Utah

A major development in immersion education in the past five to seven years is the emergence of state-funded initiatives to establish immersion programs in public schools. In Utah, the State Legislature passed Senate Bill 41 in 2008 to fund the establishment of elementary school DLI programs in Chinese, French, and Spanish, starting in first grade or, for some, in kindergarten. Other states (e.g., Delaware and Georgia) have followed Utah's example.[3] As of 2015, and following the addition of Portuguese in 2013 and German in 2014, Utah had 138 elementary and secondary DLI schools with programs in five different languages. Russian immersion is expected to start in 2016.

In addition to one-way programs in five languages, Utah schools house two-way Spanish-English programs, where at least one-third of students must be native speakers of Spanish or, conversely, of English. Since two-way programs integrate students classified as EL's, native Spanish- and native English-speaking students, they create unique demands, but one-way and two-way programs share the main goals of immersion education: For all students to succeed academically; to attain high levels of proficiency in two languages; and, ultimately, to emerge from their immersion education as bilingual and biliterate, and equipped with cross cultural competencies and understanding (Howard *et al.*, 2003).

All Utah DLI programs that receive legislative funding are required to align with the 50/50 two-teacher model. In this model, students spend half of their school day with exclusive instruction in the target language by one teacher and the other half in English with another teacher. Utah has implemented uniform language curricula that are designed to promote literacy in both languages and to meet established proficiency targets for all four skills and for each grade level. Math and all other content areas, including social studies and science, are taught in the target language in first grade through third grade, while the English classroom focuses on English language arts (ELA), though it also provides vocabulary reinforcement for math, social studies and science. In fourth and fifth grade, math and social science shift to the English classroom, but students still engage in activities and practical application related to these subjects in the target language. In sixth grade, the target language classroom includes social science, while science and math are fully taught in English.[4] As a whole, the DLI curriculum aligns with the premise that language development is at the core of students' academic learning or, put differently, that "every content lesson must be a language lesson as well" (Met, 2008: 56).

Immersion Education Research

Research in the field of bilingual and immersion education traditionally has, and continues to focus on measurable outcomes such as cognitive skills, academic achievement, and language development, which are also of great interest to educators and parents (Walker & Tedick, 2000).

In response to studies that had found detrimental effects of bilingualism on children's intellectual functions (many of which are now considered flawed), Peal and Lambert (1962) administered intelligence tests to 10-year old children from six schools Montreal, who had been unambiguously classified as bilingual or monolingual (8). Balanced bilinguals with equal proficiency in French and English outperformed monolingual peers on both verbal and nonverbal measures of intelligence. Since then, many studies have demonstrated cognitive advantages for bilinguals (Ben-Zeev, 1977; Bialystok *et al.*, 2009; Cummins, 1977; Lazaruk, 2007). To achieve a clearer understanding of the wide range of cognitive outcomes associated with bilingualism, Adesope, Lavin, Thompson and Ungerleider (2010) conducted a meta-analysis of 63 studies that examined measures such as attentional control, working memory, metalinguistic awareness, abstract symbolic representation skills, and problem solving (212). Although the meta-analysis uncovered significant variability across studies, it established an overall positive effect of bilingualism on cognitive outcomes, in particular metalinguistic awareness, attentional control, working memory and abstract and symbolic representation skills (Adesope *et al.*, 2010: 229).

A number of studies, including several that were conducted on a large scale and longitudinally, have demonstrated the beneficial effects of content instruction in the target language for academic achievement across subjects and in two languages (Bournot-Trites & Reader, 2001; Collier & Thomas, 2004; Lambert & Tucker, 1972; Swain & Lapkin, 1982; Thomas & Collier, 1997, 2002; Thomas *et al.*, 1993; Turnbull *et al.,* 1998). Moreover, programs with formal schooling of minority students through their native language have been found to be most effective for ensuring academic success of this population (Lindholm & Aclan, 1991; Lindholm-Leary & Borsato, 2005; Lindholm-Leary & Howard, 2008; Thomas & Collier, 1997a, 2002).

The most significant findings to date on the academic achievement of students in dual language programs come from a series of comprehensive cross-site studies across a variety of school program types and services for English learners (ELs). Building on research they had started in 1985, Thomas and Collier (2002) collected data from five school districts representing northern and southern regions in the U.S. and urban and rural areas from 1996 to 2001; languages included English, French, and Spanish. Of the eight program types that were included in the study, only dual language programs were found to be effective long-term, as measured by students reaching the 50th percentile in both languages in all subjects and maintaining this level of achievement through the end of their schooling (7). A follow-up study (Collier & Thomas, 2004) that focused on dual language programs in two U.S. school districts confirmed that both one-way and two-way[5] dual language programs fully close the achievement gap for all students who participate in this educational model, ranging from ELs with little or no proficiency in English to heritage students who are proficient in English, but have lost their heritage language (French or Spanish).

One-way DLI programs that educate students in both their native (majority) language of English and a foreign language also have to demonstrate that schooling in two languages does not inhibit academic achievement. A study of eight French, Japanese and Spanish immersion programs (one of them two-way Spanish) showed that in grades one, two and three immersion students performed as well or better on ELA and math achievement tests as their comparable non-immersion peers (Thomas *et al*. 1993: 178). The programs were described as "partial immersion," or what we refer to as a 50/50 model today. Mathematics was taught in the target language, confirming that students successfully learn content in and through a second language.

Immersion education research also focuses explicitly on the interaction of academic learning with language of instruction or testing in L1 or L2 (Bournot-Trites & Reeder, 2001; Turnbull *et al.,* 2001). Bournot-Trites and Reeder (2001) compared a cohort of Canadian French immersion students who received 20% of their mathematics instruction in English and 80% in French with another cohort that was instructed 50% in English and 50% in French. The native English-speaking students who received 80% of their mathematics instruction in French performed significantly better than the 50/50 cohort. However, the authors caution that factors other than language of instruction could have contributed to the differences in students' test performances. These include effect of the instructional material (which differed across the two classrooms), involvement of parents and motivation of students.

Lindholm and Aclan (1991) tested the relationship of achievement and bilingual proficiency with 249 students in grades one through four in two bilingual/immersion schools in Northern California. Of these students, about two thirds were considered native Spanish, and a little more than one third (36%) native English speakers (103). Students were classified as possessing low, medium, or high bilingual proficiency based on a score assigned by teachers through observations of students' oral language use in various situations. To measure their academic achievement, students were given four tests: English and Spanish math, and English and Spanish reading. The results showed that students who had been assigned a high level of bilingual proficiency outperformed the low and the mid groups on all four measures. The study also found that students at all three levels of proficiency performed average to high-average on their English mathematics tests, even though they had received math instruction exclusively in Spanish until fourth grade and no English reading instruction until third grade. The authors concluded, "that students at all proficiency levels were developing mathematical concepts and skills because their knowledge of Spanish was excellent" (110).

Turnbull, Lapkin and Hart (2001) compared math and literacy achievement of French immersion and non-immersion students Ontario, and also analyzed the testing data for relationships between exposure to English instruction and test scores, and English exposure to language of testing (French or English). They found that native English-speaking students in third grade French immersion temporarily lagged behind non-immersion peers on tests of English literacy, especially in programs where formal English instruction was delayed until later grades. However, these students caught up with their non-immersion peers by the sixth grade. Math performance of immersion students did not differ from non-immersion students in grade 3 or grade 6, and the language of testing did not have an effect on the resulting math scores.

DLI research, whether focused on ELs and heritage students (e.g. Collier & Thomas, 2004; Lindholm-Leary, K. & Borsato, 2005; Thomas & Collier, 1997a, 2002) or native English speakers in one-way foreign language immersion programs, has provided strong evidence that schooling in two languages supports students academically and linguistically. The recent growth of DLI programs across the U.S. affords opportunities to investigate whether prior findings hold true for large sample sizes in different DLI program types and for different immersion languages, and when variables that could mitigate effects are rigorously controlled.

The Utah Study

The current study measured the academic achievement in math of students in third grade and fourth grade in Utah's DLI programs. It builds on previous research, which has demonstrated that students succeed academically when content is taught fully or predominantly in the target language (e.g. Turnbull *et al.,* 2001; Bournot-Trites & Reeder, 2001; Collier & Thomas, 2004;

Thomas & Collier, 2012). (Recall that in Utah, students learn math exclusively in the target language in first grade through third grade.) It also expands on the scope of prior research in two ways: The study includes one-way programs for native English speakers in three languages (Chinese, French, and Spanish) as well as two-way Spanis-English programs; and it relies on propensity matching of DLI and non-DLI schools, and DLI and non-DLI students to mitigate effects of pre-existing differences at the school and student levels. To measure academic performance of students in one-way and two-way DLI programs, which employ the same instructional model (50/50 two-teacher), we formulated the following three research questions:

1. How do third grade students who are and are not in Utah DLI programs perform in math, relative to their performance in English Language Arts (ELA)?
2. How much growth in math is observed in fourth grade students in Utah DLI programs compared to similar students who are not in DLI?
3. What effect does DLI target language or program type (one-way or two way) have on third and fourth grade students' academic performance in math?

Methods

We used rigorous statistical methods to ensure an equitable comparison of students participating in DLI with students not in DLI. The methods described in this section meet the What Works Clearinghouse Evidence Standards with Reservations designation (Clearinghouse, 2009). This designation reflects the highest possible criteria for making causal inferences from data collected in non-randomized studies. As such, results showing differences between students who were in DLI and who were not in DLI can be cautiously attributed to DLI participation rather than to pre-existing differences between the two groups.

Two samples were used in this research: a third grade sample, which consisted of third grade students who attended DLI schools in the 2011-12 school year, and a fourth grade sample, which consisted of fourth grade students who were in DLI in the 2012-13 school year, and their propensity matched peers. In both samples, students from DLI programs were identified from within 26 schools. These schools included six schools with Chinese programs, five schools with French programs, six schools with one-way Spanish programs and nine schools with two-way Spanish programs.

The third grade sample was comprised of all students enrolled in third grade in the 2011-12 school year who met three inclusion criteria: 1) students were enrolled in a Utah public school that offered Utah model DLI programs to third grade students; 2) students did not change schools between first and third grades; and 3) students received third grade scores from Utah's standardized English Language Arts (ELA) test and Utah's standardized math tests.

A total of 2,524 students met the inclusion criteria. A majority of these students identified as either White (73%) or Hispanic (20%) with 37% of the students qualifying for free or reduced prices lunch (FRL) and 7% qualifying for English Learner (EL) services in the third grade. The demographic characteristics of this student sample are representative of student demographics within the state. Of the sample, 47.3% of the students (N=1,195) were enrolled in DLI programs and 52.7% of the students (N=1,329) were not.

The fourth grade sample comprised students who were enrolled in a DLI program as fourth graders during the 2012-13 school year, and who had third grade and fourth grade ELA and math scores on Utah's standardized achievement tests. The fourth grade students were paired with propensity matched students who had similar demographics and similar test scores.

Overall, 2,287 students were in the fourth grade sample, including 1,148 DLI students and 1,139 non-DLI propensity matched students. Consistent with demographics in the state, a majority of the student in the fourth grade sample identified as either White (73%) or Hispanic (21%), with 35% qualifying for free or reduced priced lunch and 7% qualifying for EL services in the fourth grade.

Students' ELA and math scores on Utah's standardized Criterion Reference Tests (CRTs) served as the academic outcomes in this study. Until recently, CRTs were administered each spring; the CRTs have been replaced by a new set of standardized tests aligned to a new set of core standards implemented in Utah. Scaled scores on the CRTs ranged from 130 to 190 points, with a mean of 160 and a standard deviation of 10. CRT scores were used to analyze math scores of third grade students and create Student Growth Percentages (SGPs) to analyze math learning during the fourth grade. SGPs were created by binning students into 100 quantiles based on their third grade test results, and calculating the percentile for each student within each bin using fourth grade test results. For example, if a student was in the 45th percentile in math in the third grade, that student's fourth grade math scores were compared to all other students in the state who were also in the 45th percentile in the third grade. Each student's SGP could range from the 1st to the 100th percentile.

To find a sample of non-DLI students for comparison with DLI students in the fourth grade study, we used a two-step process for propensity score matching. In the first step, we included school size, ethnicity, mobility, parental income, and performance of third and fourth grade students prior to the study to match DLI schools with non-DLI schools. After we had matched each DLI school with a non-DLI school, we matched students from within each DLI school with peers from within the matched non-DLI school using gender, ethnicity, qualification for free and reduced priced lunch, qualification for special education and EL services, and third grade CRT scores. As the focus of the analysis was on math learning, an exact match was required, by decile, for math scores. In other words, a student who scored in the 80^{th}-89^{th} decile in math in the third grade was matched to another student who scored in the 80^{th} -89^{th} decile on the math test in the third grade. We used caliper matching and required a maximum difference of .1 on the combination of all propensity matching variable in order to match students. Using this method, we obtained suitable matches for 1,139 of the 1,148 DLI students.

Propensity matching was assessed by comparing the absolute standardized bias (ASB) estimates of DLI students and non-DLI students prior to matching and after matching. Standardized bias is a measure that is not influenced by sample size and, thus, can be used to compare matched and unmatched samples (Stuart 2010). The use of ASB removed the sign for direction from the standardized bias results allowing for a direct comparison of the magnitude of bias and, importantly, allowed for meaningful averaging of magnitudes across covariates.

To answer the research questions related to third grade math scores, data were analyzed with multi-level regression. To answer the research questions related to fourth grade test scores, multiple regression was used to compare DLI students with their propensity matched peers, and multi-level multiple regression was used to detect possible effects of target language (i.e., Chinese, French, or Spanish) and program type (i.e., one-way or two-way immersion) on math learning. Multi-level regressions for both third and fourth grade analyses included student-level and school-level models, which were estimated with HLM software (Raudenbush *et al.,* 2004). First, null models (i.e., models with no predictors) were fit to gain a baseline estimate of variance at the student-level and to determine what proportion of the student-level variance could be accounted for at the school level. Next, means-as-outcomes models with school-level predictors (i.e., target language and program type) were run to estimate the proportion of variance in school-level averages that could be accounted for by target language and program type. Finally, full models were run. Appendix A shows the specifications of the null, means-as-outcomes, and full models for both the third and fourth grade study.

Third Grade Study

According to the Utah DLI program requirements, math is taught in the target language through the third grade year; ELA is, of course, taught in English. This policy allowed for a within-subjects design wherein student learning of ELA was used to predict student learning of math, which was taught in the target language. Using ELA scores and demographic variables to predict math scores allowed us to analyze whether DLI and non-DLI students who performed similarly on the ELA test performed differently in math, after controlling for demographic and academic characteristics. A significant positive DLI coefficient would have indicated that students in the DLI program performed better than students who were not in DLI and a significant negative DLI coefficient would have indicated that students in the DLI program performed worse than students who were not in DLI. At the second level, or school level, the target language and the program type were used to predict average math scores for each school. Significant coefficients for target language or program type would have indicated that students with similar demographics and ELA scores who were in schools with different target languages or different program types performed differently in math.

Fourth Grade Study

The fourth grade study included two analyses. The first analysis was a multiple regression that controlled student demographic characteristics (i.e., gender, qualification for free and reduced price lunch, special education, EL services, and race/ethnicity) and tested for differences in math growth between DLI students and their propensity-matched peers. The second analysis was a multi-level regression that included only the subsample of 1,148 DLI students. This regression controlled student demographic characteristics and tested for differences in math growth among the students learning in the different target languages and between students in one-and-two way programs.

Results

Third Grade Study

Results from null model (see Appendix B for all multi-level model results from third grade study) showed that approximately 10% of the variance in student math scores could be accounted for by which school the student attended. The proportion of variance accounted for was significant (X^2=322.37; p<.001), which indicated that proposed multi-level analysis was appropriate for the data. The means-as-outcomes model showed that the school-level predictors (i.e., target language and program type) accounted for about 10% of the variance among schools with program type approaching significance ($T_{(22)}$=2.055, p=.052). The full model showed acceptable reliability indices (reliability of intercept = .791; reliability of DLI slope = .679) and good model fit (R^2=.51). These statistics suggested that the model was well-specified and that the results were reliable. As shown in Appendix B, results from key DLI predictors (i.e., student participation in DLI, target language, and program type) were all non-significant. These non-significant results indicated that students who participated in DLI performed the same in math as students who did not participate in DLI when all other factors, including ELA scores, were held constant. Results also showed that, student math performance was similar across schools with different target languages (i.e., Chinese, French, and Spanish) and with different program types (i.e., one-way and two-way) when all other factors were held constant.

Fourth grade study

Average absolute standardized bias statistics were calculated for DLI and non-DLI students before and after the matching process. Bias over 10 is considered problematic. As shown in Table 1, bias associated with each of the covariates was reduced through the propensity score matching process and bias for 7 of 9 covariates was reduced from values of greater than 10 to values less than 10. Importantly, the overall bias was reduced from a value of 16.706 to a value of 6.058.

Table 1 near here

The multiple regression model predicting growth in fourth grade math (Math SGP) was significant, $F_{(7, 2279)} = 7.079$, $p<.001$ but only accounted for a small proportion of the variance in student growth ($R^2 = .02$). As shown in Table 2, participation in DLI was a significant predictor of student growth in math. The coefficient of 2.83 indicates that students who participated in DLI during fourth grade scored almost three percentile points higher on the standardized math test than propensity-matched students.

Table 2 near here

Only DLI students were included in the multi-level analyses of fourth grade growth. Results from the null model (see Appendix C for all multi-level model results from fourth grade study) showed that for the sample DLI students, approximately 18% of the variance in math growth could be accounted for by which school the students attended. The proportion was significant (X^2=262.26; p<.001), indicating that the proposed multi-level analysis was appropriate for the data. The means-as-outcomes model results showed that the school-level predictors (i.e., target language, and program type) accounted for about 22% of the variance among schools. Although this proportion of variance accounted for was significant (X^2=72.72; p<.001), none of the predictor variables showed significant independent relationships with math growth.

Results from the full model showed acceptable reliability indices for the intercept (.823) but reliability estimates for programmatic effects on English Learners (.148) and on students from Hispanic backgrounds (.073) indicated poor reliability. Thus the model was respecified to use programmatic variables to predict school averages only. The respecified model showed strong reliability (.878) but poor model fit (R^2=.02) indicating that the predictors explained little variance in student growth scores. Results from the respecified model are included in Appendix C. As shown in the results, neither target language nor program type significantly predicted student growth in math. Figure 1 shows that even after controlling for differences in demographics, average math percentiles for DLI students in schools with different target languages and program types were quite different. However, because of large amounts of variance among the schools, these differences could not be considered statistically significant.

Figure 1 near here

Discussion

We could not make direct comparisons between the math performance of third grade DLI and non-DLI students because no prior academic achievement scores were available. Instead, we considered math performance of DLI and non-DLI students in relation to ELA scores. We found that math scores were neither relatively lower nor higher than ELA scores for students who were and were not enrolled in DLI in third grade. In other words, for DLI and non-DLI students who attained the same achievement levels in ELA, DLI students, who received math instruction in the target language from first grade through third grade, performed the same as non-DLI students, who received math instruction completely in English. These results demonstrate that learning math in the target language did not impede the DLI students' academic achievement.

Once students were in the fourth grade, we were able to use students' third grade scores and their demographic characteristics to form a non-DLI control group. Each student in the control group had a third grade math score that matched the exact same decile as the third grade math score of a student in the DLI group. Students in the DLI and non-DLI control groups were also matched on ELA scores and demographic characteristics, although exact matches for those variables were not required. We found that students who were in DLI in the fourth grade grew more in math than did matched students who were not in DLI in the fourth grade. Because DLI and propensity-matched non-DLI students had similar math scores in the third grade, these results demonstrate that students in DLI demonstrated more growth in fourth grade math than students who were not in DLI.

There were substantial differences between both third grade CRT scores and fourth grade SPGs at the school level. However, neither target language nor program type could reliably predict these differences because too much variance existed within the groups of schools. For example, the six schools that taught Chinese as the target language ranged in math SGP averages from the 49th to 70th percentiles.

Post hoc tests on the multiple regression conducted to detect DLI effects on fourth grade students showed no interaction between a student qualifying for EL services and student participation in DLI. We attempted to analyze differential effects of target language and program type for both Hispanic students and students qualifying for EL services, but the differences in outcomes across target languages were not sufficient to conduct those analyses. The relatively homogeneous performance of Hispanic and EL students across different target languages and program types suggests there were no program level effects on these students. From our findings, we were able to draw the overall conclusion that, on average, students in Utah's DLI programs grew more in math in the fourth grade than similar students who were not in DLI. This effect was observed across different student types, three target languages, and two program types (one-way and two-way). Our study therefore supports findings consistent across a wide range, and a long tradition of immersion research: That content education in two languages, or exclusively in the target language has the potential to enhance academic outcomes for all students. We were able to demonstrate that beneficial effects hold for three languages, two different program types, and students with a range of demographic characteristics.

Limitations

As is true for all quasi-experimental studies, causal inferences from the analyses presented in this paper should be made cautiously. We used the most rigorous methods available to make equitable comparisons between DLI and non-DLI students, but demographic and academic covariates used in the propensity matching analysis could not account for all individual differences between DLI and non-DLI students or between parents of DLI and non-DLI students. Non-measured factors such as parental support may have threatened the internal validity of our conclusions.

The Utah DLI students used in our sample represent the first cohort to be enrolled in Utah's DLI program. Conceivably, DLI curriculum, available pedagogical supports, and the nature of the DLI students may have changed as the program developed and expanded. Finally, although our analyses allowed us to describe the academic performance of this sample of DLI students, and compare their performance to non-DLI students, we were unable to consider pedagogical differences between math teaching in DLI and non-DLI classroom. For this study, we also did not address learning of the target language expected to co-occur with academic content learning in the DLI classrooms.

Ongoing and Future Research

This paper reports on our recently completed research of Utah DLI and non-DLI students' academic achievement in third and forth grade math. The results robustly demonstrate the benefits of dual language immersion for academic achievement, specifically for students' performance on standardized math tests. For our ongoing and future research we are pursuing two trajectories. One trajectory will build on preliminary analyses of students' performance in fourth grade and fifth grade science, which will also provide an occasion to investigate the impact of curricular structures. While math is taught as a stand-alone subject, and exclusively in the target language through third grade, science instruction is joined with social studies in Utah's so-called *Interconnections* curriculum in first grade through third grade. The *Interconnections* curriculum is also taught in the target language, and reinforced in English, but instructional times are considerably less than math. In fourth grade, science is taught separately and is also subject to statewide assessment. Our initial findings indicate that DLI students score lower than non-DLI students in fourth grade science, but then grow at a greater rate than their non-DLI peers in fifth grade. To gain a better understanding of possible differences in math and science performance, we will replicate the pilot study, but we will also examine the potential impact of curricular structures and instructional times.

The second research avenue will focus on the relationship between academic achievement and language proficiency. Previous research has found that students with high levels of bilingual proficiency, typically in English and Spanish, also perform well on academic achievement tests (e.g. Lindholm and Aclan, 1991; Alanis, 2010; Collier, 2002; 2012). In Utah, DLI students participate in a combination of formative and summative assessments to determine whether they reach the proficiency targets set at each grade level for each DLI target language. We will be able to draw on data that have been collected since 2014[6] to investigate whether the language proficiency of DLI students in both one- and two-way programs, and across three languages (Chinese, French, and Spanish), interacts with performance on academic achievement tests. Together with our current math study, our research of outcomes in science, and the interaction of academic performance and language proficiency will establish a comprehensive account of DLI as an effective educational model that should be available to all students.

Notes

[1] Immersion education is, of course, also considered a form of bilingual education. The term bilingual education continues to be used in Canada, Europe, and South America, while immersion education has taken hold in the U.S.

[2] The terminology to classify immersion program types continues to evolve. For example, total or full immersion programs are now often referred to as 90/10, and partial immersion has largely been replaced by 50/50.

[3] Delaware's *World Language Expansion Initiative* started four programs in 2012, with an annual investment of $1.9 million; currently, some 2300 students are enrolled in Mandarin Chinese and Spanish immersion programs (Delaware.gov, 2015). In 2013, the Georgia Department of Education awarded grants for DLI to six elementary schools (Georgia Department of Education, 2013).

[4] See the Utah DLI homepage for more detailed information on the DLI instructional model: http://utahdli.org/instructionalmodel.html

[5] In the context of their dual Thomas and Collier (2004) define one-way programs as those that serve students from one language group, often heritage speakers, whose linguistic backgrounds may vary widely. In two-way programs, students who speak English natively join their "bilingual and ELL peers in an integrated bilingual classroom (3)." They also point out that two-way classrooms should be open to all students, including those who have lost their heritage language and only speak English.

[6] Utah uses the ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL), beginning in third grade with the interpersonal mode of listening/speaking. Test takers participate in computer-based role-play scenarios that simulate real world tasks. In 2014, 3000 students took the test.

References

Adesope, O.O., Lavin, T., Thompson, T. and Ungerleider, C. (2010) A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research* 802, 207-245.

Alanis, I. (2000) A Texas Two-Way Bilingual Program: Its Effects on Linguistic and Academic Achievement. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education* 24(3), 225-248.

Bialystok, E., Craik, F.I., Green, D.I. and Gollan, T.H. (2009) Bilingual Minds. *Psychological Science in the Public Interest* 10(3), 89-129.

Bournot-Trites, M. and Reeder, K. (2001) Interdependence revisited: Mathematics achievement in an intensified French immersion program. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 58, 27–43.

Clearinghouse, What Works (2008). *WWC procedures and standards handbook.* Washington, DC.

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

Cummins, J. (1977) Cognitive Factors Associated with the Attainment of Intermediate Levels of Bilingual Skills. *The Modern Language Journal* 61(1-2), 3-12.

Delaware.gov. (2015) World Language Immersion Program Expands for 2015-2016 School Year. Online at http://news.delaware.gov/2015/01/07/world-language-immersion-program-expands-for-2015-16-school-year/

Fortune Williams, T. and Tedick, D.J., Eds (2008) *Pathways to Multilingualism. Evolving Perspectives on Immersion Education*. Clevedon: Multilingual Matters.

Howard, E.R., Olague, N. and Rogers, D. (2003) *The Dual Language Program Planner: A Guide for Designing and Implementing Dual Language Programs*. Center for Research on Education, Diversity and Excellence. Santa Cruz, CA.

Lazaruk, W. (2007) Linguistic, Academic, and Cognitive Benefits of French Immersion. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 63(5), 605-628.

Lindholm, K.J. and Aclan, Z. (1991) Bilingual Proficiency as a Bridge to Academic Achievement: Results from Bilingual/Immersion Programs. *The Journal of Education* 173(2), 99-113.

Lindholm-Leary, K. and Borsato, G. (2005) Hispanic High Schoolers and Mathematics: Follow-Up of Students Who Had Participated in Two-Way Bilingual Elementary Programs. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education*, 29(2), 641-652.

Lindholm-Leary, K. and Howard, E.R. (2008) Language Development and Academic Achievement in Two-Way Immersion Programs. In T.W. Fortune and D. J. Tedick (eds) *Pathways to Multilingualism. Evolving Perspectives on Immersion Education* (pp. 177-200). Clevedon, UK: Multilingual Matters.

Met, M. (2008) Paying Attention to Language: Literacy, Language and Academic Achievement. In T.W. Fortune and D. J. Tedick (eds) *Pathways to Multilingualism. Evolving Perspectives on Immersion Education* (pp. 49-70). Clevedon, UK: Multilingual Matters.

Peal, E. and Lambert, W.E. (1962) The relation of bilingualism to intelligence. *Psychological Monographs: general and applied* 76(27), 1-23.

Raudenbush, S.W., Bryk, A.S. and Congdon, R. (2004) HLM 6 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Stuart, E. A. (2010) Matching methods for causal inference: a review and a look forward. *Statistical Science* 25(1), 1-21.

Swain, M. (1996) Discovering successful second language teaching strategies and practices: From program evaluation to classroom experimentation. *Journal of Multilingual and Multicultural Development* 17, 89–104.

Swain, M. and Lapkin, S. (1982) *Evaluating bilingual education: A Canadian case* study. Clevedon, England: Multilingual Matters.

Thomas, W.P. and Collier, V. (1997a) School effectiveness for language minority students. Online at http://crede.berkeley.edu/research/llaa/1.1_final.html

Thomas, W.P. and Collier, V. (2002) A national study of school effectiveness for language minority students' long-term academic achievement. Online at http://www.crede.ucsc.edu/research/llaa/1.1_final.html

Thomas, W.P. and Collier, V. (2012) *Dual Language Education For a Transformed World*. Dual Language Education of New Mexico-Fuente Press: Albuquerque, NM.

Turnbull, M., Lapkin, S. and Hart, D. (1998) Time on task and immersion graduates' French proficiency. In Lapkin, S. *(ed)* French second language education in Canada: Empirical studies (pp.31-55). Toronto: University of Toronto Press.

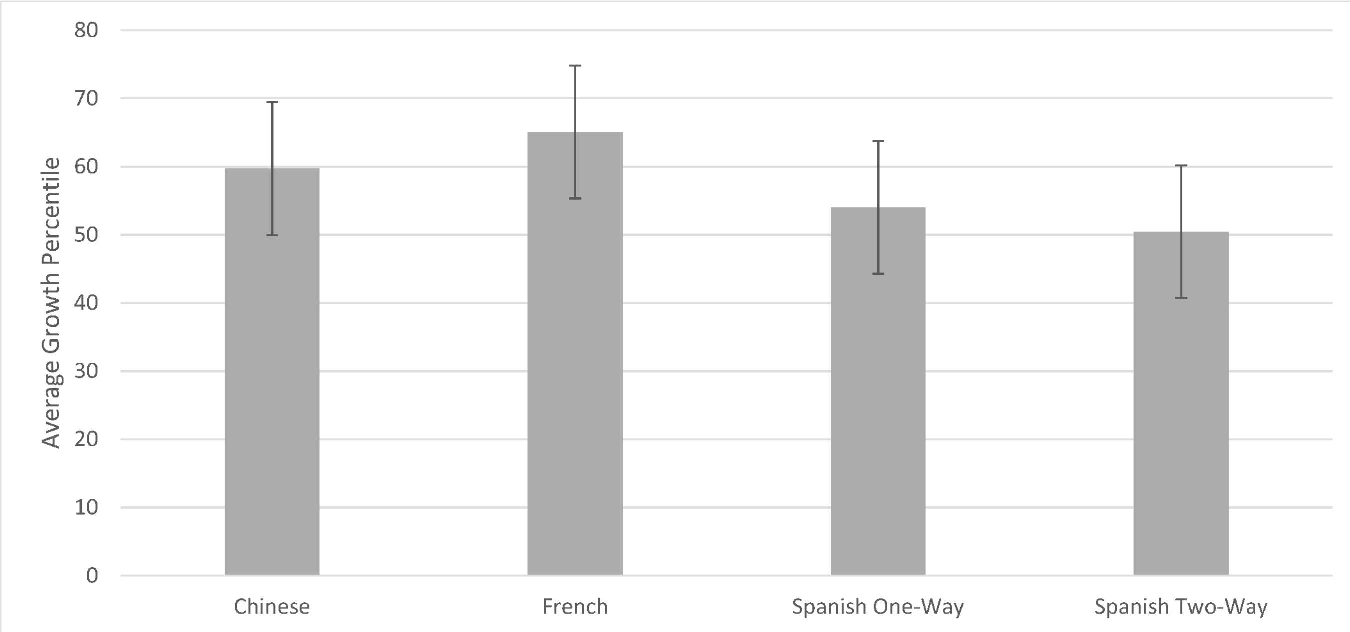
Turnbull, M., Lapkin, S. and Hart, D. (2001) Grade 3 immersion students' performance in literacy and mathematics: province-wide results from Ontario 1998-99. *Canadian Modern Language Review/La Revue canadienne des langues vivantes* 58(1): 9-26.

Table 1: Absolute Standard Bias Before and After Matching

Variable	Before Match	After Match
Mobile	11.401	5.372
Chronically Absent	19.51	3.032
Female	11.118	8.122
White	12.684	3.985
Hispanic	18.12	4.719
FRL	18.257	12.702
EL	4.718	5.605
Special Ed	26.328	5.204
ELA CRT	26.167	11.837
Math CRT (decile)	18.759	0
Average	**16.706**	**6.058**

Table 2. Multiple Regression Results Predicting Fourth Grade Growth in Math

Predictor	Unstandardized Coefficient	Standardized Coefficient (Beta)	P value
(Constant)	55.13		0.00
DLI student	**2.83**	**0.05**	**0.02**
female	-4.06	-0.07	0.00
Low Income	-1.61	-0.03	0.26
English Learner	-7.71	-0.07	0.00
Special Education	-1.17	-0.01	0.61
Hispanic	-3.16	-0.05	0.08


80
70
60
50
40
30
20
10
0
Average Growth Percentile
Chinese
French
Spanish One-Way
Spanish Two-Way