INTERNATIONAL JOURNAL OF BILINGUAL EDUCATION AND BILINGUALISM, 2016
http://dx.doi.org/10.1080/13670050.2016.1214675

Routledge
Taylor & Francis Group

Academic achievement of students in dual language immersion

Johanna Watzinger-Tharp[a], Kristin Swenson[b] and Zachary Mayne[c]

[a]Department of Linguistics, University of Utah, Salt Lake City, UT, USA; [b]Utah Department of Human Services, University of Utah, Salt Lake City, UT, USA; [c]Utah Education Policy Center, University of Utah, Salt Lake City, UT, USA

ABSTRACT

This article reports on a study that investigated achievement in math of third and fourth grade dual language immersion (DLI) students, building on research that has demonstrated the academic achievement of students who receive content instruction predominantly in the target language. Our study expands the scope and methodology of prior research by including one-way programs in three languages (Chinese, French and Spanish) and two-way Spanish-English programs; and by relying 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 English Language Arts (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 a target language, performed at the same level as their non-DLI peers in third grade math tests given in English. For the fourth grade study, we compared DLI students to a propensity-matched non-DLI group. DLI students grew more in math than their counterparts not in DLI. The results from this natural experiment indicate that students in a DLI program that has been implemented state-wide were able to succeed academically in math.

ARTICLE HISTORY
Received 26 August 2015
Accepted 13 July 2016

KEYWORDS
Dual language immersion; academic achievement; content and language integrated learning; bilingual education; content-based instruction; immersion education

Introduction

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] St. Lambert Elementary, 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, 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. 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 US each represent a particular immersion education program type: One-way programs that serve students who speak the majority language (e.g. English native speakers moving toward proficiency in Chinese or German), 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

CONTACT Johanna Watzinger-Tharp j.tharp@utah.edu

a mutually beneficial learning environment, two-way programs require a balanced distribution of students' language backgrounds, usually operationalized as roughly a one-third to two-thirds minimum ratio.

Educational policy in the US 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 US 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 stigmatize students not only as linguistically, but also as academically deficient. In contrast, two-way immersion programs, which started to grow in the 1990s, explicitly aim at the integration of language minority students with majority English speakers, while also aligning with the goals of high levels of proficiency and achievement in academic areas (Christian 1996).

Since these early beginnings, a significant number of programs that fall under the broad rubric of immersion education[2] have been established in the US. 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 US. For example, North Carolina's dual language/immersion programs have grown from nine to over 90 since 2005. In Utah, DLI programs more than doubled from 58 to 138 between 2011 and 2015. Indeed, a 'majority of states in the United States reported that, during the 2012–2013 school year, districts in their state were implementing at least one dual language program' (Boyle et al. 2016, x).

Background

Research related to bilingualism, bilingual and immersion education addresses a wide range of interconnected issues, but of particular interest for the present study are theoretical and experimental studies on the cognitive benefits of bilingualism, and studies that examine the effects of dual language immersion education on academic achievement, in particular those conducted on a large scale.

Bilingual and immersion education are grounded in a significant body of research that has demonstrated cognitive advantages of bilinguals. After a number of studies had issued warnings about potentially detrimental effects of bilingualism on students' intellectual functioning, Peal and Lambert's (1962) seminal study showed that balanced bilinguals with equal proficiency in two languages were able to outperform monolingual peers on various verbal and nonverbal tests of intelligence. Since that time, many studies have demonstrated cognitive advantages of bilinguals, which, in turn, may enhance academic performance, in particular in math and reading (Bialystok 1991; Bialystok and Majumder 1998; Bialystok et al. 2009; Bialystok and Craik 2010; Esposito and Baker-Ward 2013; Foy and Mann 2014). From their review of literature on the relationship between bilingualism and cognitive processes, Bialystok et al. (2009) conclude that bilingual speakers, by necessity, develop an enhanced ability to inhibit or suppress material that is not relevant to a task at hand, and to select information that is. Together with other functions such as monitoring, the processes of inhibition and selection contribute to bilinguals' advantage in executive control.

To be able to estimate the effects of cognitive correlates associated with bilingualism, Adesope et al. (2010) conducted a meta-analysis of 63 studies that had reported on measures of cognitive benefits such as attentional control, working memory, metalinguistic awareness, and problem solving (212). Though effect sizes varied significantly, this systematic review of a wide range of studies confirmed the positive association of bilingualism with cognitive benefits, in particular metalinguistic and metacognitive awareness, attentional control, and problem solving.

The potential of bilingualism to impact students' academic performance positively is manifested in the focus of observational immersion education research on academic achievement (Swain and Lapkin 1982; Thomas and Collier 1997, 2002; Collier and Thomas 2004). For minority language students, programs with formal schooling in their native language have been found to be most effective for ensuring long-term academic success (Lindholm and Aclan 1991; Thomas and Collier 1997, 2002; Alanis 2000; Lindholm-Leary and Howard 2008; Lindholm-Leary and Block 2010).

In the US, research on the effects and effectiveness of dual language immersion education, as measured by academic outcomes, is often situated in the context of the persistent English Learner (EL) academic achievement gap in reading and math. Some of the most significant findings on academic achievement of ELs come from a series of comprehensive cross-site studies which examined a variety of schooling types, including bilingual and immersion. Building on research that had started in 1985, Thomas and Collier (2002) collected data on language minority student achievement from five school districts representing northern and southern regions, and urban and rural areas in the US between 1996 and 2001. Languages included English, French, and Spanish. Their study included a total of eight major program types: Four developmental bilingual or dual language immersion programs (90/10 and 50/50, and one-way and two-way[3]); two transitional bilingual (90/10 and 50/50); ESL through academic content; and English mainstream. Two-way bilingual immersion (90/10 and 50/50) and 90/10 one-way developmental bilingual education were found to be effective long-term, as measured by language minority students reaching and maintaining the 50th percentile in both languages in all subjects (7). In contrast, transitional, ESL and English mainstream programs were shown not to be effective in closing the achievement gap for ELs. Based on their consistent results, Collier and Thomas (2004) conclude that dual language schooling, both one- and two-way, can close 70–100% of the achievement gap for ELs by grade 5 (15).

A number of studies in Canadian and US contexts have demonstrated that immersion students perform at similar levels, or outperform their non-immersion peers in tested academic content areas, most commonly English literacy and math (Turnbull, Lapkin, and Hart 2001; Lazaruk 2007; Lindholm-Leary and Hernandez 2011; Marian, Shook, and Schroeder 2013; Padilla et al. 2013; Steele et al. 2015). In their province-wide study in Ontario, Turnbull and his colleagues (2001) found that third grade French immersion students, enrolled in a range of program types, performed as well on reading, writing and math tests as English students. Immersion students who had not received any English instruction by the time they took the tests slightly lagged behind in literacy, but for all others, even if they had only started English instruction in third grade, there was no evidence that the level of English instruction had affected their test performance. In mathematics, students did as well as their non-immersion peers, even when they had received no instruction in English. Furthermore, it did not matter whether students took the math tests in English, their native language, but not the language of instruction, or whether they took the tests in French, their second language and the language of instruction.

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 equally in English and French (50% each). The native English-speaking students who received 80% of their mathematics instruction in French performed significantly better than the 50/50 cohort.

The existing body of research suggests that students in dual language immersion programs are able to achieve academically, and perhaps even outperform their peers in monolingual programs. Lindholm-Leary and Genesee (2014) conclude from their meta-analysis of international research on student outcomes that dual language immersion education, in one-way, two-way and indigenous programs, clearly confers academic and linguistic benefits. The benefits have been found for majority and minority language speakers, for students from a range of ethnic backgrounds and for students who have special educational needs (Lindholm-Leary and Genesee 2014, 175).

However, as Steele et al. (2015) point out, the methodology of studies on the effects of dual language immersion education is often not sufficiently robust to control for pre-existing conditions or selection bias, which limits the ability to identify causal program effects or to generalize beyond a study's particular context (9). In their large-scale longitudinal study of students in grades eight through kindergarten in Portland Public Schools, Steele and her colleagues capitalized on lottery system data to control for unobserved differences between students randomly assigned to dual language immersion in kindergarten and students not selected for the program (4). The dual language immersion cohort in this study outperformed the non-immersion cohort in reading, or English Language Arts (ELA), with their advantage equivalent to about seven months of instruction in grade five, and roughly one academic year in grade eight (21). The positive effects on reading were observed for both native speakers of English and students classified as ELs, and across the different programs and languages (predominantly Mandarin and Japanese one-way, and Spanish two-way). Positive immersion effects were not found on math or science. However, it is worth noting that target language instruction in these content areas, fully or partially until grade five, also did not disadvantage students (Steele et al. 2015, 25).

The Portland study, as one of few conducted with a large randomized sample and multiple languages, provides particularly strong evidence of the positive effects of dual language immersion education. It also underscores the need for additional systematic research that tightly controls for selection bias or pre-existing between-group differences. Before we turn to our study, we next provide an overview of the Utah dual language immersion context.

Utah's dual language immersion education model

A major development in immersion education during the past five to seven years is the emergence of state-funded initiatives to help establish immersion programs in public schools. Until 2009, Utah, much like other states, had created bilingual education programs without sufficient funding or infrastructure for some 30 years (Leite and Cook 2015). With the passage of Senate Bill 41 in 2008, the Utah legislature funded the establishment of public 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 since followed Utah's example.[4]

As of 2015–2016, and following the addition of Portuguese in 2012 and German in 2014, Utah had 111 elementary, grades 1–6, and 27 secondary schools, grades 7–9, in 5 different languages; Arabic and Russian programs are expected to start in 2017. Elementary DLI programs are located across the state in 22 urban and rural school districts, out of a total of 36 districts with elementary schools. Utah schools establish a DLI program by submitting an application to the Utah State Board of Education requesting one of two program types, one-way or two-way, and one of the five immersion languages. Schools and districts are free to use their own procedures to enroll students once their application has been approved. When demand exceeds available slots, a few districts use a lottery system, but most establish waiting lists.

In addition to one-way programs in five languages, Utah schools house two-way elementary Spanish programs (30 out of 57 in 2015–2016), which also employ the 50/50 two-teacher model. Two-way programs require that at least a third of the enrolled students must be native speakers of the target language (Spanish or English), although a 1:1 ratio is desired. One-way and two-way programs share the main goals of dual language 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 bilingual and biliterate, and equipped with cross cultural understanding (Howard, Olague, and Rogers 2003).

To meet its significant demand for qualified teachers, Utah relies on both domestic and international guest teachers. All DLI teachers must be licensed to teach in public schools, and domestic teachers must also hold a world language and a dual language immersion endorsement when hired, or obtain these credentials within a specified time frame. Utah DLI teachers participate in mandatory pre- and in-service training and professional development together with their English classroom counterparts. In addition, regular meetings of the state-wide advisory council serve to reinforce the DLI model with school- and district-level principals and administrators.

All DLI programs in Utah that receive legislative funding are required to align with the 50/50 two-teacher model and to demonstrate fidelity to 9 program assurances for grades 1 through 6.[5] 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 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. Students' development of proficiency in the target language is measured with the ACTFL Assessment of Performance of Proficiency in Languages (AAPPL) test in grades 3–9, with the goal of 80% of DLI students meeting the proficiency benchmarks for each grade level. Academic content curricula and materials for math, social studies and science have been translated into each immersion language, and are supported by lesson plans. The original and the translated versions align with the common (or state) core, and DLI and non-DLI students alike have to meet content standards. Math and all other content areas, including social studies and science, are taught nearly exclusively in the target language in first grade through third grade. Students receive about 70 minutes of daily math instruction, which accounts for roughly 20% of the school day, as shown in Figure 1. The English classroom focuses on building English literacy, with ELA accounting for about 35% of the school day, which translates into some 140 minutes each day. Reinforcement in English of all content areas that are taught in the target language (math, science and social studies) accounts for about 15% of the day.

In fourth and fifth grade (see Figure 2), math instruction changes to 60 minutes daily in the English classroom, and 30 minutes in the target language focused on activities and practical application to

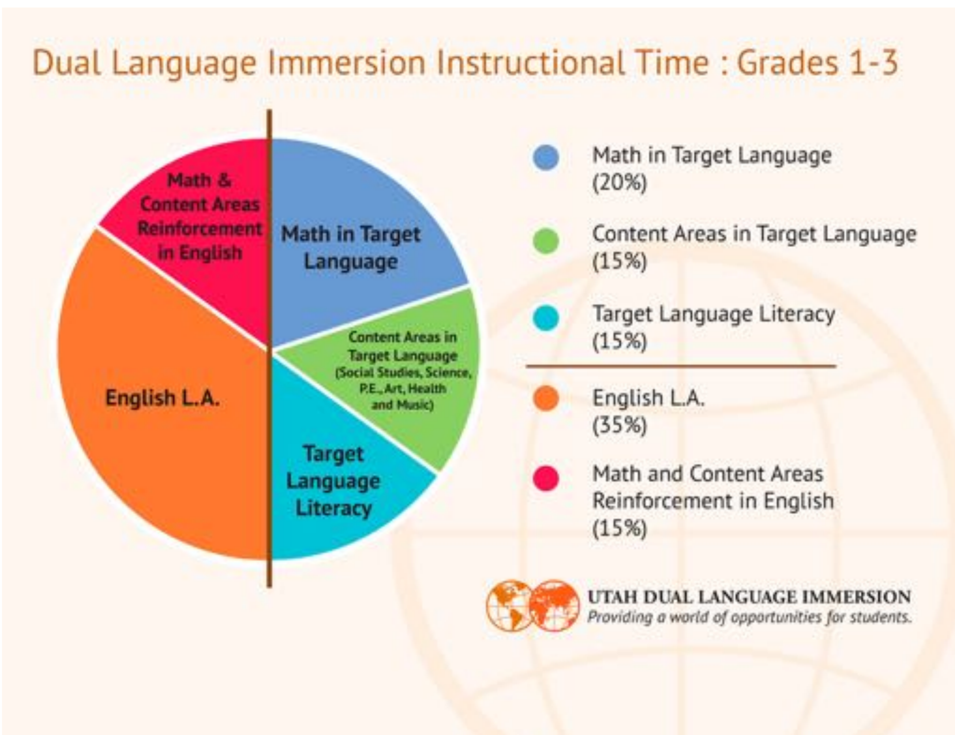

Figure 1. Dual language immersion instructional time: grades 1–3.

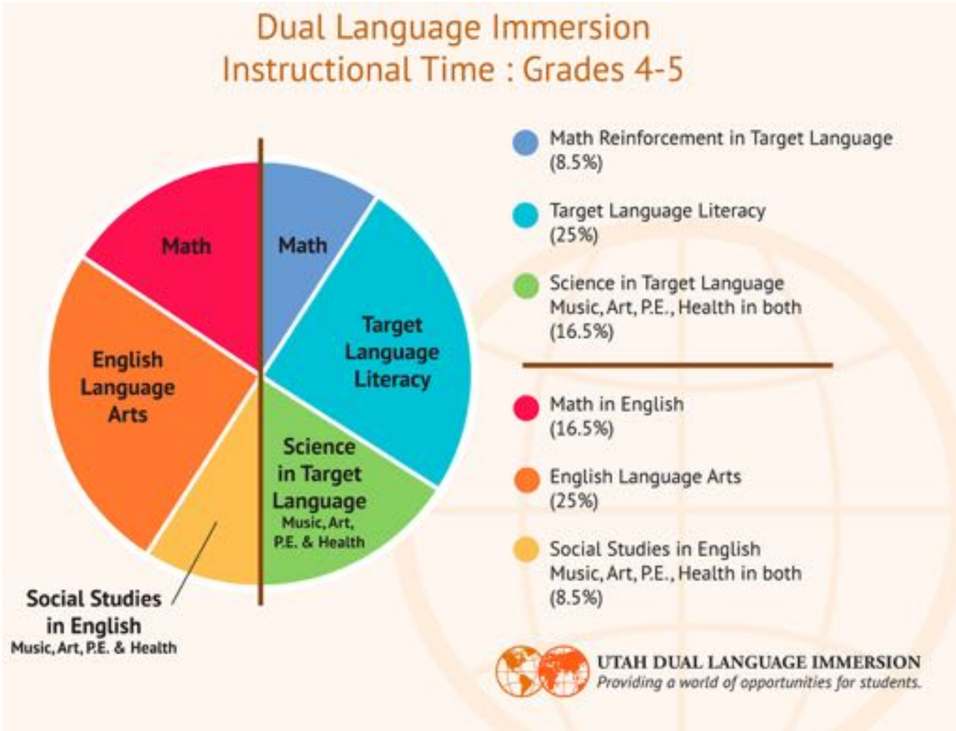

Figure 2. Dual language immersion instructional time: grades 4–5.

promote rich language use. Target language literacy instruction increases from 15% in grades one through three to 25% (about 90 minutes of daily instruction) in grades four and five.

As a whole, Utah's DLI curriculum promotes literacy instruction in English and in the target language, and its integration with content instruction. It 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).

The study

Our study sought to determine the effect of dual language immersion on academic achievement in math across three different dual language immersion languages (Chinese, French, and Spanish) and two program types (one-way and two-way). It builds on previous research, which has shown that students in dual language immersion succeed academically when content is taught fully or predominantly in the target language (e.g. Bournot-Trites and Reeder 2001; Collier and Thomas 2004; Thomas and Collier 2012; Marian, Shook, and Schroeder 2013; Steele et al. 2015). It also expands on the scope of prior research in two ways: The study examined the achievement of students enrolled in a DLI program that has been implemented state-wide, across three languages and with two program types (one-way and two-way); and it employed propensity matching of DLI and non-DLI schools, and DLI and non-DLI students to mitigate potential effects of pre-existing differences at the school and student levels. This research is also motivated by the need to demonstrate to stakeholders that students who participate in dual language immersion, and in a relatively recently established program, are performing at the same level as their non-immersion peers in content areas, as measured by state-wide accountability testing.

To measure academic performance of students in one-way and two-way Utah DLI programs, we formulated the following 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 ELA?
(2) How much growth in math is observed in the fourth grade year for 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. In the third-grade study, students provided within-subject control groups for themselves and we were able to consider a student's math performance relative to his or her performance in ELA. In the fourth-grade study, we used propensity matching. This method identified student pairs that included two students, one DLI student and one non-DLI student, who were similar academically and demographically in the third grade. The analysis then focused on differences in growth in math from the third to the fourth grade year. The propensity matching approach eliminated pre-existing differences in demographic characteristics, and test scores in ELA and math, between DLI and non-DLI students prior to fourth grade.

Samples

Two samples were used in this research: a third grade sample, which consisted of third grade students who attended DLI schools in the 2011–2012 school year, and a fourth grade sample, which consisted of fourth grade students who were in DLI in the 2012–2013 school year, and their propensity-matched peers. In both samples, students from DLI programs were identified from within 26 schools, which included 6 Chinese, 5 French, 6 one-way Spanish and 9 two-way Spanish programs.

Third grade sample

The third grade sample was comprised of all students enrolled in third grade in the 2011–2012 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 ELA test and Utah's standardized math tests.

A total of 2524 students met the inclusion criteria. A majority of these students identified as either non-Hispanic White (73%) or Hispanic ethnicity (20%), with 37% of the students qualifying for free or reduced prices lunch (FRL) and 7% qualifying for 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 ($N = 1195$) were enrolled in DLI programs and 52.7% of the students ($N = 1329$) were not.

Fourth grade sample

The fourth grade sample comprised students who were enrolled in a DLI program as fourth graders during the 2012–2013 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, 2287 students were in the fourth grade sample, including 1148 DLI students and 1139 non-DLI propensity-matched students. Consistent with demographics in the state, a majority of the students in the fourth grade sample identified as either non-Hispanic White (73%) or Hispanic ethnicity (21%), with 35% qualifying for free or reduced priced lunch and 7% qualifying for EL services in the fourth grade.

Academic outcomes

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 fourth grade SGP could range from the 1st to the 100th percentile.

Propensity matching

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 identified a similar non-DLI school for each DLI school by matching on school size, percent of students from traditionally marginalized races or ethnicities, mobility-rate, percent of students who qualified for free or reduced price lunch, and performance of third and fourth grade students prior to the study. After matching each DLI school with a non-DLI school, we matched DLI students from within each DLI school with peers from within the matched non-DLI school. Student matching was conducted using gender, ethnicity, qualification for free and reduced lunch, qualification for special education and EL services, and third grade CRT scores. As the focus of the analysis was math learning in the fourth grade, an exact match was required, by decile, for math scores. In other words, a student who scored in the 80th–89th decile in math in the third grade was matched to another student who scored in the 80th–89th decile on the math test in the third grade. We used caliper matching and required a difference of no more than .1 on the combination of all propensity matching variable in order to pair students. Using this method, we obtained suitable matches for 1139 of the 1148 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.

Data analysis

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, Bryk, and Congdon 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 to estimate the effect of DLI participation on student math scores.

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 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 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 similar students in schools with different target languages or 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 1148 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 the null model 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 the 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 the program type variable 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. Results from key DLI predictors (i.e. student participation in DLI, target language, and program type) were all non-significant. In response to research question one (How do third grade students who are and who are not in Utah DLI programs perform in math, relative to their performance in ELA?), these results indicate that students who participated in DLI performed the same in math as students who did not participate in DLI

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**

when all other factors, including ELA scores, were held constant. Results also show that student math performance was similar across schools with different target languages (i.e. Chinese, French, or Spanish) and with different program types (i.e. one-way or two-way) when all other factors were held constant.

Fourth grade study

Propensity score results

Average ASB 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 an average value of 16.706 to a value of 6.058.

Multiple regression results

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. In response to research question two (How much growth in math is observed in fourth grade students in Utah DLI programs compared to similar students who are not in DLI?), 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 their propensity-matched peers did.

Multi-level regression results

Only DLI students were included in the multi-level analyses of fourth grade growth. Results from the null model showed that for the DLI students, approximately 18% of the variance in math growth could be accounted for by which school the student attended. The proportion was significant

Table 2. Multiple regression results predicting fourth grade growth in math.

Predictor	Unstandardized coefficient	Standardized coefficient (β)	*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
EL	−7.71	−0.07	0.00
Special education	−1.17	−0.01	0.61
Hispanic	−3.16	−0.05	0.08

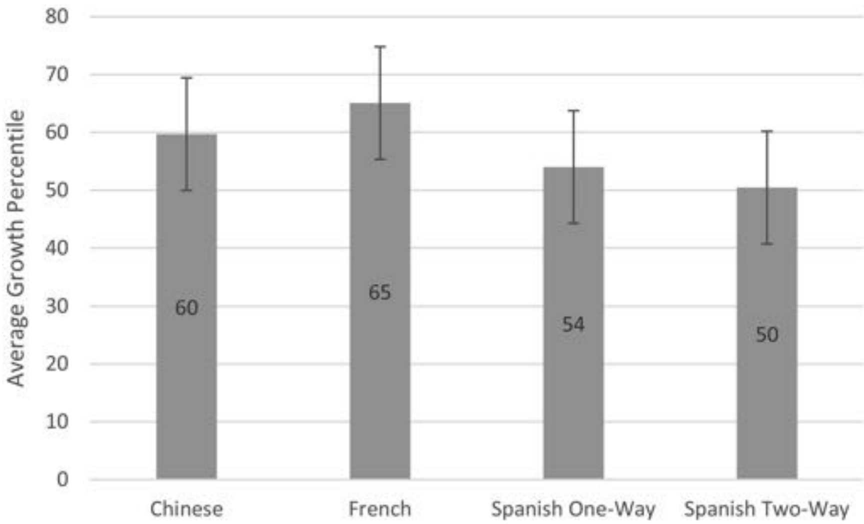

Figure 3. Average growth percentiles by target language and program type.

($X^2 = 262.26$; $p < .001$), 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 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 ELs (.148) and on students from Hispanic backgrounds (.073) indicated poor reliability. Thus the model was respecified to use only programmatic variables to predict school averages. 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. In response to research question three (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?), neither target language nor program type significantly predicted student growth in math. Figure 3 shows differences in average SGPs for students in schools with different target languages and different program types. However, because of large amounts of variance among the schools, these differences are not statistically significant.

Discussion

We conducted two studies to measure Utah DLI students' achievement in math, one in third grade and another in fourth grade. Our third grade study found that DLI students, who had received math instruction from first grade through third grade nearly exclusively in the target language, performed the same in math relative to their ELA scores as non-DLI students who received math instruction in English. We found no effect of target language (Chinese, French, or Spanish) or program type (one-way or two-way). Because no prior math achievement scores were available to permit us to rule out the possibility that DLI and non-DLI students differed prior to participation in the DLI program, we considered math performance of DLI and non-DLI students in relation to their 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.

For our fourth grade study, we were able to use students' third grade scores and their demographic characteristics to form a non-DLI control group. Each student in the non-DLI control group had a third grade math score that matched 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, who received instruction in English and in the target language, grew more in math than did matched students not in DLI. Because DLI and propensity-matched non-DLI students had similar math scores in the third grade, these results indicate that students in DLI showed more growth in fourth grade math than students who were not in DLI. 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 demographic or test score differences between the two groups.

We conducted post hoc tests to detect DLI effects on fourth grade students by EL status. The majority of students classified as EL, which made up 7% of our fourth grade sample, came from Hispanic backgrounds. Results of the post hoc tests showed that DLI and non-DLI students who were classified as EL performed similarly on CRT tests and grew similarly in the fourth grade. There was insufficient variance across the two program types to determine whether Hispanic or non-Hispanic DLI students who qualified for EL services performed differently in one-way versus two-way programs. Although we could not directly test effects, the relatively homogeneous performance of EL students across different target languages and the two program types suggested no program level (language, or one-way versus two-way) interactions with these students.

We found significant differences in both CRT scores and SGPs among the 26 schools. CRT scores and SGPs were different for schools with different target languages and different program types, but because variance *within* those groups of schools was so great, differences *between* the groups of schools were not significant. For example, SGP averages for DLI students in the six schools teaching in Chinese ranged from the 49th percentile to the 70th percentile. Given that amount of variance, many more schools would need to be sampled to detect differences between target languages or program types.

As is true for all quasi-experimental studies, causal inferences from the analyses presented in this paper should be made cautiously. We used rigorous methods to make equitable comparisons between DLI and non-DLI students, but demographic 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 could certainly have threatened the internal validity of our study. However, the matching process ensured that students matched in the third grade, serving as a good (although not perfect) control for the factor of parental support and other pre-existing factors.

Generalizability of our findings may be affected by programmatic factors such as training and quality of teachers, pedagogical practices and classroom materials. The DLI math curriculum is fairly uniform, with all DLI programs using the same math program (EnVision Math) except for Spanish, which has a second option. Moreover, training and lesson plans that are provided to DLI teachers for math instruction in the target language further help to achieve uniformity, and quality, though individual teaching practices may still vary. The differences in growth in fourth grade math between DLI and non-DLI students are not simply accounted for by the fact that students in DLI programs were taught for half of the day in a target language. It is plausible that differences in the way that students were taught accounted for differences in growth.

In contrast to DLI schools, non-DLI schools use a wide variety of math curricula and materials, and teachers may or may not receive the kind of additional support that is built into DLI math. The nature of DLI math curricula, materials and instructional practices, more standardized and more uniform than non-DLI, may have benefitted the DLI students' performance on tests. Thus, replicability of the study would be affected by factors such as uniformity.

Standardization of instruction or, put differently, adherence to the key features of an instructional model, contributes to fidelity of implementation, which, in turn, is likely to yield strong results. As Li et al. (2016) point out, immersion programs, like any other educational program, succeed when implemented well, with success for dual language immersion education principally measured by the extent to which students achieve academically in both languages (33). Additional research is needed to determine in what ways instructional practices that are specified for Utah's DLI model might contribute to the positive effect of DLI on student outcomes for math.

Another limitation to the generalizability of our findings may be the study's population. We drew our sample largely from Utah's inaugural DLI student cohort, which enrolled in the program in 2009–2010, and then participated for the first time in state-wide third grade math tests in 2012. The Utah DLI instructional model and assurances still hold, but some aspects of the DLI curriculum have been slightly modified as the program has developed, and, as the program has significantly expanded since 2009, the demographics of DLI students have slightly changed (Swenson, Mayne, and Watzinger-Tharp 2015). Follow-up studies will have to be conducted to determine whether the findings for the DLI cohort examined in this study, that is, students who completed third grade in 2012, and fourth grade in 2013, hold for subsequent DLI student populations.

Implications and conclusion

This paper reports on our recently completed research of Utah DLI and non-DLI students' academic achievement in third and fourth grade math. We found that third grade DLI students, who received math instruction almost exclusively in the target language, were at no disadvantage, performing at the same level as their non-DLI peers on the state-wide math CRT tests. From our findings for fourth grade, we were able to draw the cautious conclusion that, on average, DLI students grew more in math in their fourth grade year than similar students who were not in DLI. This effect was observed across three target languages, and two program types (one-way and two-way).

We conducted this study as a critical first phase of a comprehensive and on-going research program on Utah's DLI program, which was established relatively recently in 2009. The goal of the study reported here was to establish whether students who participate in a state-wide DLI program are able to achieve academically in math. Our findings for third and forth grade math establish a baseline for future research that will continue to evaluate DLI academic outcomes in elementary math in grades five and six, and also in science, which is first tested in grade four. In contrast to math, which is taught as a stand-alone subject beginning in first grade, science instruction is joined with social studies in first through third grade. This joined curriculum, like math, is translated and taught in the target language, and reinforced in English. Initial findings from a pilot conducted in 2014 indicate that DLI students scored lower than non-DLI students in fourth grade science, but then grew at a greater rate than their non-DLI peers in fifth grade. Since science is not subject to state-wide testing until it is taught separately in fourth grade, we will apply the methodology we used for the two math studies to compare DLI and non-DLI students' performance in science in fourth and fifth grade.

An important goal of dual language immersion research is to determine to what extent DLI benefits ELs as an intervention that helps close the achievement gap. Our current study was not designed to measure the achievement of ELs separately, though we were able to establish that ELs, who made up about 7% of both the third and fourth grade samples, achieved at levels comparable to their native English-speaking peers in one-way and two-way DLI programs. Our next step is to measure separately how well ELs who participate in DLI perform on state-wide ELA tests. Ideally, ELs are placed in two-way programs, but measures of DLI program benefits for ELs have to include one-way programs as well. Examining outcomes for ELs in one-way and two-way dual language immersion is especially important for the Utah context, with dual language immersion implemented across the state, and with demographics that create the conditions to establish two-way programs in Spanish, but not other languages.

The study's methodology and results hold important implications for dual language immersion education. We relied on propensity matching to produce a DLI sample that was closely matched with a non-DLI control group, which, in turn, allowed us to compare their performances on the math test. Our study of academic achievement of students in 17 one-way programs in Chinese, French and Spanish, and 9 Spanish two-way programs complements large-scale studies such as Thomas and Collier (1997, 2002) and the more recent Portland study (Steele et al. 2015). It also expands on prior research with a study of dual language immersion programs across an entire state, and methodology that allowed between-group comparisons. Consistent findings of studies that use large sample sizes, control for pre-existing differences, and are able to overcome selection bias will help advance the strong case for dual language immersion as an instructional model that supports students' linguistic and non-linguistic academic achievement, while also moving them toward bilingualism and biliteracy. Our analysis suggests that the DLI model of early math instruction in a target language, as implemented in Utah, may outperform monolingual math instruction.

Notes

1. Immersion education is considered a type of bilingual education. In Canada, Europe, and South America, programs are typically referred to as bilingual; dual (language) immersion has taken hold in the US though the term is applied inconsistently. A recent report on immersion education in the US advocates for uniform use of "dual language immersion" to refer to all programs that provide instruction in two languages, and to distinguish between one-way and two-way (Boyle 2016, 97).
2. The terminology to classify immersion program types continues to evolve. For example, total or full immersion instructional models are now typically referred to as 90/10, and partial immersion has largely been replaced by 50/50. Some states use 'dual language' for two-way programs only, referring to one-way as 'foreign language.' Utah refers to both 50/50 one-way and two-way immersion programs as dual language immersion, or DLI.
3. In the context of their studies of language minority students, Collier and Thomas (2004) define one-way programs as those that serve students from one language group, often heritage speakers, whose linguistic backgrounds may vary widely. In the two-way programs they included in their study, native English-speaking and minority language populations form an integrated bilingual classroom.
4. 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. In 2013, the Georgia Department of Education awarded grants for DLI to six elementary schools (Georgia Department of Education 2013).
5. These assurances are accessible on the Utah DLI homepage: http://www.utahdli.org/assurances.html.

Acknowledgement

The authors wish to acknowledge the assistance provided by the Utah Education Policy Center at the University of Utah, and the Utah State Board of Education.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Initial research for the study reported here was supported with funding from The Vice President for Research, University of Utah.

Notes on contributors

Johanna Watzinger-Tharp is an Associate Professor in the Department of Linguistics at the University of Utah. She earned her MA and PhD degrees at the University of Texas at Austin. Her research focuses on L2 methodology and teacher education, dual immersion, and language variation, and has been published in *Foreign Language Annals*, the *Modern Language Journal* and *Unterrichtspraxis/Teaching German*. Her publications also include co-edited volumes and German language textbooks. In 2010, the American Council for the Teaching of Foreign Languages (ACTFL) awarded her the Anthony Papalia Award for Excellence in Teacher Education. She serves as the co-chair of Utah's World Language Council, which produced the *Utah Language Roadmap for the 21st Century* that helped create Utah's Dual Language Immersion Program.

Kristin Swenson is a Senior Level Research Consultant at Utah's Department of Human Services, and earned her MS and PhD at the University of Utah. She specializes in quantitative analysis, focusing primarily on multivariate analyses, longitudinal analyses, and random and fixed effects models. Her areas of research in the field of cognitive psychology include implicit memory processes and reading comprehension and, in the field of education, scale development, dual language immersion, and relationships between chronic absenteeism and high school dropout. She has published in *Death Studies* and the *Journal of Gerontology: Social Science* and is currently conducting research related to issues of behavioral and mental health. As Adjunct Associate Professor in the Department of Educational Psychology at the University of Utah, she teaches advanced quantitative methods at the doctoral level.

Zachary Mayne is a Research Associate in the Utah Education Policy Center at the University of Utah and a doctoral student in the Department of Educational Psychology at the University of Utah. His educational psychology research interests include cognitive development, explicit and implicit memory processes, and intelligence. His research in education includes test validity and reliability, school climate and educator effectiveness, and dual language immersion. He has expertise in analyzing experimental and observational data, using various quantitative methods such as psychometric analysis, exploratory and confirmatory factor analyses, structural equation modeling, and analysis of variance.

ORCiD

Johanna Watzinger-Tharp http://orcid.org/0000-0001-6300-1754

References

Adesope, O. O., T. Lavin, T. Thompson, and C. Ungerleider. 2010. "A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism." *Review of Educational Research* 80: 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. 1991. "Metalinguistic Dimensions of Bilingual Language Proficiency." In *Language Processing in Bilingual Children*, edited by E. Bialystok, 113–140. Cambridge: Cambridge University Press.

Bialystok, E., and F. I. Craik. 2010. "Cognitive and Linguistic Processing in the Bilingual Mind." *Current Directions in Psychological Science* 19 (1): 19–23.

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

Bialystok, E., and S. Majumder. 1998. "The Relationship Between Bilingualism and the Development of Cognitive Processes in Problem Solving." *Applied Psycholinguistics* 19 (1): 69–58.

Bournot-Trites, M., and K. Reeder. 2001. "Interdependence Revisited: Mathematics Achievement in an Intensified French Immersion Program." *The Canadian Modern Language Review* 58 (1): 27–43.

Boyle, A., et al. 2016. *Dual Language Education Programs: Current State Policies and Practices*. American Institutes for Research. http://www.air.org/resource/dual-language-education-programs-current-state-policies-and-practices

Christian, D. 1996. "Two-Way Immersion Education: Students Learning Through Two Languages." *The Modern Language Journal* 80 (1): 66–76.

Cohen, A. D., and M. Swain. 1976. "Bilingual Education: The Immersion Model in the North American Context." *TESOL Quarterly* 10 (1): 45–53.

Collier, V. P., and W. P. Thomas. 2004. "The Astounding Effectiveness of Dual Language Education for All." *NABE Journal of Research and Practice* 2 (1): 1–20.

Esposito, A. G., and L. Baker-Ward. 2013. "Dual-Language Education for Low-Income Children: Preliminary Evidence and Benefits for Executive Function." *Bilingual Research Journal* 36 (3): 295–310.

Fortune, T. W., and D. J. Tedick, eds. 2008. Pathways to Multilingualism. Evolving Perspectives on Immersion Education. Clevedon: Multilingual Matters.

Foy, J. G., and V. A. Mann. 2014. "Bilingual Children Show Advantages in Nonverbal Auditory Executive Function Task." *International Journal of Bilingualism* 18 (6): 717–729.

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

Lazaruk, W. 2007. "Linguistic, Academic, and Cognitive Benefits of French Immersion." *The Canadian Modern Language Review* 63 (5): 605–627.

Leite, J., and R. Cook. 2015. "Utah: Making Immersion Mainstream." In *Building Bilingual Education Systems*, edited by P. Mehisto and F. Genesee, 83–96. Cambridge: Cambridge University Press.

Li, J., J. Steele, R. Slater, M. Bacon, and T. Miller. 2016. "Teaching Practices and Language Use in Two-Way Dual Language Immersion Programs in a Large Public School District." *International Multilingual Research Journal* 10 (1): 31–43.

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

Lindholm-Leary, K., and N. Block. 2010. "Achievement in Predominantly Low SES/Hispanic Dual Language Schools." *International Journal of Bilingualism and Bilingual Education* 13 (1): 43–60.

Lindholm-Leary, K., and F. Genesee. 2014. "Student Outcomes in One-Way and Two-Way Immersion and Indigenous Language Education." *Journal of Immersion and Content-Based Language Education* 2 (2): 165–180.

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

Lindholm-Leary, K., and A. Hernandez. 2011. "Achievement and Language Proficiency of Latino Students in Dual Language Programmes: Native English Speakers, Fluent English/Previous ELLs, and Current ELLs." *Journal of Multilingual and Multicultural Development* 32 (6): 531–545.

Marian, V., A. Shook, and S. R. Schroeder. 2013. "Bilingual Two-Way Immersion Programs Benefit Academic Achievement." *Bilingual Research Journal* 36: 167–86.

Met, M. 2008. "Paying Attention to Language: Literacy, Language and Academic Achievement." In *Pathways to Multilingualism: Evolving Perspectives on Immersion Education*, edited by T. W. Fortune and D. J. Tedick, 49–70. Clevedon: Multilingual Matters.

Padilla, A. M., L. Fan, Z. Xu, and D. Silva. 2013. "A Mandarin/English Two-Way Immersion Program: Language Proficiency and Academic Achievement." *Foreign Language Annals* 46 (4): 661–679.

Peal, E., and W. E. Lambert. 1962. "The Relationship of Bilingualism to Intelligence." *Psychological Monographs* 76 (27): 1–23.

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

Steele, J., B. Slater, G. Zamarro, T. Miller, J. Li, S. Burkhauser, and M. Bacon. 2015. *The Effect of Dual-Language Immersion on Student Achievement: Evidence from Lottery Data*. University of Arkansas. http://www.uaedreform.org/working-papers/

Stuart, E. A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25 (1): 1–21.

Swain, M., and S. Lapkin. 1982. *Evaluating Blingual Education: A Canadian Case Study*. Clevedon: Multilingual Matters.

Swenson, K., Z. Mayne, and J. Watzinger-Tharp. 2015. Dual Language Program Participation. Utah Education Policy Center, University of Utah.

Thomas, W. P., and V. P. Collier. 1997. *School Effectiveness for Language Minority Students*. University of California, Berkeley, CA. http://crede.berkeley.edu/research/llaa/1.1_final.html.

Thomas, W. P., and V. P. Collier. 2002. *A National Study of School Effectiveness for Language Minority Students" Long-Term Academic Achievement*. University of California, Berkeley, CA. http://escholarship.org/uc/item/65j213pt.

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

Turnbull, M., S. Lapkin, and D. Hart. 2001. "Grade 3 Immersion Students" Performance in Literacy and Mathematics: Province-Wide Results from Ontario 1998–99." *Canadian Modern Language Review* 58 (1): 9–26.