Foreign Language Annals
Published by the American Council on the Teaching of Foreign Languages

A Mandarin/English Two-Way Immersion Program: Language Proficiency and Academic Achievement

Amado M. Padilla
Stanford University

Lorraine Fan
Stanford University

Xiaoqiu Xu
Pearson Knowledge Technologies

Duarte Silva
Stanford University

Abstract: *A Mandarin/English two-way immersion elementary program is described from its inception and implementation through the fifth grade, the culminating year of the program. All students in all grades were assessed on their oral/listening, reading, and writing performance in Mandarin using program-created assessment measures. Fifth-grade students also took the Mandarin STAMP 4Se test online to assess their oral and literacy performance. In addition, all second- through fifth-grade students participated in the mandated California Standards Tests for English language arts, writing (fourth grade), math, and science (fifth grade). Results showed that across grades, Mandarin immersion students acquired high-level performance in oral/listening, reading, and writing in Mandarin. A comparison of Mandarin heritage students with non-heritage students in the immersion program initially favored heritage students in the acquisition of Mandarin; however, this advantage was not statistically significant in the later grades. On the mandated California standardized tests, the non-immersion students from the same school in the second and third grades had higher scores on the English language arts and math test,*

Amado M. Padilla (PhD, University of New Mexico) is Professor of Psychological Studies in Education, Stanford University, Stanford, CA.

Lorraine Fan (MS, Stanford University) is a PhD student in the Stanford University Graduate School of Education, Stanford, CA.

Xiaoqiu Xu (PhD, Stanford University) is a test development specialist at Pearson Knowledge Technologies, Sunnyvale, CA.

Duarte Silva (EdD, University of San Francisco) is an Executive Director of the California Foreign Language Project at Stanford University, Stanford, CA.

Foreign Language Annals, Vol. 46, Iss. 4, pp. 661–679. © 2013 by American Council on the Teaching of Foreign Languages.
DOI: 10.1111/flan.12060

but in the upper grades, Mandarin immersion students scored higher than their non-immersion peers in these two subject areas.

Key words: *Mandarin, academic achievement, elementary, immersion, language performance, standardized tests*

With the rapid rate of globalization and China's increasingly important role in the world economy, there has been a dramatic increase in interest in the study of Chinese[1] in the United States. The number of Chinese language programs in the United States, from elementary through adult programs, tripled from 1995 to 2005 and continues to expand (Asia Society, 2006; Dobuzinskis, 2011; Neely, 2011). Among those programs, Mandarin two-way immersion programs are in high demand, and the number of such programs is growing (Rogers, 2012): In 2012, there were approximately 125 (mostly elementary) schools that had Mandarin two-way immersion programs in the United States (Worthen, 2012). The largest number of such programs is found on the West Coast of the United States, with additional school districts announcing their intention to open Mandarin two-way immersion programs in the near future (Worthen, 2012).

Background

Most of what is known about two-way immersion programs comes from researchers who have studied French immersion programs in Canada and Spanish immersion programs in the United States (Lindholm-Leary, 2001). Although there is now a surge in the popularity of English/Mandarin two-way immersion programs, there have been very few published papers that have summarized the issues and complexities involved in learning Mandarin and English for children in such programs. Two exceptions are a chapter by Lindholm-Leary (2011) that presented students' self-ratings for Chinese oral language proficiency and a study by Chang (2011) that examined the impact of Mandarin instruction on academic, attitudinal, and cognitive development. The Lindholm-Leary language outcome data can only be construed as interesting because of the limitations of student self-ratings of language proficiency. Further, while Chang stated that parents and teachers believed that student academic work was enhanced with instruction in Mandarin, no quantitative data were offered to show that instruction actually led to increased academic or cognitive development.

Given the dearth of information on Mandarin/English two-way immersion programs, it is critical to understand the unique challenges that learning Mandarin and English in a two-way immersion program presents, particularly when compared to similar, and more typical, French/English or Spanish/English programs. As a starting point, it is important to first delve into the question of linguistic complexity. Languages that are taught at the Defense Language Institute Foreign Language Center in Monterey, CA, are categorized by difficulty level and length of instruction in terms of the number of weeks that are required for a learner to acquire an intermediate level of proficiency in the language.

- Category I languages are closely related to English, and an intermediate level of proficiency may generally be acquired in 26 weeks (e.g., French and Spanish).
- Category II languages exhibit more linguistic differences from English, and learners generally need 35 weeks of instruction to reach the intermediate level (e.g., German and Indonesian).
- 48 weeks of instruction are required for Category III languages (e.g., Russian, Persian-Afghan, Urdu).
- Category IV languages are exceptionally difficult for native English speakers and require 64 weeks of instruction for learners to reach the intermediate level (e.g., Chinese, Arabic). (H. Sung-Frear, personal communication, September 26, 2013).

One must also keep in mind that students who attend the Defense Language Institute

are adults, are admitted based on rigorous selection criteria, possess strong aptitude for language study, and are taught in small classes of typically no more than six learners. It would be expected that the number of weeks required to reach an intermediate level of proficiency could be much greater for diverse groups of young learners in public school, two-way immersion programs (Odlin, 1989).

Furthermore, it is important to consider the differences between English and some of the languages that are popular in two-way language immersion programs and to contrast these to Mandarin. For example, French and Spanish are both Romance languages and, while they differ in major ways from English, whose roots lie in the Germanic languages, these languages are all in the family of Indo-European languages (Fortson, 2004). As Spanish, French, and English are all in the same language family, they share a number of similarities in grammar, lexicon, and phonology. In addition, all share a Romanized alphabet and orthography where the mechanics of writing are the same—from left to right and top to bottom. There are many differences among these languages, as any language learner will attest, but the similarities over time outweigh the differences, and the second language learner can make use of metalinguistic knowledge from the first language to understand how a new language works (Odlin, 1989). Thus, a young child learning to read in a Spanish/English immersion class has only to learn a single alphabet in order to begin the process of acquiring reading and writing in not one, but two languages.

In contrast, a Mandarin/English two-way immersion program offers a specific set of challenges for teachers and students in comparison to French/English or Spanish/English programs. Mandarin belongs to the Sino-Tibetan language family, which accounts for the many differences between Mandarin and English (Li & Thompson, 1981). First, unlike English, Spanish, or French, Mandarin is a tonal language where the meaning of a word changes based on the tone used. For example, the tone that is used—high-level *mā* (mother), rising *má* (hemp), falling-rising *mǎ* (horse), and falling *mà* (scold)—completely changes the meaning of the word. Thus, the English-speaking child learning Mandarin not only has the challenge of learning a new language with a different set of phonological rules but also has the added complexity of listening for and producing tones in the new language. Another complication is that, while the English-speaking child begins to learn grammatical rules while reading and writing in English (e.g., verb tenses, subject-verb-object placement, adjective-noun placement), in Mandarin such rules are not parallel. For instance, there are no conjugations, declensions, or other inflections in Mandarin. In addition, the same verb form is used for all persons (first, second, third), and tense is expressed through adverbs of time such as *today, yesterday*, or *tomorrow* (Li & Thompson, 1981; Yip & Don, 2004). A third major difference has to do with orthography: Children in a two-way immersion program have to learn Chinese characters, a writing system that, unlike English, French, or Spanish, is not phonetic. Mastery of Chinese characters is usually considered difficult because of the large number of non-phonetic, visually complex symbols that constitute the character orthography of the language (Packard, 1990)—there are a total of 28 distinguishable types of strokes, and the number of strokes in a particular character may vary from 1 to 30 (Shen, 2005). Although the number of strokes seems to be controllable, the combination and layout of the strokes vary across Chinese characters and make them particularly challenging to write and remember for older learners and much more so for children (Xu & Padilla, 2013). Finally, in addition to learning characters, students in two-way Mandarin/English programs must also learn to read and write using pinyin, a Roman alphabet that was adopted in 1958 to facilitate the spread of Mandarin and the learning of Chinese

characters. Pinyin uses the 26 Roman letters, plus ü and four diacritics for tones.

In sum, Mandarin presents multiple layers of linguistic complexity that go beyond those found in learning Spanish or French as second languages for a native English speaker. For this reason, it is particularly important to study how two languages as disparate as English and Mandarin can be acquired by children in a two-way immersion program.

In two-way immersion programs, it is important to have a near balance of heritage and non-heritage language learners enrolled such that both groups of children have peer native language models while they learn each other's language. In addition, it is critical that in addition to gaining increasing proficiency in both languages, children also demonstrate mastery of grade-level specific content in language arts, math, science, and social studies. The purpose of this study was to track the progress of heritage language and non-heritage language learners enrolled in a public Mandarin/English two-way immersion program as they moved from kindergarten through the fifth grade and to compare their academic performance with same-age peers attending the same school but not enrolled in the Mandarin program.

Methods

Program Development

Parent interest in a Mandarin immersion program began in 2005 in a suburban school district of approximately 12,500 students in an upper-middle-income community in Northern California. The district is known for its high academic performance, with a graduation rate of 98%, and from which most high school graduates continue on to higher education. The district is composed of 13 elementary schools, including two schools of choice, three middle schools, and two high schools. The district also houses a Spanish immersion program in one of its elementary school buildings.

Although the initial plan for a Mandarin immersion program was not approved by the school board, the school district, in collaboration with the California World Language Project at Stanford University, applied for and received a Foreign Language Assistance Program (FLAP) grant to implement Mandarin instruction at the elementary and secondary levels. The FLAP grant was awarded to the district in 2007, and the school board approved the opening of a mixed-grade (K–1, 2–3, 4–5) Mandarin immersion program offered as a public elementary school-of-choice option in 2008.

Participants

Two K–1 mixed-grade classes, each including 10 kindergarten and 10 first-grade students (a total of 40 students), were enrolled in the program. A pool of children whose home language included Mandarin and whose parents expressed an interest in placing their child in the first immersion classes were given an oral interview to determine their age-appropriate proficiency in Mandarin. On the basis of the interview, 10 entering kindergarteners and 10 first graders were selected as the native Mandarin speakers. The names of non-Mandarin-speaking children were entered into a lottery from which another 10 kindergarteners and 10 first-grade students were selected. In both cases, the first-grade students already had a one-year kindergarten experience in a non-Mandarin immersion program. The researchers followed the students through their fifth-grade year. In subsequent years, a lottery system was put in place to fill the entering kindergarten class. Since the beginning of the program in fall 2008, the school has added 20 new kindergarten students each year. Table 1 shows the enrollment of students in the Mandarin two-way immersion program across grade levels. Although when they enroll their children, parents are discouraged from withdrawing their children from the program, some attrition is inevitable; thus, the actual enrollments varied by grade level.

The teachers were all native Mandarin speakers who had completed most of their

TABLE 1

Possible and Actual Student Enrollment in the Immersion Program Across Time

Year	Classes	Possible #	Actual #
2008–09	K–1, K–1	40	40
2009–10	K–1, K–1, 2	60	66
2010–11	K–1, K–1, 2–3, 2–3	80	88
2011–12	K–1, K–1, 2–3, 2–3, 4	100	106
2012–13	K–1, K–1, 2–3, 2–3, 4–5, 4–5	120	124

education in China and who received a multiple-subject credential in bilingual education from universities in California with a specialization in Mandarin.

Program Design

Students received instruction in Mandarin and English in mixed grade levels according to the following approximate balance over the K–5 program:

- K–1: 80% Mandarin instruction, 20% English instruction
- 2–3: 60% Mandarin instruction, 40% English instruction
- 4–5: 50% Mandarin instruction, 50% English instruction

The program has several additional attractive and innovative features. First, following the mixed-grade philosophy of the school, there is frequent collaboration between the older and younger children within each classroom and thus each younger group of students has regular and sustained contact with older students who can serve as linguistic and content guides. Second, curriculum and instruction are designed to be meaningful and relevant to students so that they become thinking, probing learners who search for understanding rather than memorization. Furthermore, social and emotional development are regarded as being as important as academic development. Finally, parent involvement is viewed as an important and crucial element of the program, and parents are actively involved in the program by assisting teachers in teaching subjects, organizing activities, or maintaining classroom discipline.

Students in the Mandarin two-way immersion program receive instruction in the core curriculum based on the California and district content standards for elementary school. In addition, prior to the start of the program, a curriculum was created that infused best practices in structuring Mandarin instruction in a developmentally appropriate manner. The Mandarin curriculum called for using simplified characters and pinyin and included the following features: explicit language instruction in reading and writing Chinese characters, correct form and stroke order for writing in Chinese, content-based instruction, and real-life opportunities to use Mandarin.

Assessment

The purpose of this study was to (1) examine the listening/oral, reading, and writing performance in Mandarin of students who had participated in a Mandarin two-way immersion program for varying lengths of time, and (2) compare immersion students' academic achievement in English language arts, math, and science with scores from students from the same elementary school who were not in the Mandarin program on statewide achievement tests that are administered to all students in California.

Mandarin Proficiency Assessment
In order to assess students' language acquisition across the four linguistic skills, a Mandarin Proficiency Assessment (MPA) was developed through collaboration between the Mandarin immersion teachers and researchers at Stanford University. The MPA was developed as an end-of-school-year assessment tool and initially was created for the K–1 class. In the second year of the program, the second-grade assessment was added to the MPA, with the third-, fourth-, and fifth-grade levels developed in successive years. The MPA consisted of three major sections: oral language, reading, and writing.

- Oral assessment: In the oral section, students had to first successfully complete a listening task prior to beginning the conversation task. Only students who successfully completed the conversation task proceeded to the picture-based storytelling task. The oral proficiency rubric (see Appendix) assessed oral performance at six levels, from naming objects to describing them in context.
- Reading assessment: The reading assessment involved word- and sentence-level reading for levels 1 through 4, vocabulary and story comprehension for levels 5 and 6, and story comprehension for levels 7 and 8. The reading rubric (see Appendix) assessed reading competency at eight levels, from recognizing characters/words to comprehending short stories.
- Writing assessment: The writing assessment required students to write about a topic, such as introducing themselves or their families. The writing rubric (see Appendix) assessed student performance using an assessment rubric designed to evaluate students on their content, vocabulary, organization, mechanics (grammar, punctuation), penmanship, completion, and effort. A maximum of 20 points were equally distributed among these five areas.

Classroom teachers administered the MPA near the end of each academic year. Teachers also scored the MPA for each student. For the descriptions of the assessment levels in each area, please refer to the Appendix.

Standards-Based Measure of Proficiency (STAMP)
The STAMP test (Avant Assessment, 2012) was selected as a standardized external measure of students' Mandarin performance and was used to assess end-of-program acquisition of Mandarin. In spring 2013, 14 fifth graders took the STAMP 4Se assessment. The computerized assessment was designed and rated by Avant Assessment.

California Student Achievement Measures
Students' performance in reading/English language arts, writing, math, and science was assessed employing the California Standardized Testing and Reporting (STAR) measure that is administered annually to all students in the state starting in the second grade. This assessment provides individual performance results for all students.

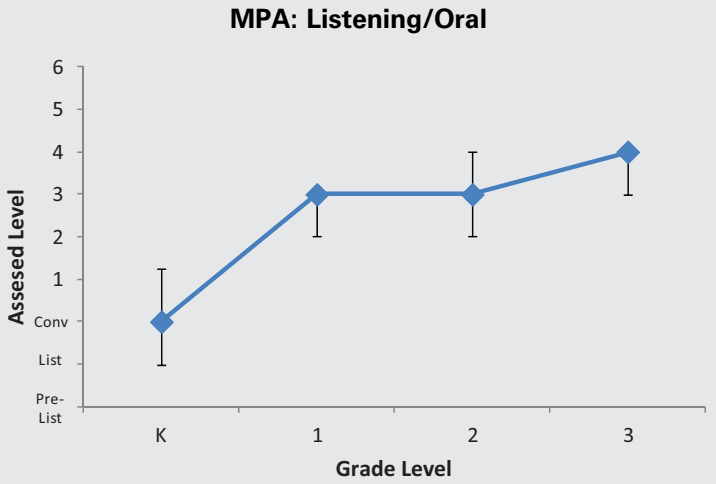
Results

In order to examine immersion students' progress in Mandarin, a cross-sectional view of their language development from kindergarten through fifth grade using data from 2010 to 2013 is provided.[2]

Listening/Oral Performance

Students' scores on the listening and speaking sections of the MPA are presented in Figure 1. The data in Figure 1 demonstrate that, at the end of kindergarten, the median student was only able to engage in basic conversations, whereas 31% of the students were merely managing to listen to and follow some instructions. However, a year later, the median student was already able to describe what happened in a story illustrated by a picture (level 3). The median remained the same in second grade, but a larger percentage of students were able to perform at or above level 3. By the end of third grade, the median student was able to answer teachers' questions with one to two

FIGURE 1



Note: In all figures with markers and error bars, the markers represent the medians while the error bars indicate the upper and lower quartiles.

sentences (level 4); specifically, the median student could reflect on the story in the picture and describe how and why something happened and what would follow.

Heritage and non-heritage learners' scores are presented in Figure 2. As shown in Figure 2, heritage language speakers performed moderately better than non-heritage language speakers. Although their median levels differed by zero to one level from kindergarten through third grade, heritage language speakers tended to concentrate more in the higher levels for each grade, while non-heritage language speakers were proportionately more concentrated in the lower spectrum.

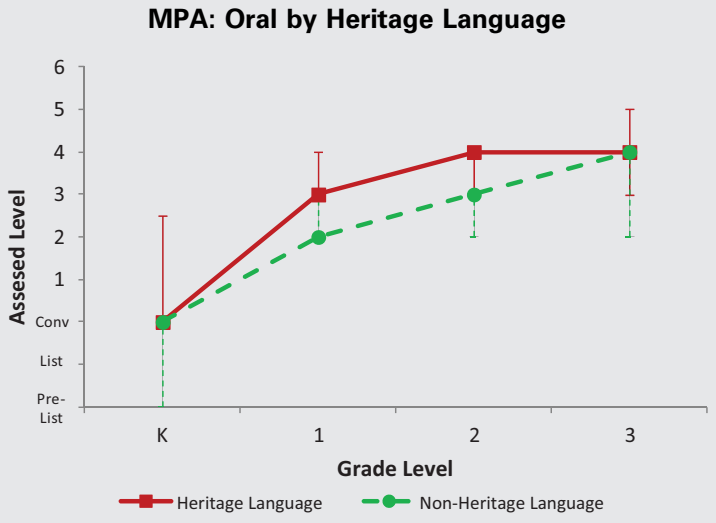
STAMP scores for fifth graders are presented in Figure 3. Of the 14 fifth graders who took the STAMP4Se assessment on listening, 57% achieved level 6, whereas the remaining students were evenly distributed between levels 3 through 5. According to the assessment rubric, students who score at level 6 are able to "understand and use language for straightforward informational purposes" (Avant Assessment, 2012, n.p.).

Speaking samples on the STAMP 4Se were collected from 13 fifth graders, as shown in Figure 4. Sixty-two percent of the students were rated level 4, which corresponds to Intermediate Low on the STAMP assessment scale. According to the developers of the STAMP 4Se test, this level is characterized by good accuracy with formulaic sentences and added detail. Further, good control is expected with a majority of the responses and, while some errors may occur as a student attempts higher-level skills, the important thing is that the student reaches beyond his or her level of comfort (Avant Assessment, 2012). This finding is encouraging and shows that the fifth-grade students had acquired an intermediate level of oral competence in Mandarin by the time they exited the two-way Mandarin immersion program.

Reading Performance

The MPA also revealed increases in students' reading proficiency in Mandarin, as shown in Figure 5. Overall medians were level 1, indicating that students could identify at least 40 out of 50 test words, chosen from a pool of approximately 100 basic high-frequency words, in kindergarten.

FIGURE 2



Students progressed to level 2 in first grade, level 3 in second grade, and level 4 in third grade, and made a large jump to level 7—reading the words of a 290-word story aloud and answering five comprehension questions—by fourth and fifth grade. At the conclusion of kindergarten, 45% of the students were not ready for level 1 words, and 38% were rated at level 1. In contrast, 66% of the fourth and fifth graders combined were able to orally express their understanding of the level 7 material. Heritage language speakers as a whole outperformed their non-heritage language peers on the reading portion of the MPA by about one level from kindergarten to third grade (Figure 6). Nonetheless, the small samples of the two groups were similar in fourth and fifth grades.

Results for the 14 fifth graders who took the STAMP4Se assessment in reading are presented in Figure 7. The fifth graders' reading scores were roughly evenly distributed among levels 3 through 6 on the STAMP 4Se. According to the assessment rubric, Novice students at level 3 rely on basic vocabulary and phrases and can recognize the purpose of basic texts. On the other hand, students who score in the Intermediate

FIGURE 3

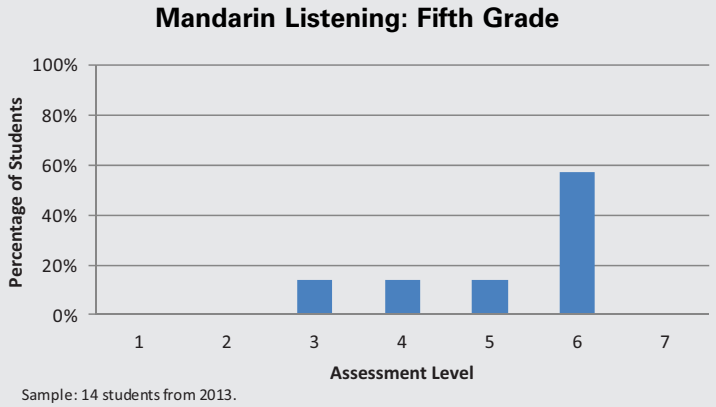
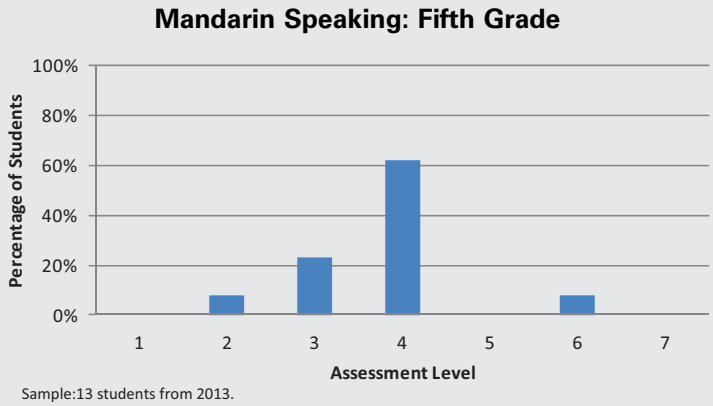


FIGURE 4



Low to High range (levels 4 to 6) in reading "understand the main ideas and explicit details in everyday language … and understand information in everyday materials … answer questions about the main idea and explicitly stated details" (Avant Assessment, 2012, n. p.). This is encouraging because it demonstrated that Mandarin two-way immersion students who were about to exit from the program had made significant advances in understanding written Mandarin (reading comprehension). Interestingly, the Spearman rank correlation between the MPA and STAMP 4Se assessments on reading was 0.77.

Writing Performance

The majority of kindergarteners and first graders were not expected to be able to write complete sentences in Chinese characters. Therefore, the writing assessment commenced in second grade. As mentioned earlier, writing samples were evaluated on five dimensions: content; vocabulary; organization; mechanics (grammar, punctuation); and penmanship, completion, and effort. The results from 2012 and 2013 are summarized in Figure 8. The median score out of a maximum of 20 points was 11 in second grade, 14 in third grade, and 16 in fourth and fifth grade. While only 7% of the second

FIGURE 5

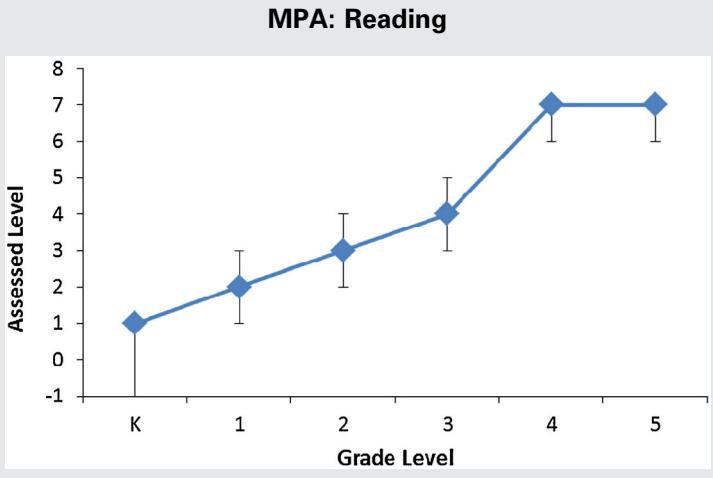
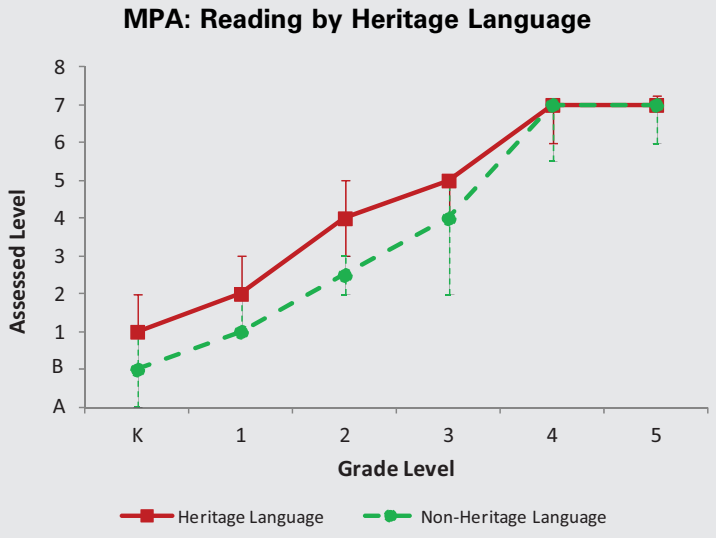


FIGURE 6

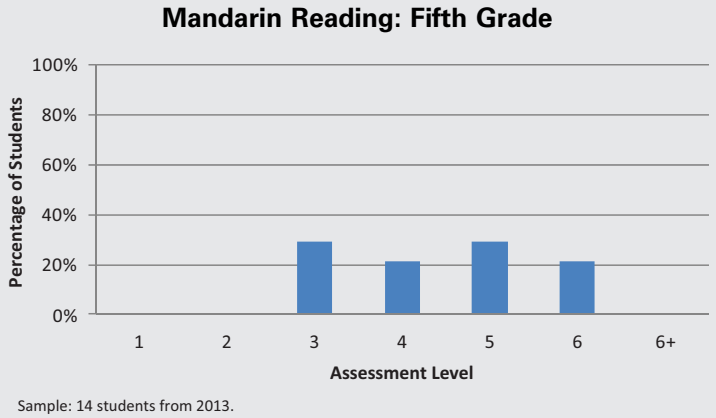


graders were barely ready for the writing assessment and 21% of them received low scores, the large majority of students scored in the middle and upper ends of the spectrum in subsequent grade levels. Overall, in fourth and fifth grades, 94% of the students received 13 points or more.

Data for heritage and non-heritage learners are presented in Figure 9. Heritage and non-heritage language speakers scored similarly, although the former group enjoyed a slight edge in terms of score distributions.

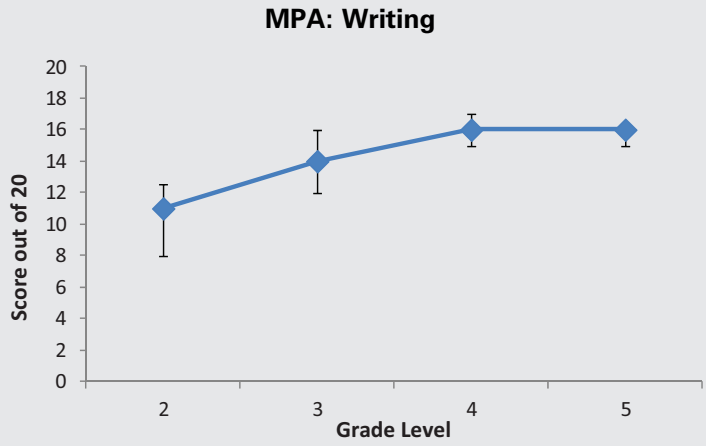
Writing samples on the STAMP 4Se were also collected from 14 fifth graders and are shown in Figure 10. Seventy-one percent of the 14 fifth graders were rated level 4. This level is characterized by "good accuracy with formulaic sentences and added detail" but with occasional errors upon attempts at higher-level skills, and a possible lack of flow in delivery. This result was consistent with that of speaking. The Spearman rank correlations of the STAMP 4Se were 0.53 with MPA scores, and 0.61 with MPA levels.[3]

FIGURE 7



Sample: 14 students from 2013.

FIGURE 8



Sample: 43 second-grade, 42 third-grade, 21 fourth-grade, and 14 fifth-grade students in 2012 and 2013 combined.

Academic Achievement in English Language Arts and Writing

In California, all students are required to take the STAR exams beginning in second grade. STAR test results for Mandarin immersion students are presented in Table 2 and compared with the scores of students who were enrolled in the same elementary school but were not in the immersion program. These data are particularly important in helping to understand how the Mandarin immersion students performed in English language arts on California's standardized tests even though only a portion of the school day was spent using English and in English language instruction. A standard score was computed for each student test outcome to represent the difference between the original score and the mean of that grade across the immersion and non-immersion programs of the school in the corresponding test year in terms of number of standard deviations. Standard scores were used because student composition could vary

FIGURE 9

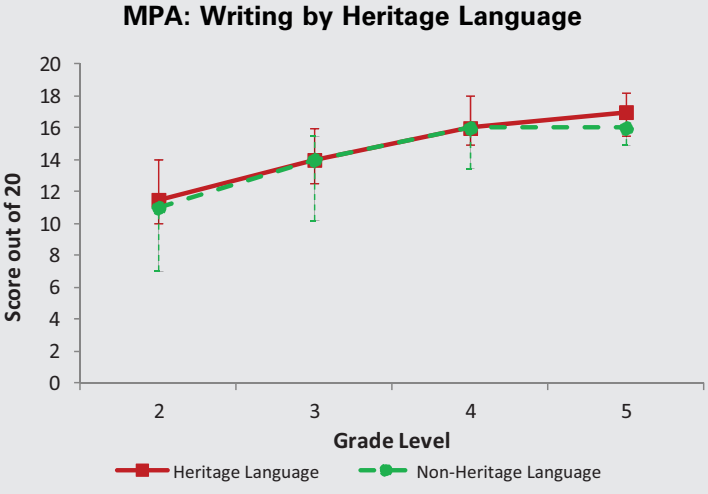
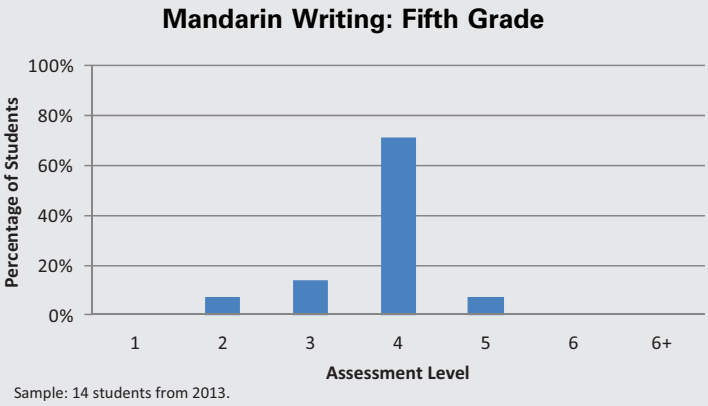


FIGURE 10



from year to year, and such variance could present a particular concern for small samples. The data were then combined across test years for each grade. Using the standard scores, immersion and non-immersion groups could be compared, giving each student equal weight. In addition, within the immersion program, comparisons could be made between heritage speakers and non-heritage speakers.

In second grade, the mean z-score of the immersion students on the STAR tests for English language arts was 0.33 lower than that of the non-immersion students, $t(131.11) = 2.70$, $p = 0.01$.[4] In contrast, by fourth grade, the average z-score of the immersion students was 0.51 higher, $t(47.56) = 2.56$, $p = 0.01$. The differences in immersion and non-immersion group means in third and fifth grades were not statistically significant at $p < 0.05$, although it should be noted that the small sample (13 fifth-grade immersion students) could have contributed to the lack of a significant finding. None of the differences between the heritage and non-heritage language students in the immersion program were statistically significant.

When comparing the students in the immersion and non-immersion programs, the proportion of students who were

TABLE 2

Distribution of STAR Results* for English Language Arts, 2010–13

Grade	Immersion	Far Below Basic	Below Basic	Basic	Proficient	Advanced	N
2	Y	4%	5%	17%	39%	36%	84
	N	0%	2%	17%	37%	45%	300
3	Y	0%	0%	16%	35%	49%	55
	N	0%	2%	16%	36%	46%	304
4	Y	0%	0%	3%	11%	86%	36
	N	1%	1%	6%	21%	72%	156
5	Y	0%	0%	0%	23%	77%	13
	N	0%	0%	4%	23%	73%	82

*Individual scores on the different standardized tests are then assigned to one of five different performance levels: Advanced, Proficient, Basic, Below basic, and Far below basic, which indicate students mastery of state standards.

considered proficient or advanced in English language arts never differed by more than six percentage points. As can be seen in Table 2, the number of immersion and non-immersion students, respectively, who were evaluated to be proficient or advanced were 75% and 81% in second grade, 84% and 82% in third grade, 97% and 92% in fourth grade, and 100% and 96% in fifth grade.

Fourth graders also took the STAR assessments in writing. Test scores were reported using different scales in 2012 and in 2013. In 2012, scores were 2, 4, 6, or 8, and in 2013, the scores were reported as 1, 2, 3, or 4. Students' scores on the 2012 and 2013 writing test are presented in Table 3. In 2012, the immersion and non-immersion students performed similarly, each with approximately a quarter of the students receiving the second highest possible score, and three-quarters receiving the maximum score (Table 3). In 2013, the immersion students performed slightly better than the non-immersion students: Fifty-nine percent of the immersion students attained the maximum score when writing in English, in comparison with their non-immersion counterparts' 43%.

Academic Achievement in Math and Science

On standardized tests of mathematics, immersion and non-immersion students performed similarly in second and third grades. Subsequently, immersion students began to outperform their non-immersion peers, as shown in Table 4. In fourth grade, the mean *z*-score of the immersion students was 0.85 higher than the non-immersion students' scores, $t(47.96) = 4.47$, $p < 0.001$. In fifth grade, the mean *z*-score of the immersion students was 1.05 higher than the non-immersion students' scores, $t(16.85) = 3.95$, $p < 0.01$. None of the differences between the heritage and non-heritage language immersion students were statistically significant at $p < 0.05$.

The proportion of students who were considered proficient or advanced with math generally suggested an edge for students in the immersion program over students in the non-immersion program. The number of immersion and non-immersion students, respectively, evaluated to be proficient or advanced were 96% and 88% in second grade, 84% and 86% in third grade, 92% and 79% in fourth grade, and 100% and 83% in fifth grade (Table 4).

In fifth grade, students also took the STAR test in science. The difference in scores between the immersion and non-immersion groups was not statistically significant: The proportions of students who were assessed to be proficient or advanced were 92% for the immersion group and 93% for the non-immersion group.

TABLE 3

Distribution of STAR Results for Writing in Fourth Grade, 2012–13

Year	Immersion	2	4	6	8	Not Scored*	*N*
2012	Y	0%	0%	23%	77%	0%	13
	N	0%	3%	22%	72%	3%	76
Year	**Immersion**	**1**	**2**	**3**	**4**	**Not Scored**	***N***
2013	Y	0%	0%	41%	59%	0%	22
	N	0%	3%	52%	43%	3%	77

*One student submitted a blank paper, while a second student wrote on a topic other than the assigned writing prompt.

TABLE 4

Distribution of STAR Results in Math, 2010–13

Grade	Immersion	Far Below Basic	Below Basic	Basic	Proficient	Advanced	*N*
2	Y	0%	2%	1%	26%	70%	84
	N	0%	3%	9%	28%	61%	301
3	Y	0%	2%	14%	11%	73%	56
	N	0%	3%	11%	25%	60%	303
4	Y	0%	0%	8%	17%	75%	36
	N	1%	3%	18%	30%	49%	155
5	Y	0%	0%	0%	8%	92%	13
	N	0%	2%	15%	36%	47%	81

Discussion

Over the last two decades, two-way language immersion programs have become increasingly available in elementary schools throughout the United States. Spanish/English two-way immersion programs constitute the most prevalent form of such programs and have the longest history (Lindholm-Leary, 2001). Because of the growing importance of China as an economic power, there has been a marked interest on the part of parents, language educators, and policy makers in Mandarin/English immersion programs, especially on the West Coast of the United States. Although these programs present interesting challenges for school administrators due to the lack of both credentialed Mandarin/English teachers and curricular materials (Chang, 2011), these challenges are not substantially different from challenges experienced in other two-way language programs. In addition, one of the crucial concerns across two-way immersion programs in any language is that students not only emerge as bilingual and biliterate, but they also demonstrate academically equivalent gains in the core content areas of English language arts, math, and science when compared to their peers who have not been enrolled in a two-way immersion program.

The results reported here present strong evidence that students can attain a high level of performance in Mandarin as assessed by both teacher-developed and nationally recognized measures of oral skills, reading, and writing. The data further indicate that learning progresses in an orderly fashion from kindergarten through fifth grade for oral, reading, and writing tasks. An interesting, but not completely surprising, finding was that heritage language students initially showed higher levels of attainment in Mandarin than did non-heritage students. However, by fourth and fifth grade these differences diminished, although heritage learners still performed slightly higher. For example, in assessing oral Mandarin, the rubric called for accuracy in language use, which included both accuracy of tones and fluency in speaking. Results from both sets of assessments indicated that non-heritage learners mastered language use in similar ways as heritage learners. Finally, even though the fifth-grade cohort was small ($n = 14$), an objective assessment—the STAMP 4Se—supported the teacher-developed measures in indicating that students were indeed performing at a comparable level in Mandarin listening, speaking, reading, and writing.

An equally important question for program developers and parents addresses the extent to which students in a two-way immersion program achieve at comparable levels on state-mandated accountability measures in the core content areas of English language arts, math, and science as same-age peers who are not enrolled in language immersion programs. From a policy perspective, educators want to be certain that the resources spent on two-way immersion programs produce students who are not only bilingual but are also academically prepared in the essential core areas as defined by state accountability measures. While parents may be convinced of the long-term benefits of a bilingual program, they too are concerned that their children not lag behind peers who receive their schooling entirely in English.

The results from this study show that students who are taught in Mandarin for much of the school day generally achieve at levels on California-mandated tests in English language arts, writing, math, and science that are as high as, or sometimes higher than, their non-immersion peers who attend the same school. These results are reassuring because they demonstrate that, when students receive instruction in two languages, they are not only developing as bilinguals but also do not fall behind their peers on the essential content. These findings also support recent reports that show cognitive advantages due to bilingual instruction (Bialystok, Craik, Green, & Gollan, 2009; Kluger, 2013).

It is important to note that this study is not without its limitations. A limitation of all naturalistic studies such as this one is the relatively small sample size in each grade. This was especially true in the fifth grade, where, due to attrition, the sample contained only 14 of the original 20 students. However, despite this problem, many of the findings were statistically significant. The study is also limited due to the use of program/teacher-developed and teacher-administered assessments of Mandarin, a problem that was unavoidable for two reasons. First, due to the shortage of Mandarin-speaking professionals at the school, it was necessary to have the teachers conduct the assessments of Mandarin. Second, there is a shortage of standardized assessment instruments for use in elementary Mandarin/English two-way immersion programs particularly because the STAMP 4Se is an online assessment measure developed for older learners and thus could only be administered with fifth-grade students. While the potential of teacher bias in the assessment process should be recognized, the correlations between the teacher-reported assessment results and the fifth-grade students' scores on the standardized STAMP 4Se suggest that teacher bias was not a significant source of concern. Ideally, future research with Mandarin/English programs will employ assessments that are independently developed and that can be used beginning in kindergarten.

Conclusion

This article presents five years of performance data from the first cohort of students who completed a two-way Mandarin immersion program. Results suggest that, although Mandarin is a non-alphabetic language, English-dominant students can attain desirable levels of linguistic performance in Mandarin while heritage speakers are able to maintain and enhance their listening and speaking skills in their home language while simultaneously developing important literacy skills. An equally important finding is that when compared to non-Mandarin immersion peers from the same school, Mandarin immersion students performed as well on standardized tests that were given in English and that assessed English language arts and other core content areas even though they had much less instructional time in English. This is particularly informative for educators, parents, administrators, and policy makers who are interested in preparing students for today's global economy and society.

Acknowledgments

The authors gratefully acknowledge the support of a FLAP grant awarded by the U.S. Department of Education that provided funding for the evaluation study of this Mandarin immersion program. In addition, the authors are appreciative of the contributions of several individuals who collaborated with the authors to successfully conduct this study, including district- and school-level administrators; teachers who taught in the program; and the students who actively acquired Chinese language, culture, and literacy skills. We are also thankful to the program's parent community who enthusiastically supported the study's evaluation activities. The authors also wish to thank the *Foreign Language Annals*' editor and reviewers, whose encouragement and comments strengthened our article.

Notes

1. Chinese includes two major common languages: Mandarin and Cantonese. As Mandarin is the standardized language used in China, it has become increasingly popular worldwide. In this article, *Chinese* and *Mandarin* are used interchangeably.
2. The data on the cohort of the fourth grade in 2012 could not be retrieved because of the unexpected departure of the classroom teacher.
3. In order to compute a Spearman rank correlation, five levels were created based on STAMP4Se scores. Levels were as follows: 0–4, 5–8, 9–12, 13–16, and 17–20.
4. A Welch *t* test for unequal variances was used for this comparison.

References

Asia Society. (2006). *Creating a Chinese language program in your school*. New York: Asia Society.

AVANT Assessment. (2012). STAMP Proficiency Assessment: Web based language assessment in Chinese. Retrieved October 8, 2013, from http://avantassessment.com/support#stamp

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*, 89–129.

Chang, W. D. (2011). *Perceptions on the impact of K-5 Mandarin foreign language instruction on students' academic, attitudinal, and cognitive development* (Unpublished doctoral dissertation). California State University, Northridge.

Dobuzinskis, A. (2011). *More U.S. students learning Chinese as school language programs expand*. Retrieved April 20, 2012, from http://www.huffingtonpost.com/2011/04/21/more-us-students-learning_n_852093.html

Fortson, B. W. (2004). *Indo-European language and culture: An introduction*. Oxford: Blackwell.

Kluger, J. (2013). The power of the bilingual brain. Retrieved October 8, 2013, from http://www.time.com/time/subscriber/printout/0,8816,2147707,00.html#

Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

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

Lindholm-Leary, K. (2011). Student outcomes in Chinese two-way immersion programs: Language proficiency, academic achievement, and student attitudes. In D. Tedick, D. Christian, & T. Fortune (Eds.), *Immersion education: Practices, policies, possibilities* (pp. 81–103). Avon, UK: Multilingual Matters.

Neely, B. (2011). *More Americans learn their ABCs in Chinese*. Retrieved April 20, 2012, from http://www.npr.org/2011/01/19/133031008/american-interest-in-learning-chinese-skyrockets

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. New York: Cambridge University Press.

Packard, J. L. (1990). Effects of time lag in the introduction of characters into the Chinese language curriculum. *Modern Language Journal*, *74*, 167–175.

Rogers, M. (2012). Demand high for dual immersion programs in Utah. *The Salt Lake Tribune* [Electronic version]. Retrieved October 8, 2013, from http://www.sltrib.com/sltrib/news/./language-programs-chinese-dual.html.csp

Shen, H. H. (2005). Linguistic complexity and beginning-level L2 Chinese reading. *Journal of*

the Chinese Language Teachers Association, *40*, 1–28.

Worthen, B. (2012, August 22). More area schools embrace Chinese immersion method. *The Wall Street Journal*, Retrieved October 14, 2013, from http://www.wsj.com

Xu, X., & Padilla, A. M. (2013). Using meaningful interpretation and chunking to enhance memory: The case of Chinese character learning. *Foreign Language Annals*, *46*, 402–422. doi:10.1111/flan.12039.

Yip, V., & Don, R. (2004). *Chinese: A comprehensive grammar*. London: Routledge.

Submitted April 2, 2013

Accepted October 2, 2013

APPENDIX

Mandarin Proficiency Assessment

Oral Proficiency Rubric. In order to assess oral proficiency, a storytelling method was used whereby the teacher showed the student a picture and asked if the child could do the following (see table below). The proficiency level of the student was determined by the accuracy with which s/he performed the task. There were six levels for assigning an oral proficiency.

Level	Description
1	Name the objects in the picture: What do you see in the picture?
2	Describe the objects in the picture (color, shape, size).
3	Describe the story in the picture: What is happening in the picture?
4	Think about the story and describe how and why something happened and what will happen next.
5	Answer all four questions with supporting details, accurate pronunciation, and fluency; make almost no grammatical errors; sentence structure is varied.
6	Use connection words for extended description of the pictures; use vivid descriptive words and some idioms to enhance meaning; story has a clear progress of ideas including clear beginning, middle, and end.

Reading Proficiency Rubric. Reading proficiency was determined by requiring students to read characters, sentences, and paragraphs. The level designates how proficient the student was in reading Chinese.

Level	Vocabulary	Sentence/Paragraph
1	50 characters/words	N/A
2	50 characters/words	A few simple sentences on a topic
3	50 characters/words	A few longer sentences on the same topic
4	50 characters/words	A short paragraph on the same topic (five to six sentences)
5	50 characters/words	A reader was used with 91-character story
6	50 characters/words	99-character story
7	–	290-character story
8	–	274-character story

- Students attaining reading proficiency levels 1, 2, and 3 were required to master the 150 most basic characters/words before they advanced to level 4. The 50 sentences/words at each level were selected based on their difficulty and frequency level—as a student progressed up the levels, the characters became more difficult because of their complexity and frequency. Level 1 students were not required to read sentences. Level 4 students and higher went directly to sentence/paragraph reading, but continued to include additional vocabulary words.
- For characters/words, students had to score at or above 80% on accuracy and comprehension to be able to move to the next level. For sentences, there were certain characters/words that students were allowed to miss. Students were also allowed to make a mistake in one additional character/word to be able to move up to the next level.
- The sentences/paragraphs at the various difficulty levels were all based on the same topic but were sequential in terms of length and the difficulty level of lexicon and syntax. Teachers could add levels onto the list along with their curriculum and students' language progress.
- In order to assign a proficiency level to students who were just beginning Mandarin instruction, it was necessary to make an adjustment in level 1. Accordingly, students who scored between 0 and 19 correct in level 1 characters were marked as level A, and students who scored 20–39 correct were marked as level B.

Writing Rubric Starting From 2012. *Note:* The assessment rubric of STAMP4Se is available at http://www.avantassessment.com/sites/default/files/STAMP4Se%20Benchmarks%20and%20Rubric%20Guide%202012%20Mar.pdf

	1	2	3	4
Content	Off topic; most of the information is irrelevant or unintelligible.	Partially addresses the topics; demonstrates basic concepts/ideas but lacks detail.	Addresses the topics; may include some supporting details.	Fully addresses the topics with appropriate and concrete details.
Vocabulary	Limited word choice (most could be the hint words); may substitute with English words; incorrect and confusing words are frequent.	May use basic words and/or substitute with pinyin; some words may be overused; incorrect and confusing words are obvious.	Appropriate use of words; some varieties in word choices; rare use of English/pinyin substitution; errors are few.	Appropriate use of descriptive and vivid vocabulary; may include idioms that enhance meaning; almost no errors.
Organization	Repetitive simple sentences, fragments, or isolated words; sequence of information is difficult to follow.	Scattered information; individual sentences may be understandable but lack coherence for the whole piece.	There is a main idea, but may include disorganized details; may use some connection words, but transition may not flow with the whole text.	Clear progression of ideas (clear beginning, middle, and end) and well-connected discourse; appropriate usage of connection words.

(Continued)

	1	2	3	4
Mechanics (Punctuation/ Grammar)	Many grammatical errors; may frequently miss punctuation; awkward sentence structures significantly interfere with meaning.	Noticeable inaccurate punctuation/ grammatical errors, but they do not cause significant misunderstanding.	Few inaccurate punctuation and grammatical errors; sentence structures are mostly appropriate.	Almost no incorrect characters or punctuation and grammatical errors; sentence structures are varied.

Copyright of Foreign Language Annals is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.