



A High School Intensive Summer Mandarin Course: Program Model and Learner Outcomes

Xiaoqiu Xu
Pearson Education

Amado M. Padilla
Stanford University

Duarte Silva
Stanford University

Norman Masuda
Palo Alto Unified School District

Abstract: *This article describes a STARTALK intensive summer high school Mandarin language and culture program that was conducted for three summers. Participants across the three years included 40 Mandarin Level II and 53 Mandarin Level III high school students. Quantitative and qualitative data are presented to show the effectiveness of the program. Students' language proficiency scores showed significant improvement between pre- and posttests for listening, speaking, reading, and writing. Interview and observational data are also discussed to show students' improvement in all four Mandarin language skills, as well as their increased knowledge of Chinese culture. A detailed program description including program goals, curriculum, instruction, and assessment is also shared for those who are interested in replicating the program model.*

Key words: *Mandarin, high school summer program, intensive language program, learning Mandarin as a foreign language, model program*

Xiaoqiu Xu (PhD, Stanford University) is a Research and Test Development Manager at Pearson Education, Inc., Palo Alto, California.

Amado M. Padilla (PhD, Stanford University) is a Professor of Developmental and Psychological Sciences at Stanford University, Stanford, California.

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

Norman Masuda (MA, Stanford University) is an Instructional Supervisor in the Palo Alto Unified School District, Palo Alto, California.

Foreign Language Annals, Vol. 45, Iss. 4, pp. 622–638. © 2013 by American Council on the Teaching of Foreign Languages.

DOI: 10.1111/flan.12005

Introduction

Although Chinese was first introduced as a language of instruction in the United States in 1879 (Lum, 2008), it was not commonly taught in American schools and colleges until China's recent economic rise to prominence (Xinhua, 2008). Recent studies by Asia Society and the College Board indicate that there is a strong and expanding interest in Mandarin language study among U.S. students (Asia Society and the College Board, 2008). Currently, it is estimated that about 60,000 K–12 students are studying Chinese (Robelen, 2010, n.p.). The U.S. government has paid particular attention to boosting Chinese language instruction among other foreign languages and has designated Chinese as a critical language for study. The number of Chinese language programs in the United States, from elementary through adult programs, has tripled in 10 years (Ruethling, 2005) and continues to expand (Dobuzinskis, 2011; Neely, 2011).

Starting in 2007, STARTALK, a nationwide program funded by the National Security Language Initiative (NSLI), has provided summer intensive language and cultural programs to a wide range of students and instructors learning and teaching Arabic, Chinese, Hindi, Persian, Swahili, Turkish, and Urdu. Over a 4-year period (2007–2010), STARTALK funded hundreds of institutions, including both public and private universities, high schools, and heritage schools, with thousands of students and teachers.

The purpose of this article is to describe the Summer Mandarin Chinese Language and Culture Program, a STARTALK program located at a large suburban public high school in Northern California, and to evaluate the effectiveness of the program based on current information on “best practices” used in intensive language program models. This summer program has (as of this writing) completed three continuous years with a joint collaboration among the high school, the school of education at an adjacent university, and the California Foreign Language Project. In

2009, this program was selected by the STARTALK center as an exemplary program and asked to share its practices and lessons. To be more specific, this article will provide a description of the program goals, curriculum, implementation, and results of student learning followed by a discussion of the challenges of implementing a summer intensive Mandarin program and suggestions for overcoming some of the challenges.

Literature Review of Intensive Language Programs

Definition of Intensive

Although the STARTALK Mandarin program is a recent addition to existing intensive language programs, intensive learning programs have existed for about a century (Serrano, 2007). The term *intensive* has several connotations. From the perspective of time distribution, intensity refers to time concentration in which instructional time is significantly extended every day and is condensed over a period of time. According to the history and origin of intensive language programs, intensity indicates an “intensive method” in which mimicry and repetition were heavily adopted in such courses in the 1940s and 1950s (Serrano, 2007). With the development of second language (L2) learning methodology and pedagogy, current intensive language programs no longer emphasize mimicry or repetition, but rather aural-oral skills, as well as “a determination on both teacher and student on the acquisition of speaking fluency, again irrespective of time distribution” (McMullen as cited in Serrano, 2007, p. 67). The appeal of current intensive language programs to the American public is the concept of time concentration and the promise that one can acquire a certain level of competence in a foreign language in a minimal amount of time.

Effects of Intensive Language Programs

Studies of the effects of intensive language programs in general have been widely

conducted in two main research areas: one within cognitive and educational psychology and the other within language program evaluation (Collins, Halter, Lightbown, & Spada, 1999). Findings regarding the effects of intensive learning programs have been controversial. Within the psychological literature, research findings have tended to support distributed language programs rather than intensive language programs. Dempster (1996) concluded that the superior effects of distributed practice over intensive practice are “among the most dependable, robust, and ubiquitous phenomena in the entire psychological literature” (p. 338). However, Collins et al. (1999) pointed out that the implications of these findings for language learning are “not readily apparent,” because the learning targets in the psychological studies are mostly “discrete items such as nonsense syllables, uncommon or specialized vocabulary and mathematical operations” (p. 657). None of the learning conditions in the psychological studies were analogous to real classroom language learning settings in which the learning targets were presented in different rather than repeated contexts.

Within the language program evaluation literature, findings have often supported intensive language programs. For example, Spada and Lightbown (1989) found that grade 9 and 10 students in an intensive ESL program achieved comparable results on both oral production and listening/reading comprehension as compared to their peers enrolled in the regular distributed program. In addition, they also observed some beneficial effects of the intensive ESL program on students, including greater autonomy, greater degree of self-esteem, and great cooperation and responsibility for their own learning (Spada & Lightbown, 1989). In a program of learning French as a second language, Wesche, MacFarlane, and Peters (1994) reported positive attitudes and gains of grade 5 and 6 students in oral proficiency. Likewise, Hinger’s study (2006) of Spanish as a foreign language postulated that concentrating classroom time enhanced

group cohesion, leading to a remarkable variety of group-building utterances, which maximized students’ language production and acquisition.

Successful Models for Intensive Language Programs

Although the above findings show that intensive language programs can be effective across languages, the advantages do not come effortlessly. Previous literature has discussed successful models for intensive language programs. In a successful intensive language program, both the instructor and the students need to make a commitment to work harder than under the traditional schedule, and to be enthusiastic about teaching and learning, respectively (Benseler & Schulz, 1979). Benseler and Schulz (1979) described the following aspects as necessary for the successful implementation of an intensive program:

- Extended daily exposure to the language (from 2 to 8 hours a day);
- Larger number of classroom contact hours (increases of up to 75% in instructional time);
- Usually small classes, 10 students maximum (in big classes practice is promoted in small groups);
- Focus on oral/aural skills and spoken language, but also some attention to the grammar necessary for clear communication in the L2;
- Frequent use of the language laboratory for practice;
- Extracurricular activities in the target language; and
- Staff willing to devote more time and energy than what is required for regular courses.

Similarly, Imel (2002) pointed out some of pedagogical strategies that have been frequently employed in successful intensive language programs, including nonthreatening teaching-learning environments, instructors and learners as equals in

collaborative environments, the use of small groups, and the importance of learners' experience as a resource (Imel, 2002).

Regarding the implementation of intensive language programs, research has also found that language teaching methodology and pedagogy can vary across different language programs. Germain, Lightbown, Netten, and Spada (2004) compared intensive English (IE) and intensive French (IF) programs in Canada in terms of the total time allotted to intensive periods, the selection of students, the models of delivery, and learning outcomes. In spite of the similar features that the IE and IF programs shared, there were many differences between the IE and IF programs including the compacting of the regular curriculum, the role of reading and writing, the emphasis on accuracy and fluency, and the use of cognitively demanding tasks.

With regard to models of intensive Chinese programs, successful models have not yet been validated. Actually, few studies have been conducted specifically targeting this critical language in the United States. Most programs that are described in the literature pertain to Chinese study-abroad programs, rather than intensive classroom-type language programs in the United States (Tseng, 2006; Zhang & Yu, 2008). In the study-abroad Chinese programs, students usually have ample opportunities to converse with native speakers in authentic, real-life situations (Tseng, 2006; Zhang & Yu, 2008). This is a programmatic feature that is extremely difficult to embed into an intensive language program and makes the comparison of the two program types very difficult. Therefore, it is critical to discuss the important elements of a Chinese intensive program whose effectiveness and success can only be validated by pre- and post-program comparisons.

Limitations of Previous Literature

Several limitations exist in the literature on intensive language programs. Most of the studies have been descriptive. For those

studies that took an empirical approach, many had methodological flaws (e.g., small number of participants, lack of detailed descriptions of instruments used to collect data, and data analytic limitations). Furthermore, as aforementioned, few studies have been conducted specifically targeting an intensive Chinese program.

This study overcomes the limitations of previous studies by adopting an empirical approach with both pre- and posttest data collected over three years. In addition, a detailed program description and qualitative data are shared for readers who are interested in replicating the intensive language program model or looking at a detailed map of the program design and effectiveness.

Methods

Both quantitative and qualitative measures were employed to gather information regarding the results and effectiveness of the STARTALK intensive Chinese Language and Culture Program implemented over three consecutive summers (2008, 2009, and 2010). The qualitative data included classroom observations and field notes, formal and informal interviews with instructors and students, and students' journals. The primary author served as a classroom observer for the program during all three summers. She sat at the back of the classroom each day, observing and taking detailed notes of instruction and student interactions in the classroom. Interviews with students and instructors took place during break and lunchtime and focused on student learning experiences and teaching practices respectively. The observation and interview data were then cleaned and coded for the analysis of important features of the program and students' feedback and motivation. The quantitative data included student performance on the Standards-based Assessment & Measurement of Proficiency (STAMP) in Mandarin and the Simulated Oral Proficiency Interview (SOPI).¹ Pre- and posttest results of STAMP and SOPI were collected to examine

students' progress over the 4-week intensive program. The qualitative and quantitative inquiries, though conducted separately, are linked by a common concern, namely to understand the program model and outcomes. In this sense, the two methods are complementary and together provided a rich source of data to examine the program outcome.

Participants

The 2008–2010 STARTALK program at a large suburban public high school consisted of 93 participants in total, including 31 students enrolled in the 2008 program, 34 students in the 2009 program, and 28 students in the 2010 program. Two levels (Level II and Level III²) of Mandarin instruction were offered in the 2008 and 2009 programs, while only level III was offered in the 2010 program. Students were mostly recruited from the school district, with a few others recruited from neighboring high schools. Table 1 shows the distribution of students across levels and years.

Measurement

The STAMP and SOPI tests were the two measurement tools used to assess students' pre- and posttest performance. The STAMP pre- and posttest are a set of computerized standardized assessments developed according to the guidelines of the ACTFL that measure students' reading, writing, and

speaking language proficiency in Chinese. Based on ACTFL levels of proficiency from Novice–Low through Intermediate–Mid, STAMP administers level grades (from Level 1 to Level 6³) to test takers. Students must demonstrate significant improvement in the specific language skills in order to move from one level to the next. Unlike the STAMP test, the SOPI⁴ only measures test takers' oral language proficiency. The SOPI items were also developed according to the ACTFL speaking proficiency guidelines. The test, approximately 50 minutes long, consists of warm-up interview questions and topic-based and situation-based tasks, and is presented to examinees via a test booklet and a master tape. Because the scoring rubric of SOPI does not distinguish Novice–Low, Mid, and High levels, most of the students remained in the Novice category after the 4-weeks program. In order to overcome this deficiency, we employed the Foreign Language Oral Skills Evaluation Matrix (FLOSEM) to measure the students' oral language performance using their SOPI oral interviews. The FLOSEM uses a 1- to 6-point scale to measure language proficiency, ranging from "extremely limited ability" (Level 1) through "native-like ability" (Level 6) for five language skills: comprehension, fluency, vocabulary, pronunciation, and grammar. The total possible points for SOPI tapes using the FLOSEM scoring procedure are 30 points (Padilla & Sung, 1999). Two native adult Mandarin speakers received training on the FLOSEM

TABLE 1

Number of Participants in 2008–2010 STARTALK Programs

	Level II participants	Level III participants	Total participants
2008 STARTALK	18	13	31
2009 STARTALK	22	12	34
2010 STARTALK	n/a	28	28
2008–2010 Total	40	53	93

scoring rubric and practiced listening to SOPI tapes and scoring the recordings with the FLOSEM. After the raters had attained a high level of agreement, they rated all student SOPI tapes.

In 2008, SOPI was administered to Level II students only and the STAMP was administered to Level III students only. Fifteen out of 18 Level II students took the SOPI test, and 10 out of 13 Level III students took the STAMP test. In 2009 and 2010, both SOPI and STAMP tests were given to Level II and Level III students. In 2009, 29 out of 34 students had both pre- and posttest scores for SOPI, and 32 out of 34 students had both pre- and posttest scores for STAMP. In 2010, only Level III students were recruited for the program. Twenty-four out of 28 students had both pre- and posttest scores for SOPI, and 27 out of 28 students had both pre- and posttest scores for STAMP.

Program Model

Program Overview

The STARTALK program at the high school was developed collaboratively by staff of the California Foreign Language Project, the Palo Alto Unified School District, and the Stanford University School of Education. The nonresidential program provided an intensive 4-week (5 days a week, 5 hours per day) language learning experience for high school students who wanted to accelerate their acquisition of Mandarin in order to enroll and succeed in AP Mandarin by their senior year. The program hired two instructors to co-teach each class. In each classroom, there was one experienced instructor paired with a novice instructor. In addition to learning Chinese, students also participated in a wide range of cultural experiences, including morning exercises of *Taiji*. In class, instructors engaged students in Chinese rituals and customs and sought to contextualize every language learning experience in authentic cultural settings. In the computer lab, students developed and sent each other electronic cards of Chinese festivals such as *Duanwu Jie* (The Dragon

Boat Festival) and *Zhongqiu Jie* (The Mid-Autumn Festival) and researched the different activities that Chinese families and communities experience during these traditional summer festivals. Field trips were also arranged so that students could experience authentic aspects of Chinese culture while having the opportunity to use Mandarin in real-life contexts.

Program Goals

The program established two goals for students in Level II and Level III courses:

1. Sponsor an accelerated 4-week intensive program for students integrated with their academic year course of study that enabled students to show measurable gains in all four Mandarin language skills.
2. Enhance students' knowledge of Chinese culture by integrating classroom, field trip, and internship activities in the summer program that capitalized on the rich, authentic use of Mandarin, which would be reflected in students' qualitative evaluation of the program.

Curriculum

A theme-based curriculum was developed for the program that followed the linguistic and cultural domains addressed during the regular academic year and that provided students with a deeper and more engaging learning experience. The curriculum theme for the Level II course was "My Extra-Curricular Life," and topics included "My Birthday," "Shopping," and "Visiting a Friend." The theme for Level III was "Planning a Trip to China," and topics included "Geography of China," "Key Monuments," "Tourist Attractions and Famous Scenic Spots," "Planning a Travel Itinerary," "Visiting a Chinese Family," and "Appropriate Visitor Etiquette."

Co-curricular Activities

The program's curriculum was not limited to formal instruction in the classroom but

extended to technological learning experiences and activities outside the classroom. Within the 5-hour school day, students had a 50-minute computer lab session in which they practiced typing Chinese characters, sentences, and passages; writing e-mails to each other; and taking timed-response quizzes.

Field trips were another important feature of the program. Students went to the Stanford University Museum to view the Asian art collection, where students appreciated authentic Chinese art (e.g., Chinese calligraphy, Chinese brush paintings, Chinese embroidery) from different Chinese dynasties and historical periods. During this field trip, students completed an instructor-designed worksheet that required students to select a favorite art piece and to write a short paragraph demonstrating their understanding of it. Students also went on a field trip to a nearby shopping mall, where they shopped at Chinese stores and had lunch at a traditional Chinese restaurant. The worksheet required students to interact with shop owners while looking for or purchasing goods. Students were also required to select their favorite dish in the restaurant and write down its key ingredients. The final field trip destination was to San Francisco's Chinatown. Here students listened attentively to the tour guide's explanation of Chinatown's history, sampled tea in a Chinese tea shop, shopped in Chinese markets and stores, and visited a Chinese temple. Similar to the previous trips, students were required to respond in Chinese to written questions about the history of Chinatown, their shopping experiences, and Chinese religion.

During summer 2009, a substantial portion of the program for Level III students included internship experiences in Mandarin-speaking environments for approximately 20% of the 100-hour total program. Chinese native-speaking agencies in the neighboring communities were recruited to be internship partners of the program. The internship settings were all located in Chinese communities and included a Chinese learning center for preschoolers, a Chinese real estate office, a

Zen center, and a bookstore that served a Chinese clientele. Students were required to communicate in Mandarin with both the employees and clients (including children in the learning center), to interact with them in culturally appropriate ways, and to write journal entries after each of their internship experiences.

Instruction

The program adopted a Standards-based/communicative language teaching method. The instructors included interesting indoor and outdoor activities (e.g., *Taiji* exercises) between instructional periods. The structure and organization of the class provided for a stimulating learning environment. Instructors adopted a motivational reward system that further engaged students. If a student successfully responded to a question, volunteered to participate in a learning activity as a group member, or performed a communicative task with a high level of proficiency, he or she received one *Xinghua tongpiao* (STAR-TALK money) that was factored into the student's daily performance mark for active participation. In addition, instead of being performance monitors, the instructors appointed two *Zhirisheng* (students-on-duty) every day who monitored the class "lottery" process for selecting the order in which students would be asked to lead an activity, respond to a prompt, participate in a learning group; noted students' participation and completion of assignments; and recorded the *Xinghua tongpiao* earned by the students. The program hired two instructors, one experienced and one novice, to co-teach each class. The instructors formed a strong collaborative teaching team and alternated working with both Level II and Level III students. This lowered the instructor-student ratio, which made it possible for students to get more individual and small-group attention. Both instructors were always present in the classroom. If one was providing direct instruction, the other was moving around the classroom offering clarification, assistance, and encouragement.

Technology

In the STARTALK program, students also employed technology to enhance their language and culture learning experiences. The school's World Language Computer Lab served as the students' second classroom. Students carried out online activities such as sending electronic Chinese greeting cards to each other, learning to type short Chinese paragraphs, and completing mandated STARTALK assessments. In addition to the computer lab, students also had free access to iPods equipped with iTalk microphones for the duration of the program. The program director contacted the education branch of Apple, Inc., and the company kindly provided free iPods for students to use during the program. A majority of homework assignments were on the iPods, and students listened to native speakers' dialogues and responded to appropriate prompts by recording their own voices using the iTalk microphones and creating their own individual podcasts.

Assessment

Throughout the program, the instructors used a variety of assessment tools to assess student performance. Comprehension checks were used often to ensure that students understood the different levels of input provided by the instructors, multimedia, and guest speakers. Other types of formative assessments consisted of responding to oral, visual, and written prompts; completing different communicative tasks; timed response quizzes; and descriptive audio recordings of different cultural visuals. Summative assessments included pre- and posttests using the SOPI and STAMP. The results of these summative assessments are discussed in the Results section.

Results

Analysis of the Quantitative Data

SOPI Results

To investigate the effects of the intensive summer Mandarin program on students'

language achievement as measured by the SOPI test, a two (pretest and posttest) by two (Level II and Level III) repeated-measures analysis was conducted. As previously mentioned in the Method section, we used the FLOSEM evaluation matrix to score students' SOPI tapes (Padilla & Sung, 1999). Raters were not told under what conditions the SOPI was administered to the students in the intensive summer program.

Pre- and posttest comparisons. Results showed that students improved significantly from the pretest to the posttest on the SOPI/FLOSEM ($F(1,66) = 46.21, p < 0.001$). The mean of the SOPI/FLOSEM pretest was 15.72, and the posttest SOP/FLOSEM mean was 18.57.

Analysis of each individual language skill measured by the FLOSEM evaluation tool—comprehension, fluency, vocabulary, pronunciation, and grammar—demonstrated that students improved significantly from the pretest to the posttest on every skill. Table 2 shows pretest and posttest scores and the statistical results.

Level of Mandarin instruction. Results of the repeated-measures test showed that level of instruction was not a significant factor in the different language skills measured by the FLOSEM. Results also indicated no interaction between level of instruction and the pre- and posttests for any of the language indexes measured by the FLOSEM. This indicates that the STARTALK intensive program was equally effective for Level II and Level III students in improving their Mandarin language skills (see Table 3).

The absence of a significant interaction between level and pre-posttest was found in every skill measured by the FLOSEM on the SOPI test except for the fluency skill. Although the interactions failed to achieve statistical significance, there was a trend in the expected direction ($F(1,66) = 2.97, p = 0.089$) for fluency. This result indicated that, during the program, Level II students improved more in their fluency skill than did Level III students (see Figure 1).

TABLE 2

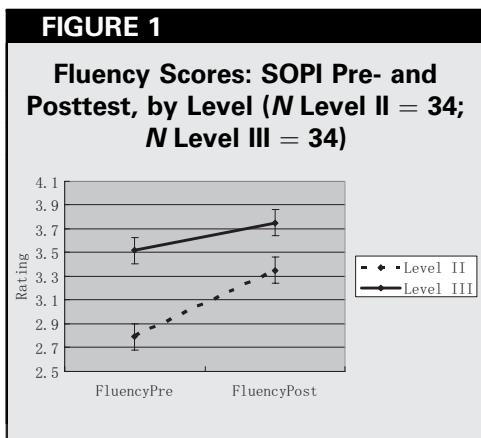
Means (Standard Deviations) of SOPI Pre- and Posttest (N = 68)

	Comprehension	Fluency	Vocabulary	Pronunciation	Grammar	Total
Pretest	3.68 (1.44)	3.07 (1.50)	2.89 (1.52)	3.26 (1.38)	2.82 (1.52)	15.72 (7.19)
Posttest	4.26 (1.32)	3.63 (1.30)	3.54 (1.29)	3.82 (1.25)	3.32 (1.40)	18.57 (6.35)
F ratio	F(1,66) = 18.94	F(1,66) = 35.50	F(1,66) = 47.47	F(1,66) = 45.02	F(1,66) = 35.86	F(1,66) = 46.21
p value	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001

TABLE 3

Means (Standard Deviations) of SOPI Pre- and Posttest, by Treatment Level (N level II = 34; N level III = 34)

	Comprehension	Fluency	Vocabulary	Pronunciation	Grammar	Total	
Level II	Pre	3.43 (.24)	2.79 (0.26)	2.60 (0.26)	3.16 (0.24)	2.56 (0.26)	14.54 (1.23)
	Post	4.15 (.23)	3.52 (0.22)	3.37 (0.22)	3.77 (0.22)	3.09 (0.24)	17.88 (1.09)
Level III	Pre	3.94 (.24)	3.35 (0.26)	3.18 (0.26)	3.35 (0.24)	3.07 (0.26)	16.90 (1.23)
	Post	4.38 (.23)	3.75 (0.22)	3.71 (0.22)	3.87 (0.22)	3.56 (0.24)	19.27 (1.09)
F ratio and p value of interaction:	F(1,66)	F(1,66) = 2.97	F(1,66)	F(1,66)	F(1,66)	F(1,66)	F(1,66)
treatment level	= 1.10 NS	p = 0.089	= 1.57 NS	= 0.28 NS	= 0.07 NS	= 1.34 NS	



STAMP Test Results

Because the STAMP test only offers categorical scores, a Wilcoxon signed-rank test analyzing the nonparametric paired-sample (dependent) comparison of STAMP pre- and posttests was computed. Results showed that students improved significantly from the pretest to the posttest on the STAMP test score ($z = 3.61, p < 0.01$). The mean of the STAMP pretest was 8.85, and the mean of the STAMP posttest was 9.65.

Analysis of each individual language skill measured (reading, writing, speaking) on the STAMP test demonstrated that students improved significantly from the pretest to the posttest on every skill (see Table 4).⁵

Level of Mandarin instruction. Similar to the analysis of SOPI results, to investigate the impact of level on student performance and interaction between level and pre/posttest performance, a two (pretest and posttest) by two (Level II and Level III) repeated-

measures analysis of the STAMP test was conducted.

Results showed that level was a significant factor for writing ($F(1,67) = 12.06, p < 0.01$), speaking ($F(1,60) = 6.33, p < 0.05$), and STAMP total rating ($F(1,67) = 6.48, p < 0.05$), but not reading ($F(1,67) = .36, p > 0.05$). This demonstrated that Level III students performed significantly better than Level II students in terms of writing and speaking skills, but not in reading Mandarin. Findings also indicated no interaction between level and pre-/posttest performance. This indicates that the STARTALK intensive program was equally effective for Level II and Level III students in improving their Mandarin language skills (see Table 5).

The absence of a significant interaction between level and pre-posttest was found for every skill measured by the STAMP test; however, there was a trend in the predicated direction ($F(1,60) = 3.14, p = 0.081$) on the speaking subtest. This finding indicated that Level II students improved more in speaking than Level III students. This is consistent with the findings on the SOPI test that showed that Level II students improved more on fluency in Mandarin, which falls into the speaking domain (see Figure 2).

Analysis of the Qualitative Data

The quantitative measurements demonstrated students' steady improvement in Mandarin acquisition. An important question is why this intensive summer program was effective. To begin to understand this, it is important to examine the unique features of

TABLE 4

Means (Standard Deviations) of STAMP Pre- and Posttest (*N* = 69)

	Reading	Writing	Speaking	Total
Pretest	1.90 (1.09)	3.49 (.97)	3.34 (.87)	8.85 (2.38)
Posttest	2.35 (1.26)	3.71 (.86)	3.54 (.95)	9.65 (2.54)
Z-score	Z = 3.52	Z = 2.58	Z = 2.17	Z = 3.61
p-value	p < 0.001	p < 0.05	p < 0.05	p < 0.001

TABLE 5

Means (Standard Deviations) of Pre- and Posttest STAMP scores by Level of Instruction (N Level II = 19; N Level III = 50)

		Reading	Writing	Speaking	Total
Level II	Pre	1.85 (0.25)	3.00 (0.21)	2.85 (0.19)	7.74 (0.54)
	Post	2.15 (0.28)	3.17 (0.18)	3.29 (0.23)	8.65 (0.59)
Level III	Pre	1.92 (0.16)	3.69 (0.13)	3.60 (0.12)	9.44 (0.34)
	Post	2.43 (0.18)	3.94 (0.11)	3.68 (0.14)	10.17 (0.38)
<i>F</i> ratio and <i>p</i> -value of interaction: treatment level		$F(1,67) = 0.72$ NS	$F(1,67) = 0.22$ NS	$F(1,60) = 3.14$ $p = 0.081$	$F(1,60) = 0.17$ NS

the program and couple this with interview data from instructors and students.

Remember that the features of the intensive summer Mandarin program included co-teaching, a variety of classroom activities, technological support, field trips, and internships. How did these program features promote students' language acquisition? And how were these elements implemented in the intensive language program? The following qualitative data, drawn from field notes and student and instructor interviews, reveal how these features contributed to the instructional program and to learning.

Co-teaching

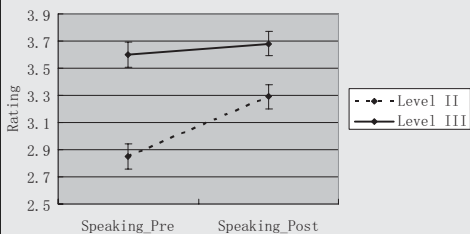
Co-teaching is two or more instructors sharing responsibility for planning, instruction, and evaluation for students. Key elements of co-teaching include (1) having

common, publicly agreed-on goals; (2) sharing a belief system; (3) demonstrating parity by alternatively engaging in the dual roles of instructor and learner, expert and novice; (4) using a distributed functions theory of leadership; and (5) using a cooperative process (Cushman, 2004). In this STARTALK program, instructors discussed and shared common goals and shifted leadership roles through a cooperative program. Co-teaching and instructor collaboration also lessened the stress of teaching for a 5-hour stretch of time. In addition, it lowered the teacher-student ratio to one to eight, which gave students more individual and small group attention and was more consistent with the successful models of intensive language program (Benseler & Schulz, 1979; Imel, 2002). While one instructor was engaged in direct instruction, the other was moving around the classroom offering assistance and encouragement. Mr. M., the most experienced instructor in the program, commented about the co-teaching:

The STARTALK language program is different from the language programs in the regular school year. It is very fast-paced and intensive. I would not have been able to plan and do all these things without the help of the other teachers. Also, the selection of the team members was wonderful. It was not random. Ms. V. and I had collaborated in last year's STARTALK program. She was very

FIGURE 2

Speaking: STAMP Pre- and Posttest, by Level (N Level II = 19; N Level III = 50)



familiar with the program model, and we collaborated in drafting the lesson plans. Ms. S. and Ms. J. were both my student teachers. They were very familiar with the students and did a good job in teaching. ... We've had a very good collaboration and the teaching just flows!

Students also liked the mode of co-teaching and thought it was beneficial to their learning. One student mentioned:

The teachers are great! They all have different personalities and teaching styles. Mr. M. is like a magician. He can always produce various toys and classroom activities. ... Ms. Z is strict and really serious about learning. She works extremely hard to improve our pronunciation and character writing. ... Ms. J really knows how to motivate students. She always gives positive feedback, and we are never afraid of making mistakes in front of her. ... Ms. S is like our friend or an older sister. She not only teaches us the language, but also cares about us.

Thus, the team teaching approach is considered critical in an intensive and accelerated Mandarin program and is recommended to anyone planning an intense multi-hour instructional program.

A Variety of Classroom Activities

As in other communicative language teaching classrooms, instructors in our program relied on a variety of activities in which to embed instruction. Students were highly motivated by their activity-based instruction. "Mr. M. makes the class so exciting because there is always something going on," said one of his students in an interview.

Regarding the reward system *Xinghua tongpiao* and the *Zhirisheng* (students-on-duty), Mr. M. suggested, "The idea of *Zhirisheng* not only reduced instructors' workload, but also motivated students by giving them power in the classroom and increased their involvement."

Students also mentioned their favorite classroom activities: Bingo, jigsaw puzzles, singing, role-playing, and group competition in writing Chinese characters. These activities enhanced students' motivation to use the target language while interacting with their instructors and peers, and lowered their affective filter (Krashen, 1982). These entertaining but meaningful language activities also allowed for frequent negotiation of meaning and maximized students' target language output.

Technological Support

Students also used technology to enhance their language and cultural learning experiences. For example, the podcasts greatly enhanced the typical 50-minute classroom language learning experience for the students. Because students had free access to iPods, after-class assignments included listening to native speakers' dialogues on their iPods and practicing the dialogues by themselves. In the post-program survey, students uniformly agreed that the use of the iPod was an important and motivating part of their language instruction. One student commented about the role of technology in the program:

It is sometimes painful to be in the classroom for a long time. The lab is a great place to continue our learning since it gives us a chance to be in a different place than always in the classroom.

Another student stated:

Having the iPods with all of the podcasts on them was really helpful, and even though lots of the dialogue we had to listen to was pretty easy, we really learned that stuff thoroughly.

In terms of the integration of technology into language instruction, Mr. M. also commented about the benefits of technology and emphasized the selection of appropriate content:

Students at this age love the iPod! It makes learning more fun and relaxing. But we teachers should be extremely careful in selecting materials. The dialogues we choose to upload into the iPod should be appropriate for students' proficiency level and relevant to their lives, such as the topics "Making Friends" and "Keep in Touch," "Going Shopping and Buying a Souvenir,"⁶ etc. We also introduced and played the movie "Riding Alone for Thousands of Miles." It was not just about the language, but also contained many cultural elements such as the relationship of Chinese family members, the beauty of Chinese local operas, and the complicated Chinese bureaucratic procedures. These elements will be very useful if the students go to China one day on their own.

Computer and iPod usage interspersed daily during the 5-hour learning session provided students with breaks and a different type of language learning exposure. The podcasts exposed students to a variety of authentic language topics as well as different native accents. The e-mail interactions and lab activities facilitated students' negotiation of meaning through online communities. Furthermore, movies immersed students in authentic cultural experiences and deepened their understanding of East-West cultural differences.

Field Trips

The opportunity to participate in relevant field trips also extended the language learning experience for students beyond the classroom and with real-life language situations that enabled them to use what they learned in authentic cultural settings. Field trips were an important feature of the summer program.

These trips were a great hit with students. In the post-program survey, all the students agreed, "The field trips we took were helpful for my language learning." For instance, one student stated:

When school starts in the fall, I'll really miss all of the field trips, because it was really fun and we all learned a lot about Chinese culture. Plus, it was fun to get out of the classroom, shop, and eat.

Another student commented regarding field trips:

The best thing I like about the program is the field trip. We got to eat authentic Chinese food, shop in the Chinese grocery store, and meet Chinese people. It was very cool! I just hoped I could speak better Chinese so that I could talk more to them.

The field trips, while enjoyable, were an extension of the classroom because the instructors taught language and culture wherever they were and encouraged students to use their Chinese in ordering food at restaurants, as well as speaking with local businesspeople when they entered local shops and browsed for items of interest. During the language program, field trips allowed both students and instructors to get away from the classroom and to focus on situationally specific language use. In addition, students had field trip assignments to complete. According to Mr. M.,

Students need to complete a sheet on which questions are asked about possible language and cultural elements of the field trips. They play, and at the same time, learn the language and culture. The field trip to the museum enriched students' cultural and historical knowledge about China; the field trips to Cupertino Chinese dining and shopping center and San Francisco's Chinatown immersed them in an absolute Chinese environment that they could not experience in the classroom.

Internships

The internship component of the program for Level II students was also very unique. Seven internship sites where Mandarin was

used served as the base for placing students in locations where they could hear and use Mandarin in an authentic context. Students were initially challenged by their internship but came to enjoy their placements and felt comfortable with a sense of pride in their “part-time job.” Here are excerpts from two students’ journals:

In the bookstore, the boss asked me to greet every customer. It was boring at the very beginning. But then, the customers found I could speak more Chinese than *Nihao* (hello) and became very interested in me. They talked to me in Chinese, asked which school I was from and why I was there. They said my Chinese was very good.

Today I interned at Sunflower Learning Center. I was nervous and excited. Me (I) and my partner M. walked into the learning center. In the beginning, the kids were “mean” to us, but with a little help from our Chinese skills, we were able to make some new friends. Later in the day, M. and I split up to help different teachers. I helped the teacher by passing out papers and grading them. I felt very happy at the end.

The gap between textbook acquisition of Mandarin and authentic real-life use of Mandarin was reduced while also increasing students’ motivation to learn Mandarin. From a student’s journal:

We were introduced to a lady who would be helping us. She gave us forms to fill out and asked us a lot of questions in Chinese. We talked mostly in Chinese. She spoke a lot faster than our teachers and we had to ask her to repeat a lot. We had a lot of Chinese today and I thought it was very educational.

Discussion

From the quantitative data, we conclude that this intensive summer language program for high school students was rigorous

and effective because students’ test scores showed significant improvement between pre- and posttests on the outcome measures for both the Level II and Level III students. There were mixed findings about level of instruction on some language skills measured by the FLOSEM and the STAMP. For instance, analysis of the interaction between level of instruction and pre-posttest comparisons indicated that Level II students improved more in speaking (especially in fluency) than Level III students. Nonetheless, the absence of interaction effects between level of instruction and the different language outcome measures supports the conclusion that the program was equally effective for lower-level (Level II) and higher-level (Level III) students, backing previous studies (e.g., Spada & Lightbown, 1989) that reported similar findings.

In addition, the qualitative data also showed students’ learning of Mandarin language skills as well as their increased knowledge of Chinese culture. Students liked the co-teaching model because instructors possessed different talents and teaching techniques that complemented and supported each other in constructing an engaging language learning environment for students. In addition, students also favored field trips and internships, the unique features of the program, through which they could translate what they had learned in the classroom to Mandarin usage in authentic cultural settings.

The STARTALK Mandarin high school program described here can serve as a model for an intensive world language program for educators interested in launching or improving a similar language program. The highlights and successful features of the program were consistent with the key elements identified as necessary for the successful implementation of an intensive program (Benseler & Schulz, 1979): for example, extended daily exposure to the language, small classes, focus on oral/aural skills, frequent use of language lab, and extracurricular activities. A few key program features that serve as suggestions for the

development of future intensive language programs are as follows: (1) in the classroom, co-teaching lowers the student-to-instructor ratio and permits more time for individual student instruction; (2) the activity-based instruction and communicative language teaching enhances students' attention and keeps them highly motivated; (3) the incorporation of technology into language instruction augments instructional time and content and complements the traditional language instruction offered by instructors; (4) more informal learning opportunities such as field trips and internships serve to connect students to authentic language and cultural usage; and (5) from a pedagogical perspective, intensive language programs need to be carefully planned and implemented. The success of an intensive program requires well-trained instructors who are flexible enough to modify their instruction (e.g., designing a variety of classroom activities and adding co-curricular activities such as lab sessions, field trips, and internships) to the challenges posed by teaching adolescents in a program that runs 5 hours per day for 4 weeks.

Despite the successful features of the STARTALK Mandarin program, as with all experimental programs, there were a few glitches that created challenges for the instructors that need to be addressed for planning and implementing future intensive summer Mandarin programs. For example, instructors need a Mandarin proficiency assessment tool that is easy to administer and score. The STAMP assessment that students took on a pre- and posttest basis was time-consuming and limited in its range of language proficiencies. The SOPI test also proved to be a very time-intensive instrument, and the scoring rubric does not distinguish well between Novice-Low, Novice-Mid, and Novice-High learners. Another challenge has to do with suitable curriculum materials for use with intermediate-level students. The textbook series *Nihao* adopted in the program appeared to work best for lower-level students, but was not as suitable for Level III students at the intermediate

level. A major problem with the book was that some lessons were too long and uninspiring. Mandarin instructors need to compare and carefully select an appropriate textbook for higher-level students. This is particularly important in intensive language programs because the materials must be pedagogically sound as well as inherently interesting to students.

Finally, it is important to recognize that the students in the program were self-selected into the program and very motivated to succeed; they may have done well no matter how the program was designed and implemented. Another limitation of the study is the lack of a comparison group. While we were able to demonstrate that students improved their language skills significantly through a 4-week intensive Chinese program, we could not address how effective the program was compared to a regular distributed classroom program that runs for a semester. Future randomized studies can be conducted to determine which program, intensive or distributed, is more effective and beneficial for foreign language learners.

Acknowledgments

The authors gratefully acknowledge the support of the STARTALK project, funded by the National Security Agency via the National Foreign Language Center. In addition, the authors are grateful for the contributions of our partnership district, the Palo Alto Unified School District, the instructors who taught in the project, the students who actively engaged in the learning of Chinese language and culture, and their parents who enthusiastically supported our program implementation and evaluation efforts. The authors also wish to thank the reviewers whose comments strengthened our article.

Notes

1. A detailed description of the STAMP and SOPI measurements is provided in the Measurement section.

2. Level II students had completed 1 year of high school Mandarin instruction, and Level III students had completed 2 years of high school Mandarin instruction.
3. See <http://www.stamptest.net/stamp0708/stamptest/>. With regard to the results of the STAMP test, level 1 represents Novice-Low, level 2 represents Novice-Mid, level 3 represents Novice-High, level 4 represents Intermediate-Low, level 5 represents Intermediate-Mid, and level 6 represents Pre-Advanced.
4. See <http://www.cal.org/topics/ta/sopi.html>.
5. Among the 69 STAMP test-takers, seven of them did not take the STAMP speaking test due to time and technological issues. They all took STAMP reading and writing tests.
6. See <http://www.podcast.com>.

References

- Asia Society and The College Board. (2008). *Chinese in 2008: An expanding field*. Retrieved January 20, 2008, from <http://www.askasia.org/chinese/publications.htm>
- Benseler, D., & Schulz, R. A. (1979). Intensive foreign language courses. *Language In Education: Theory and Practice*, 18, 1–53.
- Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, 33, 655–680.
- Cushman, S. (2004). *What is co-teaching?* Retrieved May 1, 2012, from http://www.sagepub.com/upm-data/6847_villa_ch_1.pdf
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 317–344). San Diego: Academic Press.
- 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
- Germain, C., Lightbown, P., Netten, J., & Spada, N. (2004). Intensive French and intensive English: Similarities and differences. *Canadian Modern Language Review*, 60, 409–430.
- Retrieved January 8, 2009, from http://www.mmecarr.ca/ICF/ICF_PDFs/IFrandIEngl.pdf
- Hinger, B. (2006). The distribution of instructional time and its effect on group cohesion in the foreign language classroom: A comparison of intensive and standard format courses. *System: An International Journal of Educational Technology and Applied Linguistics*, 34, 97–118.
- Imel, S. (2002). *Accelerated learning in adult education and training and development. Trends and Issues Alert*. ERIC Clearinghouse on Adult, Career, and Vocational Education, Columbus, OH. Retrieved February 1, 2009, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/ce/ac.pdf
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Lum, R. (2008, March-April). Ko K'un-hua: Brief life of Harvard's first Chinese instructor: 1838–1882. *Harvard Magazine* [Electronic version]. Retrieved December 1, 2008, from <http://harvardmagazine.com/2008/03/ko-kun-hua.html>
- 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>
- Padilla, A. M., & Sung, H. (1999). The Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM): A rating scale for assessing communicative proficiency. Retrieved January 31, 2010, from <http://www.eric.ed.gov/PDFS/ED445538.pdf>
- Robelen, E. W. (2010, October 6). Chinese aid boosts Mandarin-language instruction. *U.S. Education Week* [Electronic version]. Retrieved January 2, 2011, from http://www.edweek.org/ew/articles/2010/10/06/06chinese_ep.h30.html
- Ruethling, G. (2005, October 15). Classes in Chinese grow as the language rides a wave of popularity. *The New York Times* [Electronic version]. Retrieved January 2, 2009, from http://www.nytimes.com/2005/10/15/national/15chinese.html?_r=2
- Serrano, R. S. (2007). *Time distribution and the acquisition of English as a foreign language* (Unpublished doctoral dissertation). Retrieved December 1, 2008, from http://www.tdx.cat/bitstream/handle/10803/1667/RSS_THESIS.pdf?sequence=1
- Spada, N., & Lightbown, P. M. (1989). Intensive ESL programmes in Quebec primary

schools. *TESL Canada Journal/Revue TESL du Canada*, 7, 11–32.

Tseng, M. (2006). Language gains in the study abroad and domestic contexts. *Journal of Chinese Language Teachers Association*, 41, 57–78.

Wesche, M., MacFarlane, A., & Peters, M. (1994). The “*bain linguistique*”: A core French experiment at Churchill Alternative School 1993–94. Ottawa: Canada University of Ottawa.

Xinhua. (2008, October 24). Reform, open-up have led China to sustainable development. *China Daily* [Electronic version]. Retrieved

January 10, 2009, from http://chinadaily.cn/china/2008-10/24/content_7135592.htm

Zhang, D., & Yu, Y. (2008). Context of learning and requesting in Chinese as a second language: An exploratory study of students learning Chinese in study abroad and at home contexts. *Journal of Chinese Language Teachers Association*, 43, 1–21.

Received April 19, 2011

Accepted October 29, 2012