

A Longitudinal Study of the Development of Mandarin Chinese Learners' Oral Fluency

Wo-Hsin CHU
Department of Chinese
as a Second Language
National Taiwan
Normal University

Chian-Wen KAO¹
Department of
Applied English
Chihlee University of
Technology

Alexander SOEMER
Department of
Psychology
University of
Potsdam

Abstract

We assessed the development of oral fluency of Mandarin Chinese learners from different language backgrounds enrolled in an undergraduate program of Chinese as a Second Language (CSL) between their first and third year of study. Analyses of speech data collected during the yearly examinations revealed large increases in articulation rates, filled pauses, repetitions, and false starts from the first to the second year as well as a further increase in articulation rates from the second to the third year. We also replicated typical relationships between CSL learners' speaking proficiency and fluency indicators. Contrary to previous research, we found that the articulation rates do not always serve as a significant fluency indicator to distinguish higher level learners from lower level learners. We will discuss explanations for this pattern related to contextual and educational factors.

Keywords: longitudinal study, second language acquisition, oral fluency

¹ Corresponding author

1. Introduction

Improvements in many of the basic types of proficiencies in second language acquisition –listening, speaking, reading, and writing – are difficult to measure because they tend to happen over a relatively long period of time. Although cross-sectional studies seem to provide a viable platform for investigating some of the aspects of proficiency development, it has repeatedly been pointed out that a full understanding of L2 learning requires longitudinal designs (Ortega and Iberri-Shea 2005; Derwing, Munro, and Thomson 2008; Ortega and Byrnes 2008). To date, longitudinal studies are still rare. Thus, research has made relatively slow progress towards an understanding of the development of L2 learners' proficiency (Ortega and Iberri-Shea 2005; Chu and Kao 2016), despite it being a very important factor in terms of university admission and employment decisions.

The present study was designed to address the lack of longitudinal evidence on oral proficiency in SLA research with a focus on an increasingly important target of L2 learning – Mandarin Chinese. We report analyses of speech data taken from 26 international students at different points of time during a four-year undergraduate program of Chinese as a Second Language (CSL). The main goal of the present research was to measure various aspects of oral fluency and derive a set of indicators that reliably reflect the development of oral fluency in Mandarin Chinese.

1.1 Investigations of Oral Proficiency and Fluency

Research usually highlights three aspects of oral linguistic behavior commonly believed to adequately reflect speakers' oral proficiency: *complexity*, *accuracy*, and *fluency* (CAF; Skehan 2009). The current study – although being part of a bigger project seeking to analyze all three aspects – reports a first set of analyses focusing on the development of oral fluency in L2 learners of Mandarin Chinese. From a speaker-based perspective, oral fluency often refers to the degree of automaticity in speech production reflected in time-related aspects of an oral production (Lennon 1990; Schmidt 1992; Chambers 1998; Kahng 2014). Within this framework, analyses of oral fluency usually focus on measures of production speed such as *words/syllables per time unit*, hesitations such as the *number/length of filled and unfilled pauses*, and repair phenomena such as the *frequency of repetitions, self-corrections*, and *false*

starts (Lennon 1990; Griffiths 1991; Freed et al. 2004; Segalowitz and Freed 2004).

There is a substantial body of research showing that such quantitative performance measures reliably account for evaluations of oral fluency by native listeners. For example, Lennon (1990) and Riggensbach (1991) explored the relationship between objective measures of oral production and fluency judgments of native speakers for non-native speech samples. Lennon (1990) reported that speed of production, fewer filled pauses per T-unit, and fewer T-units followed by a pause reliably predicted English native speakers' fluency judgments of German L2 learners' English productions. Similarly, Riggensbach (1991) found that a faster speech rate and fewer unfilled pauses contributed to English native speakers' fluency evaluations of speech samples produced by Chinese speakers. Neither study, however, found support for repair phenomena as a significant predictor of fluency evaluations (but see Bosker et al. 2013). In the case of Hungarian speakers, Kormos and Dénes (2004) found that speech rate, the mean length of an utterance, the percentage of pause time with regard to total time, and the number of stressed words per minute were the best predictors of fluency scores, whereas filled and unfilled pause frequencies did not seem to be significant predictors of perceived fluency.

Assuming that time-related measures of production adequately reflect the degree of oral fluency, the natural question arises as to how these measures change over time or with proficiency level. After a 2-year study period that included a study abroad phase, Mora and Valls-Ferrer (2012) found increases in production speed and number of words between pauses, as well as generally less hesitant speech for L2 English learners with Catalan-Spanish language background. Segalowitz and Freed (2004) investigated the development of native English speakers' L2 fluency in Spanish over one semester and similarly found greater production speed and less hesitant speech. Complementing studies on the changes of oral proficiency measures over time, there are a few other longitudinal studies investigating fluency development from a listener-based perspective. For example, Derwing, Munro, and Thomson (2008) compared oral fluency of adult ESL students from Mandarin and Slavic language backgrounds over a 2-year period. They found that the Slavic but not the Mandarin language speakers showed a significant improvement in perceived oral fluency, though both groups started at the same level of proficiency (the authors attributed this to different

degrees of exposure to English outside the ESL courses between the two groups).

To date, most available data on L2 oral proficiency development concerns English or other Indo-European languages. Notable examples are two cross-sectional studies by Iwashita et al. (2008) and Iwashita (2010). Iwashita et al. (2008) investigated the relationship between ESL learners' spoken performance and their detailed features of spoken English language. Although not focusing on oral fluency, they reported that higher level learners' speech rates were higher than lower level learners'. In addition, higher level learners' unfilled pauses were fewer than low level learners'. No significant difference was evident for filled pauses and repair in terms of different proficiency levels. The other study by Iwashita (2010), part of which compared oral fluency of English (EFL) and Japanese (JFL) learners between two proficiency levels, reported that both the high level EFL and JFL students' production speed was higher than the low level students' production speed. However, the same study showed that only high level EFL students but not high level JFL students made more frequent unfilled pauses and repairs. However, less is known about the development of oral fluency in Mandarin Chinese, a language whose importance has arguably grown over the recent decades. One rare exception known to the authors is a study by Li (2014) who examined the effects of different levels of Mandarin Chinese proficiency (intermediate/advanced) on the development of pragmatically appropriate requests in a sample of American learners. Although not particularly concerned with oral fluency, Li (2014) reported that advanced but not intermediate learners' production speed increased over a study abroad period of several months.

Theoretically, there are a number of particular characteristics of this language that might quantitatively and qualitatively influence the development of oral proficiency in L2 learners. A well-known stumbling block for L2 learners of Mandarin Chinese, for example, is the status of tone or accent as a lexical feature (identical syllables may refer to different words depending on the pitch pattern accompanying their production, for example, the syllable *mai* can mean either 'to buy' or 'to sell'). Proficient speakers of Mandarin Chinese not only know the difference between the tones produced in isolation, but they are also able to integrate them into the prosodic and phonological context of natural speech (Tseng, Soemer and Lee 2013). One might therefore expect that lexical tone and other Mandarin-specific phenomena could be a

particular source for disfluency leading to different patterns of fluency development in comparison to other languages (in particular, patterns of repairs and pausing). Another factor potentially leading to results incongruent with previous research concerns the fact that the development of oral fluency was investigated in a study abroad context in a non-Western country. For example, CSL learners might speak more fluently than Chinese as a Foreign Language (CFL) learners because CSL learners have more exposures to the target language in their daily life. In other words, differences in the results compared to research conducted with English or other Indo-European L2 learners might not (only) reflect specifics of Mandarin Chinese as a target language, but also the environment in which the language is studied.

To summarize, previous research strongly suggests that greater automaticity of oral production, reflected in time-related measures such production speed, hesitations and repairs per time contribute to listeners' evaluation of L2 learners' speaking fluency. However, most research in this field has compared learners at different proficiency levels instead of following their development over time. Secondly, because most previous research has been conducted with L2 learners of English and other Indo-European languages, it is unknown whether observations based on the available data can be generalized to other increasingly popular non-European target languages such as Mandarin Chinese. The present investigation sought to address these issues.

2. Methodology

By adopting a longitudinal design, the purpose of this study is (1) to examine the changes in fluency indicators over time and (2) to investigate whether learners at different speaking proficiency levels performed differently in certain fluency indicators over time. To address the issues raised above, how the issues are addressed is elaborated in the following sections, including participants, data collection and analyses.

2.1 Participants

This investigation is based on speech samples taken from 26 international students enrolled in an undergraduate course (bachelor's degree) of "Chinese Language and Culture" at the Department of Chinese as a Second Language, National

Taiwan Normal University. Students are required to obtain at least 128 credit points for graduation and they normally graduate within 3 to 5 years. Most students already have received 3 to 12 months of Chinese training either at a Mandarin Training Center in Taiwan or in their home country when they start the course. This corresponds to a CEFR proficiency level of A1~B1. They normally advance to a proficiency level of A2~B2 by the second year, and B1~C1 by the third year.

The course consists of 14 hours of language classes per week and 4-8 hours/week of culture and history classes per week taught in Chinese. The classes are taught over 18 weeks per semester; thus, students receive around 504 hours of language training and attend around 216 hours of other Chinese-taught classes per academic year. Of the 26 participants included in the present study, 7 were from Japan, 6 from Vietnam, 3 from Thailand, 2 from South Korea, 2 from Indonesia, and 1 student each from Austria, Canada, France, Guatemala, Switzerland, and the US. The participants' age ranged from 20 and 42 years (median = 24 years).

2.2 Data Collection and Analyses

The data for the present study were taken from the yearly computer-based language examinations. During these examinations, each participant received two questions, for which they were given 60 seconds to familiarize themselves with the examination procedures and 90 seconds to record their answers. Since task difficulty might influence participants' fluency (Préfontaine and Kormos 2015), we conducted two additional analyses to investigate whether the difficulty of exam topics significantly varied across years. First, we collected two kinds of expert judgments from the department's CSL teachers. One judgment required the experts to rate on a scale of 1 to 10 to what degree the topics conformed to a number of CEFR standards (e.g. "does the exam topic require students to talk about their own experience and impressions?"). Most exam topics generally received high ratings > 7 on all items. A one-way ANOVA with repeated measures conducted with each to-be-rated item as dependent measure and *year* as fixed factor did not reveal significant differences ($F = 2.823$, $p = .124 > .05$). Another judgment required the experts to rate from a scale of 1 to 3 the appropriateness of the exam questions for the CEFR levels A1~C2 (1 = not appropriate, 2 = somewhat appropriate, 3 = fully appropriate). Both experts rated all exam questions as appropriate for CEFR levels B2, C1 and C2. Some of the exam

questions were also rated as being appropriate for B1. Importantly, however a one-way ANOVA with repeated measures conducted on these data, again, showed no significant difference across years ($F = .260, p = .773 > .05$).

Second, we classified for each year and topic the vocabulary required according to the CEFR levels and then compared whether the distribution of words across CEFR levels changed across the years (Chu and Kao 2016). The computation of the percentage of words corresponding to each CEFR level was done automatically using an online Chinese text classification tool of National Taiwan Normal University². To give an example, for the first year's question "please introduce an ordinary person you respect the most." the results are A1 = 60%, A2 = 20%, B1 = 20%, B2/C1 = 0%. Having such data for each year and topic, we tested for differences in the percentages across the years with multiple ANOVAs. If there was an increase in difficulty each year, we would expect that the percentage of words corresponding to higher CEFR levels increase across the years. However, we found no evidence for this hypothesis (see Table 1).

Table 1: Vocabulary Distribution across the Years according to CEFR Levels

CEFR Level	Year	Mean %	SD	<i>F</i> value	<i>p</i> value
A1	1	45.40	13.08	.057	.945
	2	47.92	14.23		
	3	44.87	14.84		
A2	1	5.24	8.88	2.763	.107
	2	11.11	3.93		
	3	15.39	4.83		
B1	1	22.45	14.43	.413	.672
	2	29.86	13.10		
	3	28.05	12.45		
B2	1	14.33	11.43	.750	.495
	2	6.94	4.81		
	3	7.14	14.29		
C1	1	5.09	9.03	.196	.825
	2	4.17	4.81		
	3	2.27	4.55		
None	1	7.50	9.05	1.739	.221
	2	.00	.00		
	3	2.27	4.55		

² <http://huayutools.mtc.ntnu.edu.tw/ts/TextSegmentation.aspx>

Data were collected from the students' first, second, and third year of enrollment. Each speech sample was independently transcribed by two assistants. The assistants were instructed to transcribe segments and tones, "as they hear them" (that is, without paying attention to grammar and meaning of the speech) and to indicate the occurrence of filled pauses. Furthermore, in line with recent recommendations (De Jong and Bosker 2013) silent pauses extending over 250 ms were automatically tagged using the speech software PRAAT (Boersma and Weenink 2015). These transcriptions, along with the pauses tags, were checked by the same assistants during a second session a few weeks later. Subsequently, segmented transcriptions were screened for repetitions, self-corrections, and false starts. A word was counted as a repetition, if it occurred two or more times in a sequence (e.g., “我覺得 .. 覺得” ; Pinyin: “wo3 jue2de2 ..jue2de2”; English: “I think ..think”) or if it was part of a repeated sequence (e.g., “我覺得 .. 我覺得” ; Pinyin: “wo3 jue2de2 ..wo3jue2de2”; English: “I think ..I think”) excluding repetitions that are part of a common phrase or repetitions used for emphasis (e.g., “好熱好熱” ; Pinyin: “hao3re4hao3re4”; English: “very hot very hot”). Self-corrections were defined as revisions that aimed at correcting a preceding segmental or tonal error (e.g., “我 jue1.. 覺 得 ”; Pinyin: “wo3 jue1 ..jue2de2”). False starts were defined as the disruption of an incomplete clause followed by the start of a new clause (e.g., “我覺得 .. 我希望” ; Pinyin: “wo3 jue2de2 .. wo3 xi1wang4”; English: “I think .. I hope”).

Researchers have discussed which indicators are the most appropriate to measure L2 fluency. Although their discussions have not reached consensus, it has been suggested that some composite measures (such as speech rate, pause time proportion and mean length of run) should be avoided because these measures might overlap with others (Tavakoli and Skehan 2005; Kormos 2006; de Jong, Steinel, Florijn, Schoonen and Hulstijn 2012; Skehan 2014). Therefore, in the current study, the following non-composite fluency indicator commonly used to measure fluency (e.g. Koponen and Riggensbach 2000; Skehan 2003, 2009; Tavakoli and Skehan 2005; Bosker, Pinget, Quene, Sanders and de Jong 2012; Préfontaine and Kormos 2015) were calculated for each participant and year:

- (1) *Articulation rate* defined as the number of syllables per 60 seconds excluding both filled and unfilled pauses

- (2) *Pause frequency* defined as the number of unfilled and filled pauses per 60 seconds
- (3) *Disfluency ratio* defined as the number of self-corrections/false starts/repetitions per 60 seconds

As a result of having two independent transcribers, we obtained two scores for each speech sample and measurement. To ensure the reliability of the coding of fluency indicators, data was coded by two transcribers. This was repeated for a second or third time until an intertranscriber reliability of 90% was obtained in each fluency indicator. We therefore averaged the scores for further analyses. Based on previous research (Lennon 1990; Riggenbach 1991; Mora and Valls-Ferrer 2012), we expected increases in production speed measures, a reduction of pauses and repairs. One-way ANOVAs with repeated-measure were performed to examine the changes in the fluency indicators across three years. In addition, we were interested in whether learners at different proficiency levels performed differently in each fluency indicator across three years, we divided research participants into three proficiency levels (high, intermediate and low levels) based on their speaking examination scores³. The learners at high levels referred to the upper one-third in a group below 100 students and those at low levels referred to the lower one-third in a group below 100 students (Bachman 2004). Note that, since the current study is a three-year longitudinal research design, the individual learners at each proficiency level change from year to year depending on their examination scores in each year. One-way ANOVA and one-way MANOVA with independent samples were performed to address this issue. Although these scores were not specifically based on fluency judgments, the relationship between speaking proficiency levels and the fluency indicators enabled us to study the effects of overall speaking proficiency on fluency measures.

³ The speaking performances were rated by two senior CSL teachers at National Taiwan Normal University who also serve as senior examiners for the speaking part of the Test of Chinese as a Foreign Language (TOCFL). These teachers independently rated each answer from each student on a scale of 0-100 points on a number of dimensions: pronunciation accuracy (overall number of pronunciation mistakes), speaking fluency (speed, pauses, intonation), and content (appropriate vocabulary, grammar, and meaning of the productions). For each task a mean score of speaking performance was derived by equal weighting of these three dimensions. In this study, the interrater reliability coefficients of speaking examination scores from the first year to the third year are .840, .877 and .956 respectively. Since all the reliability coefficients are more than .700, the data is considered quite reliable. The examination scores for each year were then obtained by averaging the two independent ratings.

3. Results

In this section, statistical analyses are performed and results are reported to address the issues raised in this study. The results are mainly twofold. They are: (1) the changes in fluency indicators across three years and (2) the effects of different levels of speaking proficiency on fluency indicators across three years.

3.1 The Changes in Fluency Indicators across Three Years

3.1.1 Articulation Rate

Before performing a one-way ANOVA with repeated-measures to examine the changes of articulation rates across three years, we confirmed whether the data violated the assumption of homogeneity. The data did not pass the Mauchly's Test of Sphericity with a result of .657 ($p < .05$), indicating heterogeneity of the data. Greenhouse-Geisser correction was, therefore, applied to correct the degrees of freedom. After correcting for Sphericity, the one-way ANOVA was performed. As can be seen from Table 2, articulation rate exhibited large numerical increases between the first and the second year, but small numerical decreases between the second and the third year. This pattern was statistically significant among the three years ($F = 50.639$, $p < .05$, partial eta-squared = .669). The partial eta-squared for the main effect of years accounted for 66.9% of the variance in articulation rates. The post-hoc comparisons showed that the articulation rates significantly increased from the first year ($\bar{x} = 111.442$) to the second ($\bar{x} = 160.292$) and third year ($\bar{x} = 147.335$). Note that the articulation rate, despite the slight decrease in production speed between the second and the third year, was still significantly higher in the third year as compared to the first year. What is more noteworthy is that the slight decrease from the second the third year also reached the significance ($p < .05$).

Table 2: Articulation Rates across Three Years

		1st year	2nd year	3rd year	F	p	η^2	Post-hoc
Articulation rate	Mean	111.442	160.292	147.335	50.639	.000	.669	2 > 1
	SD	20.668	33.497	20.464				3 > 1
								2 > 3

Note: 1 = the first year, 2 = the second year, 3 = the third year

3.1.2 Pause Frequency: Filled and Unfilled Pauses

To confirm the assumption of homogeneity, Mauchly's Test of Sphericity was performed. Both data of filled and unfilled pauses did not pass the test (Unfilled pauses: Mauchly's $W = .452$, $p < .05$; filled pauses: Mauchly's $W = .772$, $p < .05$), indicating heterogeneity of the data. Greenhouse-Geisser corrections were, therefore, used to correct the degrees of freedom. After correcting for Sphericity, the one-way ANOVAs were performed. Numerically, filled and unfilled pauses showed distinct patterns of changes over time (Table 3). The number of unfilled pauses per minute seemed to remain stable between the first, second, and the third year. The number of filled pauses per minute, on the other hand, increased between the first and the second year, as well as between the first and the third year. In line with this observation, one-way ANOVA did not reveal a significant change in the number of unfilled pauses between the first, second and the third year ($F = .068$, $p > .05$, eta-squared = .003). On the other hand, the one-way ANOVA revealed significant changes in the number of filled pauses among the three years ($F = 6.185$, $p < .05$, eta-squared = .198). The partial eta-squared for the main effect of years accounted for 19.8% of the variance in the number of filled pauses per minute. The post-hoc comparison analyses further showed that the number of filled pauses increased significantly during the first ($\bar{x} = 3.869$) and the second year ($\bar{x} = 5.769$), but the increase between the second ($\bar{x} = 5.769$) and the third year ($\bar{x} = 7.058$) did not reach significance. In addition, students significantly produced more filled pauses in the third year ($\bar{x} = 7.058$) compared to the first year ($\bar{x} = 3.869$).

Table 3: Unfilled and Filled Pauses across Three Years

		1st year	2nd year	3rd year	F	p	η^2	Post-hoc
Unfilled pauses	Mean	19.927	20.158	20.542	.068	.856	.003	N/A
	SD	5.127	10.102	4.090				
Filled pauses	Mean	3.869	5.769	7.058	6.185	.007	.198	2, 3 > 1
	SD	2.901	4.883	5.685				

Note: 1 = the first year, 2 = the second year, 3 = the third year

3.1.3 Disfluency Ratio: Self-corrections, Repetitions, and False Starts

To confirm the assumption of homogeneity, Mauchly's Test of Sphericity was performed. The data passed the test (Self-corrections: Mauchly's $W = .790$, $p > .05$; repetitions: Mauchly's $W = .988$, $p > .05$; false starts: Mauchly's $W = .902$, $p > .05$), indicating homogeneity of the data. After confirming the assumption of homogeneity, we performed one-way ANOVA. Inspection of Table 4 suggests that although self-corrections decreased between the second and third year, one-way ANOVA with repeated measures did not reveal any significant changes among the three years ($F = 2.292$, $p > .05$, eta-squared: .084). In contrast, both repetitions and false starts significantly increased from the first year (repetitions: $\bar{x} = 3.219$; false starts: $\bar{x} = 1.019$) to the second (repetitions: $\bar{x} = 7.681$; false starts: $\bar{x} = 3.158$) and third year (repetitions: $\bar{x} = 7.439$; false starts: $\bar{x} = 2.904$). The one-way ANOVA showed that both these patterns were significant (repetitions: $F = 16.124$, $p < .05$, eta-squared: .392; false starts: $F = 16.115$, $p < .05$, eta-squared: .392). The partial eta-squared for the main effect of years accounted for 39.2% of the variance in the number of repetitions or false starts per minute.

Table 4: Self-corrections, Repetitions and False Starts across Three Years

		1st year	2nd year	3rd year	F	p	η^2	Post-hoc
Self-corrections	Mean	.404	.450	.227	2.292	.112	.084	N/A
	SD	.384	.455	.257				
Repetitions	Mean	3.219	7.681	7.439	16.124	.000	.392	2, 3 > 1
	SD	2.796	4.459	4.801				
False starts	Mean	1.019	3.158	2.904	16.115	.000	.392	2, 3 > 1
	SD	.768	2.307	2.031				

Note: 1 = the first year, 2 = the second year, 3 = the third year

3.2 Fluency Indicators at Different Levels of Speaking Proficiency across Three Years

3.2.1 Articulation Rate at Different Proficiency Levels

A comparison was made of the articulation rates produced by the learners in the groups of different speaking proficiency levels in the first year (Table 5). Before we performed one-way ANOVA with independent samples, the Levene's test was run

to test the homogeneity of variances of the articulation rates in the three proficiency levels. The results indicated that the three levels have equal variances, showing homogeneity of the data (Levene: .331, $p > .05$). After we confirmed the assumption of homogeneity, the one-way ANOVA indicated that the main effects of proficiency levels on articulation rates did not reach significance ($F = .781, p > .05$).

Table 5: Articulation Rates at Different Proficiency Levels in the First Year

Proficiency levels	N	Mean	SD.	F	p	η^2	Post-hoc
High	8	118.913	24.859	.781	.470	.064	N/A
Intermediate	10	109.350	20.464				
Low	8	106.588	16.501				

Another comparison was made of the articulation rates produced by the learners in the groups of different proficiency levels in the second year (Table 6). The results of the Levene's test indicated that the three levels have equal variances, showing homogeneity of the data (Levene: .000, $p > .05$). After we confirmed the assumption of homogeneity, the one-way ANOVA indicated that the main effects of proficiency levels on articulation rates reached significance ($F = 4.965, p < .05$, partial eta-squared = .302). The partial eta-squared for the main effect of proficiency levels accounted for 30.2% of the variances in the articulation rates in the second year. A post-hoc comparison found a significant difference between high and low level proficiency learners (mean difference = 38.862, 95% CI = 8.675, 69.050, $p < .05$) and also between intermediate and low level proficiency learners (mean difference = 39.252, 95% CI = 10.614, 67.891, $p < .05$).

Table 6: Articulation Rates at Different Proficiency Levels in the Second Year

Proficiency levels	N	Mean	SD.	F	p	η^2	Post-hoc
High	8	172.100	31.054	4.965	.016	.302	1, 2 > 3
Intermediate	10	172.490	28.348				
Low	8	133.238	28.306				

Note: 1 = high level; 2 = intermediate level; 3 = low level

The other comparison was made of the articulation rates produced by the learners in the groups of different proficiency levels in the third year (Table 7). The results of the Levene's test showed the homogeneity of the data (Levene: .555, $p > .05$). After the assumption of homogeneity was confirmed, one-way ANOVA indicated that the main effects of proficiency levels on articulation rates reached significance ($F = 6.781$, $p < .05$, partial eta-squared = .371). The partial eta-squared for the main effect of proficiency levels accounted for 37.1% of the variances in the articulation rates in the third year. A post-hoc comparison therefore revealed a significant difference between high and intermediate level proficiency learners (mean difference = -21.360, 95% CI = -37.964, -4.756, $p < .05$) and also between intermediate and low level proficiency learners (mean difference = 27.797, 95% CI = 11.194, 44.401, $p < .05$).

Table 7: Articulation Rates at Different Proficiency Levels in the Third Year

Proficiency levels	N	Mean	SD.	F	p	η^2	Post-hoc
High	8	141.100	13.735	6.781	.005	.371	2 > 1, 3
Intermediate	10	162.460	15.835				
Low	8	134.663	20.464				

Note: 1 = high level; 2 = intermediate level; 3 = low level

3.2.2 Pause Frequency at Different Proficiency Levels

Learners' pause frequency was classified into filled and unfilled pauses. Learners' numbers of filled and unfilled pauses per minute at different proficiency levels in the first year were investigated (Table 8). Before we performed one-way MANOVA with independent samples, the Levene's tests were run to test the homogeneity of variances of the filled and unfilled pauses in the three proficiency levels. Both results showed the homogeneity of the data (Levene in filled pauses: 1.875, $p > .05$; Levene in unfilled pauses: .080, $p > .05$). After confirming the assumption of homogeneity, the one-way MANOVA indicated that the main effects of proficiency levels on both filled pauses ($F = .406$, $p > .05$) and unfilled pauses ($F = 1.527$, $p > .05$) did not reach any significance.

Table 8: Pause Frequency at Different Proficiency Levels in the First Year

Types of pauses	Proficiency levels	N	Mean	SD.	<i>F</i>	<i>p</i>	η^2	Post-hoc
Unfilled	High	8	17.538	4.705	1.527	.238	.117	N/A
	Intermediate	10	21.690	5.510				
	Low	8	20.113	4.657				
Filled	High	8	4.513	3.854	.406	.671	.034	N/A
	Intermediate	10	3.910	2.856				
	Low	8	3.175	1.923				

Analyses were carried out to investigate the learners' numbers of filled and unfilled pauses per minute at different proficiency levels in the second year (Table 9). The result of unfilled pauses showed the homogeneity of the data (Levene: 3.240, $p > .05$) but the result of filled pauses showed the heterogeneity of the data (Levene: 5.231, $p < .05$). The one-way MANOVA indicated that the main effects of proficiency levels on both filled pauses ($F = 12.659$, $p < .05$, partial eta-squared: .524) and unfilled pauses ($F = 9.193$, $p < .05$, partial eta-squared: .444) reached significance. The partial eta-squared for the main effects of proficiency levels accounted for 52.4% of the variances in the filled pauses and 44.4% of the variances in the unfilled pauses in the second year. Since the data of unfilled pauses confirmed the homogeneity assumption, a post hoc comparison using LSD revealed a significant difference between high and low level proficiency learners (mean difference = -15.413, 95% CI = -23.534, -7.291, $p < .05$), and also between intermediate and low level proficiency learners (mean difference = -13.163, 95% CI = -20.867, -5.458, $p < .05$). The effects of proficiency levels on unfilled pauses showed a different pattern from those on filled pauses. Since the data of filled pauses showed heterogeneous, a post hoc comparison analysis using Tamhane was conducted. The post hoc comparison analysis showed a significant difference between high and intermediate level proficiency learners (mean difference = 6.435, 95% CI = .831, 12.039, $p < .05$), and also between high and low level proficiency learners (mean difference = 8.388, 95% CI = 3.044, 13.731, $p < .05$).

Table 9: Pause Frequency at Different Proficiency Levels in the Second Year

Types of pauses	Proficiency levels	N	Mean	SD.	<i>F</i>	<i>p</i>	η^2	Post-hoc
Unfilled	High	8	14.550	3.896	9.193	.001	.444	3 > 1 2
	Intermediate	10	16.800	2.789				
	Low	8	29.963	13.318				
Filled	High	8	10.825	4.856	12.659	.001	.524	1 > 2 3
	Intermediate	10	4.390	3.403				
	Low	8	2.438	1.436				

Note: 1 = high level; 2 = intermediate level; 3 = low level

Learners' numbers of filled and unfilled pauses per minute at different proficiency levels in the third year were investigated (Table 10). Both results showed the homogeneity of the data (Levene in filled pauses: 1.118, $p > .05$; Levene in unfilled pauses: 3.154, $p > .05$). The one-way MANOVA indicated that the main effects of proficiency levels on filled pauses did not reach significance ($F = 1.796$, $p > .05$, partial eta-squared: .135) but the main effects of proficiency levels on unfilled pauses reached significance ($F = 4.394$, $p < .05$, partial eta-squared: .276). The partial eta-squared for the main effects of proficiency levels accounted for 27.6% of the variances in the unfilled pauses in the third year. A post hoc comparison analysis for unfilled pauses was therefore conducted. A significant difference was found between intermediate and low level proficiency learners (mean difference = -5.050, 95% CI = -8.609, -1.491, $p < .05$).

Table 10: Pause Frequency at Different Proficiency Levels in the Third Year

Types of pauses	Proficiency levels	N	Mean	SD.	<i>F</i>	<i>p</i>	η^2	Post-hoc
Unfilled	High	8	20.988	4.518	4.394	.024	.276	3 > 2
	Intermediate	10	18.100	3.394				
	Low	8	23.150	2.830				
Filled	High	8	6.300	4.456	1.796	.188	.135	N/A
	Intermediate	10	5.260	4.615				
	Low	8	10.063	7.253				

Note: 1 = high level; 2 = intermediate level; 3 = low level

3.2.3 Disfluency Ratio at Different Proficiency Levels

Learners' disfluency ratio was divided into self-corrections, repetitions, and false starts. Learners' numbers of self-corrections, repetitions, and false starts per minute at different speaking proficiency levels in the first year were investigated (Table 11). The results of self-corrections and repetitions showed the homogeneity of the data (Levene in self-corrections: .750, $p > .05$; Levene in repetitions: 1.723, $p > .05$), but the data of false starts showed heterogeneous (Levene: 3.662, $p < .05$). The one-way MANOVA was performed, indicating that the main effects of proficiency levels on self-corrections ($F = .164$, $p > .05$), repetitions ($F = .945$, $p > .05$), and false starts ($F = .350$, $p > .05$) did not reach any significance.

Table 11: Disfluency Ratio at Different Proficiency Levels in the First Year

Types of Disfluency	Proficiency levels	N	Mean	SD.	F	p	η^2	Post-hoc
Self-corrections	High	8	.363	.457	.164	.850	.014	N/A
	Intermediate	10	.460	.347				
	Low	8	.375	.396				
Repetitions	High	8	2.800	2.399	.945	.403	.076	N/A
	Intermediate	10	4.160	3.451				
	Low	8	2.463	2.176				
False starts	High	8	.950	.496	.350	.709	.030	N/A
	Intermediate	10	.920	.724				
	Low	8	1.213	1.062				

Analyses were carried out to investigate learners' number of self-corrections, repetitions, and false starts per minute at different proficiency levels in the second year (Table 12). All the results of the Levene's test showed the homogeneity of the data obtained (Levene in self-corrections: 1.924, $p > .05$; Levene in repetitions: 2.609, $p > .05$; Levene in false starts: 2.161, $p > .05$). After the assumption of homogeneity was confirmed, the one-way MANOVA was performed, showing the significant effects of proficiency levels on the number of self-corrections per minute ($F = 3.705$, $p < .05$, partial eta-squared: .244). The partial eta-squared for the main effects of proficiency levels accounted for 22.4% of the variances in the self-corrections. A post hoc comparison analysis further revealed a significant difference between high and intermediate level proficiency learners (mean difference = -.523, 95% CI = -.927, -.118, $p < .05$). On the other hand, the significant effects of proficiency levels on the number

of repetitions ($F = .643, p > .05$, partial eta-squared: .053) and false starts ($F = .543, p > .05$, partial eta-squared: .045) per minute were not evident.

Table 12: Disfluency Ratio at Different Proficiency Levels in the Second Year

Types of Disfluency	Proficiency levels	N	Mean	SD.	F	p	η^2	Post-hoc
Self-corrections	High	8	.188	.210	3.705	.040	.244	2 > 1
	Intermediate	10	.710	.534				
	Low	8	.388	.383				
Repetitions	High	8	6.363	1.641	.643	.535	.053	N/A
	Intermediate	10	7.740	5.280				
	Low	8	8.925	5.357				
False starts	High	8	2.438	.741	.543	.588	.045	N/A
	Intermediate	10	3.490	2.971				
	Low	8	3.463	2.502				

Note: 1 = high level; 2 = intermediate level; 3 = low level

Learners' numbers of self-corrections, repetitions, and false starts per minute at different levels in the third year were investigated (Table 13). All the results of the Levene's tests showed the homogeneity of the data (Levene in self-corrections: .800, $p > .05$; Levene in repetitions: 1.200, $p > .05$; Levene in false starts: .646, $p > .05$). After confirming the assumption of homogeneity, we performed the one-way MANOVA and found that the main effects of proficiency levels on self-corrections ($F = .760, p > .05$), repetitions ($F = 1.413, p > .05$), and false starts ($F = 2.986, p > .05$) did not reach any significance.

Table 13: Disfluency Ratio at Different Proficiency Levels in the Third Year

Types of Disfluency	Proficiency levels	N	Mean	SD.	F	p	η^2	Post-hoc
Self-corrections	High	8	.150	.321	.760	.479	.062	N/A
	Intermediate	10	.300	.194				
	Low	8	.213	.264				
Repetitions	High	8	5.175	3.589	1.413	.264	.109	N/A
	Intermediate	10	8.030	4.543				
	Low	8	8.963	5.822				
False starts	High	8	1.613	1.477	2.986	.070	.206	N/A
	Intermediate	10	3.180	1.933				
	Low	8	3.850	2.172				

4. Discussion and Conclusions

The present study assessed the long-term development of oral fluency among Mandarin Chinese learners from different language backgrounds. We have reported analyses of oral examination data taken from 26 international students enrolled in an undergraduate program of Chinese as a Second Language at three different points in time between their first and third year of study. In addition, we have kept track of how learners at different proficiency levels performed in each fluency indicator over three years. In what follows, we discuss the results by comparing the previous literature and the current study.

4.1 Changes in Fluency Indicators across Three Years

Consistent with previous studies (e.g., Mora and Valls-Ferrer 2012; Du 2013) we found a substantial increase in production speed as measured by articulation rate (syllables per minute excluding pauses). Since Du's (2013) three-year longitudinal study which also investigated Chinese fluency development did not take pause frequency and disfluency into considerations, to gain a more complete picture, we tested for the changes in more fluency indicators across years and found that filled pauses, repetitions, and false starts significantly increased between the first and the second year, and that there was an additional numerical trend for a further increase in the third year. Thus, it seems that students traded higher production speed at the cost of producing more interrupted speech. Despite the increase in filled pauses, repetitions, and false starts, students' L2 fluency, as indicated by temporal performance measures, significantly improved over the years.

Overall, our results partially confirm previous research showing that L2 learners increase their production speed. Consistent with previous research, the change in articulation rates seems to be reliable predictors for the perception of oral proficiency (Lennon 1990; Riggensbach 1991; Kormos and Dénes 2004; Iwashita 2010). However, the most obvious deviating pattern found concerned the increase in repetition and false start as well as the increase of fillers. It is important to consider why students increased their production speed accepting that, as a result, they would also repeat or correct themselves more frequently and pause more often. Potential explanations for these phenomena might be related to the specific participant sample, the language

being learned, or the study abroad learning context. Since exam questions were more or less related to topics discussed during the lessons, one might speculate that learners may have prepared for the exams by memorizing a portion of their answers. In such cases, occasional memory retrieval failures during the exam could result in fillers, repetitions and false starts. Note that this largely happened between the first year and the second/third year, which is consistent with the hypothesis that learners took advantage of their experience with the first exam and adjusted their preparation strategy by focusing on rote learning in the second and third year. According to the Adaptive Control of Thought (ACT) model of cognitive development (Anderson 1983), the conversion of declarative knowledge into procedural knowledge is necessary for becoming fluent in oral production. It is therefore inferred that the increase in fillers, repetitions and false starts might serve as immediate response strategy to compensate for the lack of procedural knowledge.

Alternatively or additionally, one might suspect a desire on the students' part to adapt to the new environment after arriving in Taiwan. For instance, one could speculate that changes in production speed were encouraged by the experienced CSL teachers as part of an educational philosophy or strategy (focus on speed increase). These students, therefore, developed the strategy to focus on producing more speech at the cost of also producing more interruptions from the first year to the second and third year.

4.2 Fluency Indicators at Different Levels of Speaking Proficiency across Three Years

Although the operational definitions of different levels of speaking proficiency vary across previous literature, the cross-sectional studies conducted in several learning contexts (such as ESL, EFL, JFL or CSL), have consistently shown that higher level learners' speed production was higher than lower level learners' (Iwashita et al., 2008; Iwashita 2010; Li 2014). The current study, by adopting a longitudinal research design, sheds a new light on the relationship between different levels of speaking proficiency and the development of fluency indicators.

We found that CSL learners' speed production as measured by articulation rates, pause frequency and disfluency ratio in the first year did not distinguish between any proficiency levels. A close look at articulation rates revealed that although the

articulation rates in the second year, similar to previous findings, could distinguish high level and intermediate level learners from low level learners, the articulation rates in the third year revealed a different picture concerning the influence of different proficiency levels. That is, the intermediate level learners are distinguished from high level and low level learners in terms of the third year's articulation rates.

The findings of articulation rates above should be discussed with the development of other fluency indicators including pause frequency and disfluency ratio. The low level learners seem to speak more slowly between the second and the third year because they tended to make more frequent unfilled pauses. The intermediate level learners, in the second year, increased their articulation rates by self-correcting themselves more frequently and then in the third year, they could still speak faster without self-corrections. High level learners tended to speak faster and meanwhile, made more frequent filled pauses in the second year. However, they seem to slow down the articulation rates in the third year.

It has been widely assumed that higher level learners tend to speak faster than lower level learners (Iwashita et al. 2008; Iwashita 2010; Li 2014). The current study, contrary to previous research, showed that no significant difference was evident between high level learners' and low level learners' articulation rates in the third year. The difference could be attributed to the different language characteristics investigated and different research design adopted in the current and previous studies. As to the research design, Li's (2014) cross-sectional study showed that higher level CSL learners spoke faster than lower level learners. His finding seems to be consistent with the second year's result in the current study because we found high and intermediate level learners spoke faster than low level learners in the second year. We, however, improved Li's research design by adopting a longitudinal design and found that, in the third year, it is the intermediate level learners that tended to speak faster than high and low level learners. It is suggested that speaking quickly might be a typical characteristics of intermediate learners. As for language characteristics, Mandarin Chinese is a tonal language with four distinctive tones which vary with different lexical meanings (Cipollone, Keiser, and Vasishth 1998). Compared with English and Japanese investigated in previous studies (Iwashita et al. 2008; Iwashita 2010), learners of Chinese have to pay more attention to the tones while uttering

Chinese words. It is, therefore, inferred that high level learners in the third year might have experienced speed-clarity trade-offs in the oral tasks. In other words, high level learners might prioritize clear pronunciation over fast pronunciation.

CSL learners' speed-clarity trade-offs in this study happen to reflect the debate between fluency and accuracy in the area of second language acquisition. Skehan (1998) claimed that language performance is basically comprised of three linguistic dimensions including complexity, accuracy and fluency. As to the interrelationship among these three linguistic dimensions, it has been extensively discussed whether to increase one linguistic dimension would decrease another. He further proposed a trade-off hypothesis (Skehan 2009), claiming that due to students' limited cognitive capacity, focusing on accuracy might lower the performance of fluency and vice versa. This trade-off phenomenon in the oral performance might provide pedagogical implications for CSL/CFL teachers. CSL/CFL teachers might emphasize the importance of speaking fluency in classrooms under the impression that speaking fluently is a typical indicator of proficient learners. Nevertheless, learners, especially those at higher proficiency levels, might finally notice that clear and correct utterance might be more pragmatic for successful communication in real-world contexts. CSL/CFL instructors, therefore, should help develop certain effective communication strategies to increase students' speaking accuracy or complexity, depending on the speaking tasks students will be engaged in.

Regardless of how the present data are interpreted, our investigation has shown that many of the commonly-used performance measures of oral fluency also give a meaningful picture of L2 Chinese learners' fluency development. Although the long-term development of several measures was in line with research on Indo-European languages, a few distinct patterns in the development of Mandarin Chinese learners have also been demonstrated. In particular, we have shown that although articulation rates and filled pauses can actually increase by levels in the second year, they do not always serve as a significant indicator to distinguish higher level learners from lower level learners. We, therefore, hope that future research will shed light on the potential factors that underlie the specific pattern found within our sample of Mandarin Chinese L2 learners.

References

- Anderson, R. John. 1983. *The Architecture of Cognition*. Cambridge MA: Harvard University Press.
- Bachman, F. Lyle. 2004. *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.
- Boersma, Paul, and David Weenink. 2015. Praat: Doing phonetics by computer [Computer program]. Version 5.4.22. Accessed online, February 8, 2016. <http://www.praat.org/>
- Bosker, R. Hans, Anne-France Pinget, Hugo Quené, Ted Sanders, and Nivja H. De Jong. 2012. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30: 159-175.
- Chambers, Francine. 1998. What do we mean by fluency? *System* 25: 535-544.
- Chu, Wo-Hsin, and Chian-Wen Kao. 2016. A longitudinal study on international students' Chinese pronunciation developments: A comparison between tone language and non-tone language speakers. *Journal of Chinese Language Teaching* 13: 77-121.
- Cipollone, Nick, Steven Hartman Keiser, and Shravan Vasishth. 1998. *Language Files*. Ohio State University Press.
- De Jong, H. Nivja, and Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In *Proceedings of Disfluency in Spontaneous Speech (DiSS'03)*, ed. by Robert Eklund, 17-20.
- De Jong, H. Nivja, Margarita P. Steinel, Arjen F. Florijin, Rob Schoonen and Jan H. Hulstijn. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition* 34: 5-34.
- Derwing, M. Tracey, Murray J. Munro, and Ron I. Thomson. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29: 359-380.
- Du, Hang. 2013. The development of Chinese fluency during study abroad in China. *The Modern Language Journal* 97: 131-143.
- Freed, F. Barbara, Norman Segalowitz, and Dan P. Dewey. 2004. Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition* 26: 275-301.
- Griffiths, Roger. 1991. Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics* 12: 345-364.

- Iwashita, Noriko, Annie Brown, Tim McNamara, and Sally O'Hagan. 2008. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* 29: 24-49.
- Iwashita, Noriko. 2010. Features of oral proficiency in task performance by EFL and JFL learners. *Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices*, 32-47. Somerville MA: Cascadia Press.
- Kahng, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64: 809-854.
- Koponen, Matti, and Heidi Riggenbach. 2000. Overview: Varying perspectives on fluency. In *Perspectives on Fluency*, ed. by Heidi Riggenbach, Michigan: University of Michigan Press.
- Kormos, Judit. 2006. *Speech Production and Second Language Acquisition*. Mahwah NJ: Lawrence Erlbaum.
- Kormos, Judit, and Mariann Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32: 145-164.
- Lennon, Paul. 1990. Investigating fluency in EFL: a quantitative approach. *Language Learning* 40: 387-417.
- Li, Shuai. 2014. The effects of different levels of linguistic proficiency on the development of L2 Chinese request production during study abroad. *System* 45: 103-116.
- Mora, C. Joan, and Margalida Valls-Ferrer. 2012. Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly* 46: 610-641.
- Ortega, Lourdes, and Gina Ibarra-Shea. 2005. Longitudinal research in second language acquisition: Recent trends and future directions. *Applied Linguistics* 25: 26-45.
- Ortega, Lourdes, and Heidi Byrnes. (eds.) 2008. *The Longitudinal Study of Advanced L2 Capacities*. New York: Routledge.
- Préfontaine, Yvonne, and Judit Kormos. 2015. The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal* 99: 96-112.
- Riggenbach, Heidi. 1991. Towards an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14: 423-441.

- Schmidt, Richard. 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14: 357-385.
- Segalowitz, Norman, and Barbara F. Freed. 2004. Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition* 26: 173-199.
- Segalowitz, Norman, and Barbara F. Freed, Joe G. Collentine, Barbara A. Lafford, Nicole Lazar, and Manuel A. Díaz-Campos. 2004. A comparison of Spanish second language acquisition in two different learning contexts: study abroad and the domestic classroom. *Frontiers* 10: 21-38.
- Skehan, Peter. 1998. *A Cognitive Approach to Language Learning*. Oxford UK: Oxford University Press.
- Skehan, Peter. 2003. Task-based instruction. *Language Teaching* 36: 1-14
- Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics* 30: 510-532.
- Skehan, Peter. 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In *Processing Perspectives on Task Performance*, ed. by Peter Skehan, 211-260. Amsterdam: John Benjamins.
- Tavakoli, Parvaneh, and Peter Skehan. 2005. Strategic planning, task structure and performance testing. In *Planning and Task Performance in a Second Language*, ed. by Rod Ellis, 239-273. Amsterdam: John Benjamins.
- Tseng, Shu-Chuan, Alexander Soemer, and Tzu-Lun Lee. 2013. Tones of reduced T1-T4 syllables. *International Journal for Computational Linguistics and Chinese Language Processing* 18: 81-106.

[審查：2018.7.1 修改：2018.7.30 接受：2018.8.21]

朱我芯

Wo-Hsin CHU

10610 臺北市和平東路一段 162 號 國立臺灣師範大學華語文教學系暨研究所

Department of Chinese as a Second Language

National Taiwan Normal University

No.162, Sec. 1, Heping E. Rd., Taipei City 10610, Taiwan

chuw@ntnu.edu.tw

華語文教學研究

高千文

Chian-Wen KAO

22050 新北市板橋區文化路 1 段 313 號 致理科技大學應用英語系

Department of Applied English

Chihlee University of Technology

No.313, Sec. 1, Wenhua Rd., Banqiao Dist., New Taipei City 22050,

Taiwan

briankao0706@mail.chihlee.edu.tw

孫德

Alexander SOEMER

14476 德國波茨坦卡爾李卜克內西街 24-25 號 波茨坦大學心理學系

Department of Psychology

University of Potsdam

Karl-Liebknecht-Str. 24-25 14476 Potsdam, Germany

soemer@uni-potsdam.de

華語學習者口語流暢度發展之縱貫性研究

朱我忞

國立臺灣師範大學
華語文教學系暨研究所

高千文

致理科技大學
應用英語系

孫德

波茨坦大學
心理學系

摘要

本研究探討不同母語背景之華語為二語學習者，其口語流暢度三年變化之發展。透過每年口語測驗所蒐集之語料顯示，受試者從第一年至第二年的構音速度 (articulation rates)、填充停頓 (filled pauses)、重複語 (repetitions) 及重起開頭語 (false starts) 之口語流暢指標有顯著提高，而從第二年至第三年卻只有構音速度指標顯著提高。本研究亦分析華語為二語學習者在口語能力及流暢指標之關係。結果發現，構音速度並非有效區辨高成就與低成就語言學習者之指標，此發現與過往研究結果有所出入。本研究就結果討論語境及教育相關變數對於口語流暢度之影響。

關鍵詞：縱貫性研究 第二語言習得 口語流暢度