

The Impact of Performative Language Teaching on Oral Skills in the Classroom of Chinese as a Foreign Language

El impacto de técnicas de enseñanza performativa del lenguaje en las habilidades orales de una clase de chino como lengua extranjera

戏剧化教学法在对外汉语课堂对口语能力的影响

modestocorderi@hotmail.com

Modesto Corderi Novoa*

Escola Oficial de Idiomas de A Coruña

Departamento de chino

La Coruña, España, 15011

mariapilar.garciamayo@ehu.eus

María del Pilar García Mayo**

Universidad del País Vasco (UPV/EHU)

English, German and Translation & Interpretation Department

Vitoria-Gasteiz, España, 01006

* Modesto Corderi Novoa (PhD in UPV/EHU), is a performative language teacher and researcher. He currently works as a Chinese teacher at the Official Language School EOI in A Coruña, Spain. He has hosted multiple teacher training workshops and conferences in the use of performative language teaching in Europe, Asia and North America. His research interests include performative language teaching and Chinese language teaching pedagogy.

 0000-0001-9830-4018

** María del Pilar García Mayo, is Full Professor of English Language and Linguistics at the University of the Basque Country (UPV/EHU). She is the director of the research group Language and Speech, the MA program Language Acquisition in Multilingual Settings and the editor of Language Teaching Research, and committee member of Spanish State Research Agency. Her research interests include second (L2)/third language (L3) acquisition of English morphosyntax and the study of conversational interaction in EFL. She is also interested in bilingual/multilingual education.

 0000-0002-1987-4889

Abstract: This study examines the impact of Performative Language Teaching (PLT) techniques on speech variables such as fluency, comprehensibility, and accent. 16 learners of Mandarin Chinese (L1: Spanish) from the Official Language School in the city of Vigo in Spain were divided into an experimental group (8 students that in 2018 took part in a four-month PLT-based Chinese course) and a control group (8 students who did not participate in that course). Four different tasks (first-person picture description, third-person picture description, monologue, and improvised role-play dialogue) were used to obtain speech samples from the 16 learners. The two groups were tested before (pre-test) and after (post-test) participation. 75 native Chinese raters evaluated the recorded speech samples and the development of oral skills of both groups was compared. Our findings indicate that fluency, comprehensibility, and accent improved for the treatment group, but not for the comparison group. The improvement in fluency was greater than the improvement in comprehensibility and accent, which had similar degrees of development, although accent had slightly lower improvement scores than comprehensibility.

Key Words: language teaching; applied linguistics; Performative Language Teaching (PLT); Chinese Mandarin; drama.

Resumen: Este estudio investiga el impacto de las técnicas de enseñanza performativa del lenguaje (PLT) en las variables orales de fluidez, comprensibilidad y acento. 16 estudiantes de chino mandarín (L1: español) de la Escuela Oficial de Idiomas en la ciudad de Vigo en España fueron divididos en un grupo experimental (8 estudiantes que participaron en 2018 un curso de clases de chino basado en PLT de cuatro meses) y en un grupo de control (otros 8 estudiantes que no recibieron dicho curso). Se emplearon cuatro tareas diferentes (descripción de imágenes en primera persona, descripción de imágenes en tercera persona, monólogo y diálogo improvisado) para obtener muestras de habla de estos 16 estudiantes. Los dos grupos fueron evaluados antes (pre-test) y después (post-test) de su participación. 75 evaluadores chinos nativos analizaron las muestras de habla obtenidas y se compararon el desarrollo de las habilidades orales de ambos grupos. Los resultados indican que la fluidez, la comprensión y el acento mejoraron para el grupo experimental, pero no para el grupo de control. La mejora en la fluidez fue mayor que la mejora en la comprensión y el acento, que tuvieron grados de desarrollo similares, aunque el acento tuvo una mejora ligeramente menor que la comprensibilidad.

Palabras clave: enseñanza de idiomas; lingüística aplicada; enseñanza performativa de idiomas; chino mandarín; teatro.

摘要: 本研究探讨戏剧化教学法(PLT)对学生口语流利度、可理解性和口音等变量的影响。文本以西班牙维戈官方语言学校母语为西班牙语的16名汉语学习者研究对象,将他们氛围实验组(8名学生于2018年在同一所学校参加了为期四个月的戏剧化教学法汉语课程)和对照组(另外8名学生未参加过该课程)。通过四项口语任务(以第一人称描述图片、第三人称描述图片、独白和即兴对话)得到这16名学生的语音样本。在两组学生参与项目之前和结束之后,由75名以中文为母语的评委对语音样本进行评估,并比较两组学生口语能力的提高程度。研究表明,实验组学生口语的流利度、可理解性和口音有所改善,而对照组在这些项目上则没有发生改变。同时,流利程度的改善大于可理解性和口音的改善,后两者的提高程度类似,口音的改善略低于可理解性的改善。

[关键词] 语言教学; 应用语言学; 戏剧化教学法; 普通话; 戏剧

1. Introduction

The rise of China as a superpower has led to a global increase in the number of learners of Chinese at all educational levels and ages. Not only Confucius Institutes but also universities, public and private schools, and institutions around the world have started to offer Chinese courses in response to the increasing demand to learn about Chinese language and culture. The same happens in Spain. However, there are several problems related to the quality of Chinese language teaching, such as the lack of trained teachers and adequate engaging materials. One of the main issues is the teaching pedagogy and methodology. Traditional Chinese teaching methods rely heavily on memorization and repetition, lacking communicative value. Most of the time, the materials used are not suitable for the learning styles of other countries. There is, therefore, a need to develop a modern and more effective methodology for Chinese language teaching and learning. Specifically, students of Chinese need a more collaborative and engaging approach to help them improve their oral proficiency skills. The main aim of this paper is to show how performative language teaching might be a good alternative.

2. Performative Language Teaching (PLT)

Bentley (1968) claimed that theater and drama are different, the first focusing on the product and the latter on the process. In order to overcome the drama/theater divide, Schewe (2013) introduced a new concept called Performative Language Teaching (PLT). Fleming (2016) also considered PLT to be a suitable substitute for the self-created dichotomies between drama and theater, product and process, and spectator and participant. More recently, Piazzoli (2018: 40) summarized the concept of PLT as an “embodied approach to teaching second languages”. Therefore, we can argue that PLT is an “umbrella term” for all different activities related to theater and drama including a vast range of dramatic forms that can be used for education and language teaching.

PLT fosters interaction. Within the field of second language acquisition (SLA), two main approaches have highlighted the importance of interaction. The Interaction Hypothesis (Long, 1996) stated that interaction among learners facilitates second language (L2) learning. When interacting and collaborating, learners negotiate for meaning and form. Interaction provides learners with opportunities to receive input, produce output, and receive corrective feedback from their interlocutors, all of which have been shown to facilitate L2 acquisition (García Mayo & Alcón Soler, 2013). Moreover,

interaction triggers cognitive mechanisms such as noticing (Schmidt, 1990), which have been claimed to be a key starting point for L2 acquisition.

In addition, Sociocultural Theory (SCT) (Vygotsky, 1978) states that the development of language learning and other cognitive capacities is socially co-constructed. Improvement occurs in the interaction between experts and apprentices, where the expert (e.g., parents, teachers, more knowledgeable peers) provides appropriate assistance to the novice. A range of tools mediates the provision of assistance. These tools can be material artifacts (e.g., computers, toys) or symbolic (e.g., gestures, language). The Zone of Proximal Development (ZPD) is the difference between what a student can do on their own and what the learner can accomplish with the help of others. The carefully adjusted assistance has also frequently been referred to in the literature as *scaffolding* (Wood, Bruner, & Ross, 1976).

PLT is aligned with Vygotsky's SCT because PLT is carried out collaboratively, and participants must work together to decide how to interact using the target language in a communicative way. PLT's team building reflects Vygotsky's ZPD since students can learn not only from the teachers but also from classmates when participating in a PLT activity. Some of those peers might have a more developed target language competence than others. When they interact, they will practice more by using the target language in a context, thus, improving their language skills and reach their full potential.

3. Literature Review

3.1 Experimental Research on the Use of PLT in the Language Classroom

While numerous studies have been previously conducted to examine the benefits of PLT in language learning and teaching, this topic is far from being exhausted as a research area. Stern (1980), Kao (1994), Podlozny (2000), Catterall (2002), Stinson and Freebody (2006), and the DICE Consortium (2010) confirm that there is overwhelming evidence indicating that PLT fosters language learning. PLT also facilitates student interaction and collaboration and helps learners learn the culture along with the target language in a meaningful and enjoyable context. In addition, the students and the teacher collaboratively co-create new content using the target language.

Unlike previous research, Derwing, Rossiter, Munro, and Thomson (2004) measured the variable of fluency together with comprehensibility and accent. The participants of their study were twenty high-beginner

Mandarin-speaking English as a second language (ESL) students (age range: 26-38 years, $m = 33.4$ years) who had been living in Canada for more than six months and were all taking full-time English courses. They completed three tasks: a picture narrative, a 2-minute monologue and a dialogue.

The raters were twenty-eight native speakers (age range 21-52, $m = 28.6$), all undergraduate students in an ESL course at the Faculty of Education of the University of Alberta. The rating procedure consisted of several steps. The raters judged fluency, comprehensibility, and accent using the 9-point Likert scale from previous studies (Derwing & Munro, 1997; Munro & Derwing, 1999). The findings showed that the participants performed worst in the picture description, while they had better scores in the monological and dialogical tasks. Results for fluency ratings confirmed cross-task variation in the perception of L2 speakers' fluency. The raters assigned lower fluency scores to the learners' production in the picture-description task while the ratings for the monologue or the dialogue were significantly higher. One possible explanation was that the picture narrative was more constrained than the other two tasks because the participants had to describe the story that appeared in the cartoons. They had less freedom to use familiar vocabulary or structures. However, in the monologue and dialogue, students had more control of the content, and they could rely on previous experience and the other speaker for the conversation. Results of the study confirmed that "comprehensibility and fluency ratings were highly correlated; fluency was more strongly related to comprehensibility than to accent" (Derwing et al., 2004: 656).

Galante and Thomson (2017) carried out a study in Brazil with 24 English as a foreign language (EFL) learner (L1: Portuguese, L2: English). This study was groundbreaking in the field because it used quantitative data and native raters to assess the influence of PLT on the participant's oral performance. There were two groups, a treatment, and a control group. The treatment group took part in a PLT English language program while the control group received a traditional EFL course. The researchers measured the oral skills of fluency, comprehensibility, and accent by using five different tasks: two picture narrations (in one the students had to use the first-person, and in the other the third-person), a video narration, a role-play, and a monologue. The tasks were administered before the beginning of the intervention and after the completion of the program (after four months). Speech samples were recorded and presented to thirty raters. The raters were 30 untrained native English speakers from Canada (3 m, 27 f), who listened to the samples from the participants and assigned numerical values to them. Results of the study show that PLT instruction

could “lead to significantly larger gains in L2 English oral fluency relative to more traditional communicative EFL instruction” (Galante & Thomson, 2017: 1).

The vast majority of research on the use of PLT in the classroom has been carried out with English as the target language and there are few documented research studies and experiences with PLT in Chinese language classrooms (Corderi Novoa, 2015, 2019; Meng 孟柱亿 & Wan 万玉波, 2013; Wang 王永阳, 2009; Wen, 2015; Zhang 张连跃, 2013). However, most of them lacked reliable and well-designed research instrument and their results were merely based on subjective answers from simple questionnaires. No study to date has used quantitative research methods to investigate the impact of performative language teaching (PLT) in the Chinese as a foreign language (CFL) classroom.

Consequently, there is a need for studies with more rigorous experimental designs so that a link can be established between PLT instruction in the Chinese classroom and learners’ L2 development. The current paper tries to fill that research gap by carrying out a quantitative study on the impact of PLT on the oral skills (fluency, comprehensibility, and accent) of a group of Spanish learners of CFL.

3.2 Oral Skills in an L2: Main Constructs

Speaking is one of the most challenging skills language learners have to face and also the one ability that urgently needs to be improved (Bueno, Madrid & McLaren, 2006; Segura Alonso, 2013). In many countries, such as China and Spain, learners spend many years learning English as an L2. Still, most of them will not achieve oral fluency by the time they graduate from high school or university. Several possible reasons exist, one of them being the type of instruction. Because grammar has a long-written tradition, teachers must spend their classroom time teaching their students how to write and read in an L2. This teaching style focuses on memorization and repetition. Teachers often speak most of the time while students listen without having the opportunity to use the target language while at school. In addition, in a foreign language setting there usually a lack of quality input and few opportunities to use the L2 outside the classroom. Moreover, a lack of motivation and other factors can contribute to the low level of L2 speaking skills many students have.

When assessing oral skills, three main constructs are usually assessed, namely, fluency, comprehensibility and accent. There are several definitions of the construct fluency in the literature. Some authors defined it as “the capacity to use language in real-time, to emphasize meanings” (Skehan &

Foster, 1999: 96). Others described fluency as the ability to communicate efficiently (Koponen & Riegenbach, 2000). Derwing et al. (2004) described oral fluency as the fluidity or ease with which the L2 is spoken, and the perception of listeners of fluency in speech. In our research, we will consider L2 fluency to be a perceptual phenomenon and focus on perceived fluency, following Derwing et al. (2004), Derwing., Munro., Foote, Waugh, and Fleming (2014), and Galante and Thomson (2017). These researchers determined that the ratings from phonetically unsophisticated listeners were reliable and valid. Someone would be perceived as very fluent if the words just flow with no struggle. On the contrary, someone who is not perceived as fluent would have difficulty expressing their ideas or communicating effectively.

As for comprehensibility, Munro and Dewing's (1995a) also refer to perceived comprehensibility, that is how much effort the listeners have to employ to process or understand speech stimuli. Moreover, Munro and Derwing (1995b) highlighted the idea of comprehensibility as how easy L2 speech is for a listener to understand. Later, Derwing and Munro (1997: 2) defined perceived comprehensibility as "judgments on a rating scale of how difficult or easy an utterance is to understand". In addition, Derwing, Munro, and Thomson (2008) added, "comprehensibility is the ease or difficulty with which a listener understands L2 accented speech" (p. 360). All these authors agree that the concept of comprehensibility is different from intelligibility.

Munro and Derwing (1995b) operationalized comprehensibility by asking untrained raters to listen to speech stimuli and assign perceived comprehensibility judgments using a 9-point Likert scale, in which 1 = extremely easy to understand, and 9 = impossible to understand. Other studies, mostly from the same research group (Derwing & Munro 1997; 2013; Derwing, Munro, & Thomson, 2008; Derwing et al., 2004; Galante & Thomson, 2017; Munro & Derwing 1995c; Munro & Derwing, 1999) have followed this methodology to measure perceived comprehensibility. They also determined that the ratings obtained from phonetically unsophisticated raters were reliable and valid.

Having an accent is a typical consequence of L2 learning, especially when most learners start acquiring the L2 after early childhood (Piske, MacKay, & Flege, 2001). Munro and Derwing (1995b: 289) define foreign accent as "speech that differs in some noticeable respects from native speaker pronunciation norms". Derwing and Munro (2005: 379) point out that accent is also a "complex aspect of language that affects speakers and listeners in both perception and production and, consequently, in

social interaction”. This study adopts the definition of accent in Munro and Derwing (1995b): how different an L2 speaker’s productions are from a local variety. In our study, the standard for a “local variety” we use is standard Mandarin Chinese accent.

Once we have defined the three constructs that will be measured in our study (oral fluency, comprehensibility, and accent), we will focus on the importance of tasks in SLA (Second Language Acquisition).

3.3 The Importance of Tasks in SLA

Tasks have become central to L2 acquisition research and pedagogy. Researchers can manipulate different variables to test specific theoretical claims, and teachers can employ them and promote L2 acquisition and use (García Mayo, 2007; Ellis, Skehan, Li, Shintani, & Lambert, 2020). There is extensive research that shows that tasks are essential for L2 development (Ellis, 2003; Long, 2015). Moreover, as tasks promote interaction, they generate more language learning opportunities. Tasks “prioritize meaning but do not neglect form” (Ellis et al., 2020: 12).

Research on tasks has examined different variables that have an impact on L2 oral performance. Various studies have shown that the lexical and structural complexity, fluency, and accuracy of learners’ output can be influenced if different task design features are manipulated. Thus, researchers have studied the effects of planning time (Ellis, 2005), task familiarity (Samuda & Bygate, 2008), task complexity (Gilbert, 2005), and interaction (Mackey & Goo, 2007) on learners’ oral production. Several studies have showed that there were lower fluency scores in the learners’ production in picture-description tasks compared with monologues or dialogues (Derwing et al., 2004). Also, Mora and Valls-Ferrer (2012) observed that fluency scores were better at dialogic tasks compared to monologic speech.

This study used four different tasks: first-person picture narration, third-person picture narration, monologue, and role-play. Following Munro and Derwing (1995b), we did not ask participants to read passages or sentence stimuli aloud; instead, we chose to use tasks in which the participants had to produce spontaneous speech closer to real life. Reading-aloud tasks are not suitable for our study because these tasks “are not necessarily representative of learners’ productions when they must retrieve vocabulary and grammar” and “productions in a word list or sentences may not generalize to spontaneous speech” (Thomson & Derwing, 2014: 11). However, tasks such as picture descriptions, monologues, and improvised role-play dialogues are indeed suitable because they produce speech that

better reflects naturally occurring output, and, therefore, allow us to examine the oral speaking skills under more realistic circumstances.

4. Methodology

The study had a pre-test, treatment, and post-test design with a treatment group using PLT techniques for four months and a control group that followed a traditional method also for four months. For both groups, classes were held three times a week; each class lasted 90 minutes. In total, each group had 72 hours of instruction.

The traditional classes were taught from January 2017 to May 2017. The textbook was the *New Practical Chinese Reader 3* (Liu 刘珣, 2003), which book follows the Chinese traditional teaching methods: the focus was on forms, with memorization of grammatical structures and lists of new words. There were frequent reading exercises, as well as listening to texts. Some multimedia videos and songs were played to the students. In the classroom, frequent drilling exercises were used to help students with their tone and pronunciation.

The PLT classes were taught from January to May 2018. Some of the PLT activities were created by the author or adapted from other PLT workshops. Common PLT activities were: simulation, role-play, Improvisation, etc. In most classes, students were asked to perform in front of their classmates using Chinese. The focus was on oral communication and interaction. Usually, there was a warm-up with the use of words and physical movement/games to engage students at the beginning of the class. There were also multimedia videos of commercials, short films, TV shows, etc., used as a pretext to create a context for students to act in Chinese. In the second part of the class, students were exposed to several improvised PLT activities. The classroom atmosphere was relaxed and encouraging.

4.1 Rating Procedures

In previous studies, Derwing et al. (2004) and Galante and Thomson (2017), the researchers took excerpts of initial speech samples from the beginning of the picture descriptions and the monologue tasks from the participants and prepared the stimuli for the raters. However, in our study, we decided to use full speech audio files for the speaking tasks. This methodology was chosen as a way to fulfill our goal to obtain more precise measurement.

Also, in past studies, all the raters had to be physically present in a room at the same time, where they had to listen to all the stimuli and only

had a few seconds to choose a number in the scale and assign it to each speaker. However, in our study, we used integrated online forms that could be accessed at any place in the world, at any time via the Internet. Using the most popular Chinese chatting app, WeChat, we contacted the raters using our personal contact list. We sent a message in Chinese, inviting them to participate in the rating process. In the message, there was a link of one form with a link for one task. Once they had completed the task, they were sent another for Task 2, then Task 3, and finally Task 4. This whole process could take from several days to a couple of weeks, depending on how busy the raters were. Following Derwing et al. (2004) and Galante and Thomson (2017), we used three 9-point scales for each sample, fluency (1 = very fluent to 9 = very dysfluent), comprehensibility (1 = very easy to understand to 9 = very hard to understand), and accent (1 = no foreign accent to 9 = very strong foreign accent).

At the beginning of each online form, we added recommendations to the raters to tell them how to assess the audio recordings. We wrote instructions on how to assess fluency, comprehensibility, and accent. Derwing et al. (2004) had previously used this strategy of adding clarification for raters. These instructions were necessary so that the raters did not confuse fluency with proficiency, and, for example, to remind them that the highest rating was 1, and the lowest was 9. In previous studies, researchers reported that such instructions were sufficient and resulted in reliable ratings.

4.2 Speaking Tasks

Following Derwing et al. (2004) and Galante and Thomson (2017), the following speaking tasks were used:

Task 1: First-person picture narration of a boy who is playing soccer with friends and is injured.

Task 2: Third-person picture narration of a boy who goes back home to eat with his parents.

Task 3: Monologue about describing their favorite city.

Task 4: Improvised role-play dialogue with the teacher (the author of the study) in which the student welcomes a foreign Chinese learner into the country played by the teacher.

Learners performed the first three tasks individually. Each student was given a paper with each task and instructions on how to do it. The same task was given to the participants twice, one for the pre-test and another for the post-test. The author of this research (who was also the teacher) recorded the students' oral production. Participants produced output speech individually in Tasks 1, 2, and 3. However, in Task 4, learners

have to interact with the author in a dialogue. Although the measurement did not include interaction for Tasks 1, 2, and 3, the intervention and teaching methodology of the treatment group included PLT activities that foster interaction and collaboration. The main argument we want to put forward is that PLT would help the treatment group improve oral skills and that improvement would be reflected in the findings from the post-test. Therefore, we expect that the four-month training in PLT of the treatment group will benefit the students both in the monologic and the dialogic tasks. In what follows, the four tasks will be described in detail.

4.3 Research Questions

This study addresses the following research questions:

(1) Do learners in a PLT Chinese-as-a-foreign-language program experience greater gains in their oral performance than learners in a non-PLT Chinese course? Specifically, do they experience more significant gains in fluency, comprehensibility, and accent?

(2) Do fluency, comprehensibility, and accent vary across different speaking tasks before and after the treatment (PLT program)?

Based on the literature review, our first hypothesis was that following a PLT program would improve the learners' general oral skills in Chinese. We also expected more improvement in fluency than in comprehensibility. Accent would be the least affected construct (Derwing et al., 2004; Derwing & Munro, 2013; Galante & Thomson, 2017). Our second hypothesis was that there would be cross-task variation in oral performance.

4.4 Data Collection and Data Processing

As mentioned above, the participants performed the first three tasks individually, and the fourth one with the teacher. We gave the learners one minute to prepare and become familiar with each task before starting the tests. If they needed to ask questions about the format or requirements of the test, they could do so before starting the recording. A Huawei GRA-UL10 high-quality digital audio recorder was used to obtain the speech samples. Each student produced eight digital audio recordings (four tasks in the pre-test and four tasks in the post-test). As there were 16 participants, we obtained a total of 128 digital recordings. The recordings of Tasks 1, 2, and 3 were each approximately one minute long, and the recordings of Task 4 were one to two minutes long, depending on the student. Therefore, there were between 128 and 160 mins. (2-3 hours) of raw digital audio recordings.

Figure 1 displays a flow chart summary of the data gathering and processing procedure:

Figure 1. Data collection, processing, and analysis Flow Chart



The data gathering process could be summarized in the following steps:

- (1) Digitally record students' speech samples.
- (2) Process audio files with Audacity software.
- (3) Upload the processed audio files to SoundCloud.
- (4) Create online forms with pictures, text, and embedded audio links using JotForm.
- (5) Send the online forms to the raters via WeChat.
- (6) Raters complete the online forms.
- (7) The system sends data to Gmail and Hotmail accounts and Google Drive - Google Sheets.
- (8) Export Google Sheets data to MS Excel.
- (9) Export MS Excel data to STATA statistical software for further analysis.

4.5 Setting and Participants

4.5.1 Location of the Study

Even though laboratory research can be informative, we agree with Thomson and Derwing (2014) in the following: “[...] regarding ecological

validity, the ideal study should be conducted in a classroom” (p. 2). The current study was conducted in the metropolitan city of Vigo, on the Atlantic coast of Northwest Spain, in the Autonomous Region of Galicia. The research was carried out at the Official Language Public School EOI Vigo, a well-established and recognized institution which offers courses in German, Chinese, Japanese, Italian, French, English, Portuguese, Spanish as a foreign language, and Galician. In 2018, there were 6502 students and 102 language teachers. The Chinese department had approximately 100 enrolled students and three teachers.

Escuelas Oficiales de Idiomas (EOI) are public Official Language Schools funded by each Spanish regional education government where adults (fourteen years and older) can study languages part-time for a reduced fee. In the European Union, the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) establishes the different proficiency levels learners can attain from the A1 level (beginner) to the C2 level (proficiency).

4.5.2 Participants

The sixteen participants of the study were the students of two different classes at Intermediate 2 (B1.2) level at the EOI in two different school years (2016-2017 and 2017-2018). They signed a consent form. Table 1 below presents information about the traditional and the treatment groups.

Table 1: Participants in the control group vs. the treatment group

Number of students	The control group (Traditional CFL program)	The treatment group (PLT CFL program.)
		8 students; 3 f, 5 m.
Average age	41.5 years (range=28-53)	38 years (range = 27-54)
GPA (Grade Point Average in Chinese language proficiency) of B1.1 level	8.86 (range 8-10) (highest grade: 10)	9 (range 7-10) (highest grade: 10)
Time of intervention	From January 2017 to May 2017	From January 2018 to May 2018

The learner’s age range was 27-54. This kind of age group is the most common for learners of Chinese at EOIs. Both groups had similar grade point averages before the PLT treatment started, this being very important if any improvement is to be attributed to the new methodology. Statistical tests

also corroborated that both groups were similar at the beginning regarding their oral skills in Chinese.

The same teacher, the first author of the study, taught both groups for two academic courses (2016-2017 and 2017-2018). Thus, improvement in either the treatment or the comparison group could not be attributable to a change of teacher.

Seventy five (n= 75) untrained raters, native speakers of Chinese, were contacted to judge the speech samples generated by the participants. The number of raters in the current study was much larger than in previous studies in the field. As for gender, there were 64 female raters (85.33%), and 11 male raters (14.67%). Their average age of the raters was 32.56, with a range of 21-55. Regarding the raters' current location, 50.67% lived in Spain, 29.33% in Portugal, and 20% were located in the USA. Regarding the raters' profession, 62 (67%) were Chinese language teachers, and 25 (33%) were graduate students in linguistics. Therefore, 88% of the raters have direct connections with Chinese language teaching and learning. The remaining 12% had other types of jobs. The raters were not paid for their participation in the study.

5. Results

The CFL learners' speech sample ratings were split into three categories: fluency, comprehensibility, and accent. We computed Cronbach's alpha coefficients for all 128 speech samples for each scale. The coefficient scores were .77, .78, and .75, respectively (a reliability coefficient of .70 or higher is considered acceptable in most social science research situations; UCLA Institute for Digital Research & Education, 2020). With this data, we measured the inter-rater agreement, i.e., internal consistency (across raters) in the assessment of speech samples between different observations of a group. Besides, the raters' scores for fluency, comprehensibility, and accent for each speech sample were then pooled across tasks to find the mean rating for each item.

After presenting the descriptive results of the ratings, we performed statistical tests to assess if the differences in mean scores across groups and time for each category of oral performance were statistically significant. The main statistical tests used were the Analysis of Variance (ANOVA) and t-tests. Three partially repeated measures ANOVA for fluency, comprehensibility, and accent, separately, with Task (four levels, Task 1, Task 2, Task 3, Task 4) and Time (two levels, T1 and T2) as within-subject factors and the group as a between-subject factor were carried out. When

the ANOVA test revealed significant differences, then we carried out post-hoc independent and dependent samples t-tests.

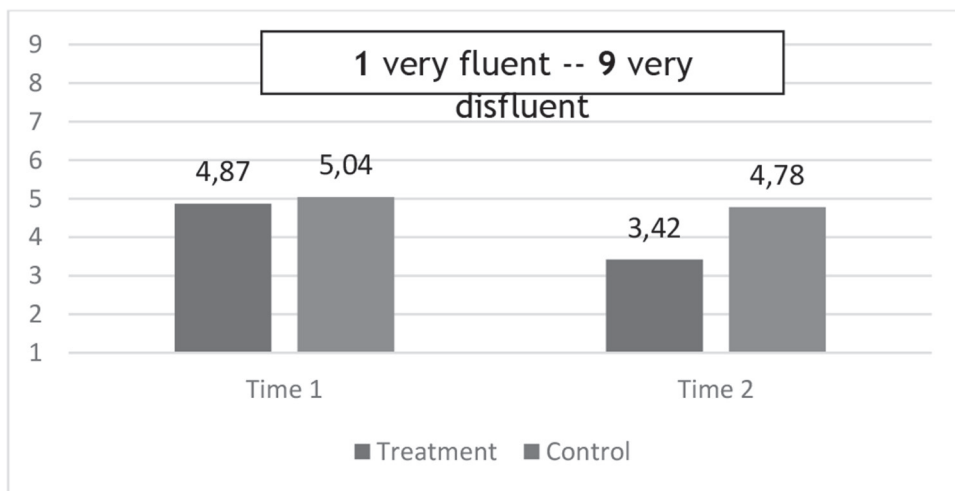
We will also present the 95% confidence interval to compare the PLT's impact across the three constructs of oral proficiency measured in this study. The confidence interval provides a range of values that have a 95% probability of containing the value of the impact of PLT.

5.1 Analysis of the Impact of a PLT Program on Learners' Fluency in CFL

Graph 1 features the observed mean fluency scores of the two groups across time. The blue bar represents the treatment group, and the red bar represents the control group; the fluency score ranged between 1 and 9, where 1 is very fluent, and 9 is very disfluent. T1 represents how both groups performed at the baseline (before the instruction program), and T2 represents how both groups performed after the program.

Graph 1 illustrates the observed mean fluency scores of the two groups across time.

Graph 1: Mean scores for fluency ratings across groups and time



The descriptive statistics showed that at T1, the treatment group had a fluency score slightly better than the control group, but the difference was close to zero. The gap widened at T2, where the treatment group had a better fluency score relative to the control group (the difference is 1.36). The lower the rating, the more fluent the performance of the learners.

Therefore, a decrease in the ratings over time indicates an improvement. Changes in fluency ratings across time for each group were as follows: the control group's scores decreased by 0.26, and the treatment group reduced by 1.45. Consequently, the treatment group's improvement over time was 1.19 (1.45 - 0.26) higher than the control group

The ANOVA tests revealed that there were significant differences in learner's fluency performance across time ($p < .001$), and across groups and time ($p < .001$). We found no significant differences across tasks and time nor groups. To further validate the ANOVA results suggesting different fluency performance across groups and time, we conducted post-hoc Bonferroni-adjusted independent samples t-tests.

When comparing the fluency scores between groups at T1 (before the start of language instruction), the t-test showed no significant difference between groups at the beginning of the program ($p = 0.592$). This result suggests that both the control and treatment groups had similar fluency levels before the start of their program of language instruction. However, at T2 (after the instruction program), the t-test showed that the observed difference in fluency scores between the treatment and control group was statistically significant ($p < .001$). The confidence interval for the gap ranges from 0.83 to 1.89.

When assessing within-group performance, how students within each group performed over time, the t-tests indicated a significant development in fluency scores for students in the treatment group from T1 to T2 ($p < .001$). However, if we analyze results for the learners in the control group, their scores did not change significantly over time ($p = .433$). These results validate the hypothesis that the PLT course had a positive impact on the participants' fluency.

In summary, results revealed a clear improvement for the treatment group in fluency scores, while the comparison group did not improve. Overall, the test suggests that the observed difference in fluency scores at T2 is statistically significant, which means that the PLT instruction program effectively improves fluency levels.

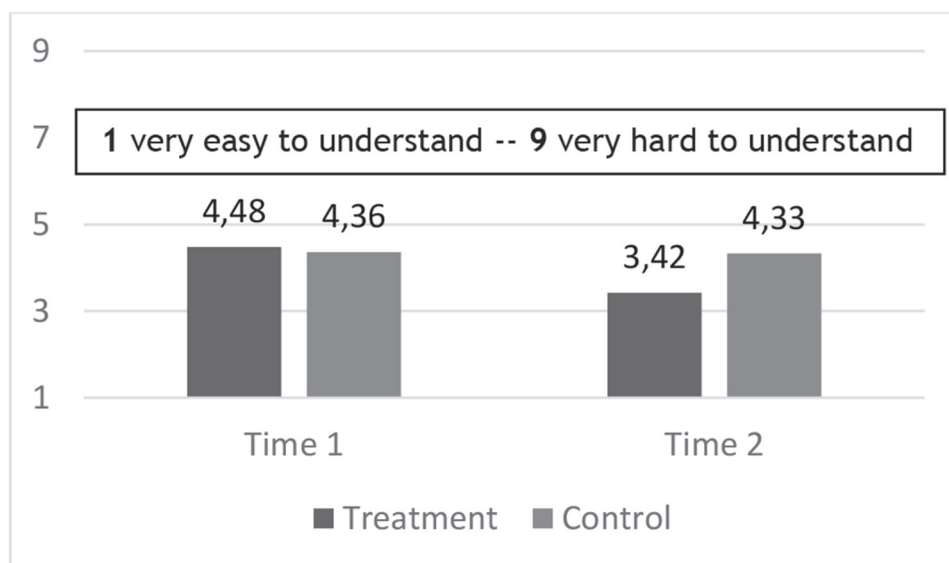
5.2 Analysis of the Impact of a PLT Program on Learners' Comprehensibility in CFL

Graph 2 presents the observed mean comprehensibility scores of the two groups across time. The comprehensibility score ranged between 1 and 9, where 1 is very easy to understand, and 9 is very hard to understand. The descriptive statistics show that at T1, the treatment group had a comprehensibility score slightly worse than the control group, but the

difference was close to zero. In contrast, at T2, the treatment group had a better comprehensibility score relative to the control group (the difference is 0.91). Similar to variations in fluency, the lower the rating, the more comprehensible the participants were. Therefore, a decrease in the scores over time indicates an improvement. Changes in ratings across time for each group were as follows: the control group's scores decreased by 0.03, and the treatment group reduced by 1.06. Consequently, the treatment group's improvement over time was 1.03 (1.06 - 0.03) higher than the control group.

The ANOVA test revealed significant differences in comprehensibility ratings across time ($p = .001$), and across time and groups ($p = .002$). We found no significant differences across tasks and time nor across groups. To further validate the ANOVA results suggesting different comprehensibility performance across groups and time, we conducted post-hoc Bonferroni-adjusted independent samples t-tests. Graph 2 below illustrates the observed mean comprehensibility scores of the two groups across time.

Graph 2: Mean scores for comprehensibility ratings across groups and time



When comparing the comprehensibility scores between groups at T1, the t-test showed no significant difference between groups at the beginning of the program ($p = .802$). Results suggest that both the control and treatment groups had similar comprehensibility levels before the start of their program of language instruction. After the instruction program,

that is, at T2, the t-test showed that students in the treatment group had better comprehensibility levels than learners in the comparison group ($p < .001$). The confidence interval for the difference ranges from 0.37 to 1.43.

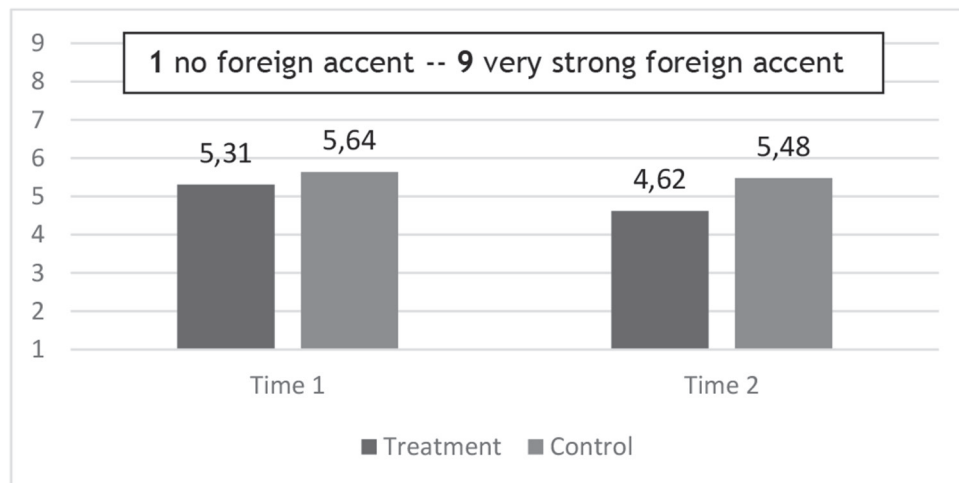
When assessing within-group performance, how students within each group performed over time, from T1 to T2, the t-tests revealed a significant development in comprehensibility scores for students in the treatment group ($p < .001$). In contrast, the mean comprehensibility scores for students in the control group did not differ significantly over time ($p = .917$). These results validate the hypothesis that the PLT course had a positive impact on student's comprehensibility performance.

In summary, results revealed a clear improvement for the treatment group in comprehensibility scores while the comparison group did not improve. Overall, the observed difference in comprehensibility scores at T2 is statistically significant, which means that the PLT instruction program effectively enhanced comprehensibility levels.

5.3 Analysis of the Impact of a PLT Program on Learners' Accent in CFL

The descriptive statistics show that, at T1, the treatment group had an accent score slightly better than the control group, but this difference was not statistically significant. At T2, this difference widened. The treatment group had a better accent score than the control group (the difference is 0.86). Similar to variations on fluency and comprehensibility, the lower the rating, the less foreign accent the learners had. Therefore, a decrease in the scores over time indicates an improvement. Changes in ratings across time for each group were as follows: the control group's ratings decreased by 0.16, and the treatment group reduced by 0.69. Consequently, the treatment group's improvement over time was 0.53 ($0.69 - 0.16$) higher than the control group. Graph 3 displays the observed mean accent scores of the two groups across time.

Graph 3: Mean scores for accent ratings across groups and time



The ANOVA test revealed significant differences in accent ratings across time ($p < .001$), and across time and groups ($p = .03$). No significant differences were found across tasks and time nor across groups. To further validate the ANOVA results suggesting different accent performance across groups and time, we conducted post-hoc Bonferroni-adjusted independent samples t-tests.

When comparing the accent scores between groups in period T1, the t-test showed no significant difference between groups at the beginning of the program ($p = .169$). This result suggests that both the control and treatment groups had similar accent levels before the start of their program of language instruction. After the instruction program (T2), the t-test shows that accent levels among learners in the treatment group were significantly better compared to the learners in the comparison group ($p < .001$). The confidence interval for the difference ranges from 0.37 to 1.43.

When assessing within-group performance, how students within each group performed over time, from T1 to T2, the t-tests revealed a significant development in accent scores for students in the treatment group ($p = .005$). In contrast, the mean accent scores for students in the control group did not differ significantly over time ($p = .479$). These results validate the hypothesis that the PLT course was beneficial for the student's accent.

In summary, results from the assessment of accent scores indicated that this dimension of oral proficiency improved for the treatment group, but not for the comparison group. Overall, the tests suggest that the

observed difference in accent scores at T2 is statistically significant, which means that the PLT instruction program effectively improved accent levels for the treatment group.

5.4 Analysis of Oral Performance by Task Type

We present the disaggregated descriptive information of oral performance ratings by task type. We first compared and analyzed the mean differences for fluency, comprehensibility, and accent at T1 and T2 in the mean scores for the control group, the treatment group, and the control vs. treatment group. With this first analysis, our goal was to obtain information about how different groups performed in various tasks at different times.

In addition, to find out whether the results were statistically significant, we carried out post-hoc Bonferroni-adjusted paired samples t-tests to compare differences in fluency, comprehensibility, and accent across speaking tasks. We took into account the group size, the different times, and the task type.

5.4.1 Analysis of fluency performance by task type

We analyzed the differences between the mean scores in fluency in all the tasks at T1 and T2. Table 2 displays the mean fluency scores and standard deviations at T1 and T2 for the control and the treatment groups.

Table 2: Mean fluency scores in Time 1 and Time 2 for control and treatment groups

Fluency	Control				Treatment			
	Time 1		Time 2		Time 1		Time 2	
	Mean	Standard Deviation	M	SD	M	SD	M	SD
Task 1	6.38	1.08	6.02	0.79	5.47	1.37	2.49	0.42
Task 2	4.68	1.24	4.61	1.07	5.08	0.33	3.81	0.67
Task 3	5.20	1.02	4.50	0.70	4.50	1.25	3.72	1.12
Task 4	3.90	1.21	4.02	1.11	4.44	0.99	3.68	0.83

In what follows, we present a comparison regarding fluency in each task for each of the groups.

At T1 the performance of both the control and treatment groups were very similar. At T2, there is a definite improvement for the treatment group, which outperforms the control group in all the tasks. For example, in Task 1, there is a very significant difference between the treatment group's results (2.49) and the control group (6.02).

We carried out post-hoc Bonferroni-adjusted paired samples t-tests to compare differences in fluency across speaking tasks. Overall, Task 4 (improvised role-play dialogue) produced significantly better fluency results than Task 1 (first-person picture narration) $t(14) = -5.88, p < .001$. When comparing fluency across tasks at each time, learners' performance in Task 4 was significantly more fluent than in Task 1 at T1 $t(14) = -4.93, p < .001$. We did not find any other statistically significant differences at each time. Finally, training significantly improved the performance of the treatment group in Task 1 and Task 2 when compared between T1 and T2 with $t(14) = 5.87, p < .001$, and $t(14) = 4.82, p < .001$, respectively. Additional t-tests were carried out to compare differences in accent across tasks by group between each time.

5.4.2 Analysis of comprehensibility performance by task type

We analyzed the differences between the mean scores by comprehensibility in all the tasks at different times. Table 3 shows the mean comprehensibility scores and standard deviations in T1 and T2 for the control and the treatment groups.

Table 3: Mean comprehensibility scores in Time 1 and Time 2 for control and treatment groups

Comprehensibility	Control				Treatment			
	Time 1		Time 2		Time 1		Time 2	
	M	SD	M	SD	M	SD	M	SD
Task 1	5.28	1.29	5.08	1.73	4.77	1.43	2.87	0.84
Task 2	4.12	0.81	4.08	0.72	4.48	0.83	3.64	0.81
Task 3	4.43	0.62	4.33	0.79	4.49	1.36	3.72	1.28
Task 4	3.62	0.86	3.83	0.90	3.98	0.70	3.48	0.80

At T1, we can see that the performance of the control and treatment groups were very similar. At T2 there is a definite improvement for the treatment group, which outperforms the control group in all tasks. Specifically, in Task 1 there is a very significant difference between the treatment group's results (2.87) and the control group (5.08).

We carried post-hoc Bonferroni-adjusted paired samples t-tests to compare differences in comprehensibility across speaking tasks. Overall, Task 4 (improvised role-play dialogue) led to significantly better comprehensibility results than Task 1 (first-person picture narration) $t(30) = -1.90, p = .039$. When comparing comprehensibility across tasks at each time, Task 4 was significantly more comprehensible than Task 1 only

at T1 $t(14) = -2.806, p = .007$. We did not find any other statistically significant differences at each time. Finally, training significantly improved the performance of the treatment group in Task 1 when compared between T1 and T2 with $t(14) = 3.24, p = .007$. Additional t-tests were carried out to compare differences in comprehensibility across tasks by group between each time. To summarize, results for tests of comprehensibility indicate that this dimension of oral proficiency improved for the treatment group, but not for the comparison group. We found a difference in comprehensibility scores for Task 4 relative to Task 1 at T1. Training appeared to differentially impact performance in Task 1 for the treatment group.

5.4.3 Analysis of accent performance by task type

We analyzed the differences between the mean scores for accent in all the tasks at different times.

See Table 4 for a description of mean accent scores at T1 and T2 for the control and the treatment groups:

Table 4: Mean accent scores in Time 1 and Time 2 for control and treatment groups

Accent	Control				Treatment			
	Time 1		Time 2		Time 1		Time 2	
	M	SD	M	SD	M	SD	M	SD
Task 1	6.49	0.84	6.23	1.00	5.66	1.35	4.21	0.69
Task 2	5.27	0.65	5.05	0.67	5.23	0.66	4.60	0.89
Task 3	5.78	0.56	5.67	0.64	5.40	1.27	4.91	1.10
Task 4	5.04	0.71	5.01	0.68	4.95	0.80	4.77	0.83

Once more, at T1 the performance of both the control and the treatment groups were very similar. The treatment group performed better in all the tasks but the difference was not very substantial. At T2 the treatment group scores were better than those in the control group in the four tasks. The control group had its best performance at Task 4 (5.01 vs. 4.77); however, there was a very substantial difference in scores for Task 1, with the treatment group's results (4.21) and the control group's scores (6.23).

We carried out post-hoc Bonferroni-adjusted paired samples t-tests to compare differences in accent across speaking tasks. Overall, Task 4 (improvised role-play dialogue) produced significantly better accent results than Task 1 (first-person picture narration) $t(30) = -2.16, p = .024$. When comparing accent across tasks at each time, accent was significantly better in Task 4 than in Task 1 only at T1 $t(14) = -2.806, p = .007$. We did not

find any other statistically significant differences at each time. Training significantly improved the accent performance of the treatment group in Task 1 when compared between T1 and T2 with $t(14) = 2.71$, $p = .021$. Additional t-tests were carried out to compare differences in accent across tasks by group between each time. In sum, results for tests of accent indicate that this dimension of oral proficiency improved for the treatment group, but not for the comparison group. The statistical results of accent performance across tasks are similar to those of comprehensibility.

6. Conclusions

Results of the current research show a significant improvement for the treatment group in oral fluency, comprehensibility, and accent compared to the comparison group. The improvement in fluency was greater than the improvement in comprehensibility and accent, which had similar degrees of development. The statistical tests suggest that the observed difference in scores at T2 is statistically significant, which means that the PLT instruction program was effective in improving the three constructs that were measured when assessing the students' oral skills. Our findings are also in line with Galante and Thomson's (2017) study, where the oral fluency and comprehensibility levels of the treatment group improved more than those of the control group, thanks to PLT. However, unlike in Galante and Thomson (2017), in our study accent also slightly improved. This finding could be explained because they used PLT techniques only 50% of class time and the rest of the class time followed a traditional format while, while in our study PLT was used 100% of the time in the treatment group.

Overall, participants obtained significantly better fluency, comprehensibility, and accent results in Task 4 (improvised role-play dialogue) than in Task 1 (first-person picture narration). Each time, the fluency, comprehensibility, and accent were compared across tasks, and the learners' performance in Task 4 was significantly better than in Task 1 at T1. Finally, training improved the treatment group's fluency performance in Task 1 and Task 2 when comparing T1 and T2. Moreover, for the treatment group, their comprehensibility performance improved most in Task 1. At the same time, training improved the accent performance of the treatment group in Task 1 when comparing T1 and T2. These findings are in line with Derwing et al. (2004) and Mora and Valls Ferrer (2002). To sum up, these results confirm our hypotheses and suggest that PLT positively impacts the development of oral skills, and that performance varies across task types.

Chinese language teaching around the world needs a methodological revolution and there is a need to develop a modern and more effective teaching and learning methodology for CFL. Specifically, students of Chinese need a more collaborative and engaging approach to improve their oral proficiency skills. On the basis of our findings, PLT could be claimed to be an alternative to more traditional methodologies.

We hope that more and more language teachers around the world start using the PLT methodology in their classrooms. Educational centers worldwide should promote the use of PLT in language courses. Textbook editors and writers should include PLT activities in their materials, and new PLT textbooks should be developed. We need more PLT teacher training seminars at a local, regional, and national level. Finally, we also hope that The Confucius Institute Headquarters will upgrade and change the design of the HSKK® oral exam, including a dialogue/interview with an examiner to assess the learner's oral proficiency skills in a more appropriate way.

REFERENCES

- Bentley, E., 1968. *What Is Theater? Incorporating the Dramatic Event, and Other Reviews, 1944-1967*. New York: Limelight Editions.
- Bueno, A., Madrid, D., & McLaren, N. (Eds), 2006. *TEFL in Secondary Education*. Granada: Editorial Universidad de Granada.
- Catterall, J. S., 2002. Research on drama and theater in education. In R. Deasy (Ed.), *Critical Links: Learning in the Arts and Student Academic and Social Development* (pp.58-64). Arts Education Partnership.
- Corderi Novoa, M., 2015. *Research study about the use of drama and improvisational theater games to teach Chinese as a second language*. Beijing Language and Culture University master's thesis.
- Corderi Novoa, M., 2019. Improchinese: Using improv theater in a Chinese as a foreign language experimental classroom. *Revista de Artes Performativas, Educación y Sociedad*, 1(1): 147-163.
- Council of Europe, (online). Common European Framework of Reference for Languages. Council of Europe Language Policy Portal, <https://www.coe.int/en/web/language-policy/home>, consultado el 12/05/2022.
- Derwing, T. M., & Munro, M. J., 1997. Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19: 1-16.
- Derwing, T. M., & Munro, M. J., 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39: 379-397.

- Derwing, T. M., & Munro, M. J., 2013. The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, 63: 163-185.
- Derwing, T. M., Munro, M. J., & Thomson, R. I., 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29: 359-380.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I., 2004. Second language fluency: Judgments on different tasks. *Language Learning*, 54: 655-679.
- Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J., 2014. Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, 64: 526-548.
- DICE Consortium, (online). The DICE has been cast. Research findings and recommendations on educational theater and drama. European Union. www.dramanetwork.eu, consultado el 03/05/2022.
- Ellis, R., 2003. *Task-Based Language Learning and Teaching*. Oxford: Oxford University Press.
- Ellis, R., 2005. Principles of instructed language learning. *System*, 33: 209-224.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C., 2020. *Task-based Language Teaching. Theory and Practice*. Cambridge: Cambridge University Press.
- Fleming, M., 2016. Exploring the concept of performative teaching and learning. In S. Even & M. Schewe (Eds.), *Performative Teaching, Learning, Research—Performatives* (pp.189-205). *Lehren, Forschen* Berlin: Schibri Verlag,
- Galante, A., & Thomson, R. I., 2017. The effectiveness of drama as an instructional approach for the development of second language oral fluency, comprehensibility, and accentedness. *TESOL Quarterly*, 51 (1): 115-142.
- García Mayo, M.P. (Ed.), 2007. *Investigating Tasks in Formal Second Language Learning*. Clevedon, U.K.: Multilingual Matters.
- García Mayo, M.P. & Alcón Soler, E., 2013. Negotiated input, output/interaction. In J. Herchensohn & M. Young-Scholter (Eds.), *The Handbook of Second Language Acquisition* (pp.209-229). Cambridge: Cambridge University Press.
- Gilabert, R., 2005. Task complexity and L2 narrative oral production. Universitat de Barcelona Doctoral dissertation, Spain.
- Kao, S., 1994. *Classroom interaction in a drama-oriented English conversation class of first-year college students in Taiwan: A teacher-researcher study*. Ohio State University Doctoral dissertation, USA.

- Koponen, M., & Riggensbach, H., 2000. Overview: Varying perspectives on fluency. In H. Riggensbach (Ed.), *Perspectives on Fluency* (pp. 5-24). Ann Arbor: University of Michigan Press.
- Liu, Xun 刘珣. 2003. 新使用汉语课本 3. *New Practical Chinese Reader 3*. 北京语言大学出版社. Beijing Language and Culture University Press.
- Long, M. H., 1996. The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). New York: Academic Press.
- Long, M.H., 2015. *Second Language Acquisition and Task-Based Language Teaching*. Oxford: Wiley - Blackwell.
- Mackey, A., & Goo, J., 2007. Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: a series of empirical studies* (pp. 407-453). Oxford: Oxford University Press.
- Meng, Zhuyi & Wan, Yubo 孟柱亿, 万玉波. 2013. 演话剧学汉语特色口语课程分析——以《春香新传》教学、演出为案例. Specialized speaking course “Learn Chinese by performing a play” - Taking the case of Chunxiang Xinzhuang teaching and acting as an example. 国际汉语教育 *Journal of International Chinese Education*, 2: 93-106.
- Mora, J.C., & Valls-Ferrer, M., 2012. Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46 (4): 610-641.
- Munro, M. J., & Derwing, T. M., 1995a. Processing time for native and foreign-accented speech. *Language and Speech* 38 (3): 289-306.
- Munro, M. J., & Derwing, T. M., 1995b. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45: 73-97.
- Munro, M. J., & Derwing, T. M., 1995c. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38: 289-306.
- Munro, M. J., & Derwing, T. M., 1999. Foreign accent, comprehensibility, & intelligibility in the speech of second language learners. *Language Learning*, 49 (1): 285-310.
- Piazzoli, E., 2018. *Embodying Language in Action: The Artistry of Process Drama in Second Language Education*. Palgrave Macmillan. Dublin: Ireland.
- Piske, T., MacKay, I. R. A., & Flege, J. E., 2001. Factors affecting the degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29: 191-215.
- Podlozny, A., 2000. Strengthening verbal skills through the use of classroom drama: a clear link. *Journal of Aesthetic Education*, 34 (3-4): 91-104.

- Samuda, V., & Bygate, M., 2008. *Tasks in Second Language Learning*. Basingstoke: Palgrave Macmillan.
- Schewe, M., 2013. Taking stock and looking ahead: Drama pedagogy as a gateway to a performative teaching and learning culture. *Scenario: Journal for Performative Teaching, Learning and Research*, 8 (1): 5-23.
- Schmidt, R., 1990. The role of consciousness in second language learning. *Applied Linguistics*, 11: 129-158.
- Segura Alonso, R., 2013. *The importance of teaching listening and speaking skills* (Unpublished master's thesis). Universidad Complutense de Madrid.
- Skehan, P., & Foster, P., 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49: 93-120.
- Stern, S., 1980. Drama in second language learning from a psycholinguistic perspective. *Language Learning*, 30: 77-100.
- Stinson, M., & Freebody, K., 2006. The DOL project: The contributions of process drama to improved results in English oral communication. *Youth Theater Journal*, 20: 27-41.
- Thomson, R. I., & Derwing, T.M., 2014. The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36: 326-344.
- UCLA Institute for Digital Research & Education, (online). What does Cronbach's alpha mean? Statistical Consulting, <https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/>, consultado el 06/05/2022.
- Vygotsky, L. S., 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wang, Yongyang 王永阳. 2009. 试论戏剧化教学法在汉语作为第二语言教学中的运用——以澳大利亚的一个课堂教学为例 On the application of drama in teaching Chinese as a second language - Classroom practice in Australia. *世界汉语教学 Chinese Teaching in the World*, 2: 233-243.
- Wen, I-C., 2015. *The effectiveness of learning through drama in teaching Chinese as a second language* (Unpublished master's thesis). Miami University. Oxford: Ohio.
- Wood, D., Bruner, J., & Ross, G., 1976. The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17: 89-100.
- Zhang, Lianye 张连跃. 2013. 戏剧在二语教学中的整合功能——基于一项海外汉语教学的案例分析 Drama's integration function in TCSL: A case study of Chinese language teaching at Bard College at Simon's Rock. *语言教学与研究 The Journal of Chinese Teaching and Research*, 1: 23-30.

