

January 2021

The Use of an Elicited Imitation Test to Measure Global Oral Proficiency of L2 Chinese at the Postsecondary Classroom

I-Huei Lee
University of Northern Colorado

Follow this and additional works at: <https://engagedscholarship.csuohio.edu/cltmt>



Part of the [Chinese Studies Commons](#), and the [Educational Assessment, Evaluation, and Research Commons](#)

[How does access to this work benefit you? Let us know!](#)

Recommended Citation

Lee, I-Huei (2021) "The Use of an Elicited Imitation Test to Measure Global Oral Proficiency of L2 Chinese at the Postsecondary Classroom," *Chinese Language Teaching Methodology and Technology*. Vol. 4 : Iss. 1 , Article 4.

Available at: <https://engagedscholarship.csuohio.edu/cltmt/vol4/iss1/4>

This Article is brought to you for free and open access by the Confucius Institute at EngagedScholarship@CSU. It has been accepted for inclusion in Chinese Language Teaching Methodology and Technology by an authorized editor of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

The Use of an Elicited Imitation Test to Measure Global Oral Proficiency of L2 Chinese at the Postsecondary Classroom

Cover Page Footnote

Footnote: The video is available at <https://www.youtube.com/watch?v=Gdt3l0X1bE8> Acknowledgements: I would like to thank Darren Ilett of University of Northern Colorado and Diane Neubauer of University of Iowa for assistance with this manuscript.

The Use of an Elicited Imitation Test to Measure Global Oral Proficiency of L2 Chinese at the Postsecondary Classroom Level

I-Huei Lee

University of Northern Colorado

Abstract

With a growing number of foreign language studies on proficiency outcomes, it is imperative to address the challenge of measuring students' proficiency development in a language program where standardized proficiency testing is not readily available. This article reports administering a Chinese elicited imitation test (EIT) by an instructor to track students' global oral proficiency development in a small language program in a mid-size U.S. public university. The test results from the EIT of second language (L2) Chinese suggest that this tool can provide the instructor with valuable insights into students' oral proficiency. This study also discusses the potential practical value of using this EIT in a language program with limited resources for standardized proficiency assessment. The hope is that this study will encourage language educators who are not already doing so to start using empirical evidence from a valid and reliable proficiency measurement tool to reflect on, improve, and guide their instructional practices.

Keywords: Chinese language proficiency testing; elicited imitation; L2 global oral proficiency; classroom-based research

Introduction

The current proficiency movement in U.S. foreign language teaching and testing can be traced back to the first publication of the ACTFL Proficiency Guidelines in 1986 (Malone and Tschirner, 2012; Martin, 2014). As more and more teachers have moved toward communicative proficiency-based language teaching, the field of foreign language education research has expanded into studies investigating proficiency outcomes across instructional contexts and course levels (Burkhauser et al., 2016; Davin et al., 2014; Fall et al., 2007; Magnan, 1986; Schmitt, 2016; Xu et al., 2015), exploring contributable factors (Strawbridge et al., 2019; Vyn et al., 2019; Winke & Gass 2018) and seeking to determine the relationship of proficiency between different modalities (Bernhardt et al., 2015; Hubert, 2013; Tschirner, 2016). It is important to conduct empirical studies on proficiency because test results are useful for setting and adjusting expected language benchmarks and for improving curriculum design and instruction (Goertler et al., 2016; Soneson & Tarone, 2019). For students, results from a valid and reliable assessment tool can help to identify areas for improvement. In the field of L2 Chinese, however, scarce attention has been devoted to oral proficiency assessment and classroom research (e.g., Liao, 2018; Yuan & Li, 2019).

While the studies mentioned above provide valuable information on students' proficiency outcomes and development, they were mostly conducted by external researchers. Furthermore, many of these studies used assessments such as ACTFL's Oral Proficiency Interview (OPI), Oral Proficiency Interview-Computer (OPIc), and Avant Standards-Based Measurement of Proficiency (STAMP) tests, or assessments independently designed by individual school districts. However, at the college level, language programs with scarce

resources and limited budgets might not afford to implement the OPI/OPIc to measure students' oral proficiency, especially when there is a need to administer it more than once. This action research was conducted in this kind of institutional constraint by an instructor wanting to use a language assessment tool with validity and reliability and can be feasibly administered to students under her teaching. To achieve this goal, this study adopted a Chinese elicited imitation test (EIT) developed by Wu and Ortega (2013), a quick and reliable measure that was demonstrated to be psycholinguistically valid to measure oral proficiency in second language acquisition research. This study presents and analyzes the scores obtained from administering that test, and then discusses the potential practical value of using this EIT for Chinese language programs with limited resources for oral proficiency testing. Based on the useful information provided by the EIT, the author recommends the use of the EIT to Chinese language teachers interested in knowing students' global oral proficiency but facing the challenge of lacking a valid and reliable measure. Given that this Chinese EIT has parallel versions in other languages, this study might be of interest to teachers of other target languages as well.

Review of the literature

Proficiency assessments in U.S. foreign language education

Researchers in foreign language education have used a number of performance and proficiency assessments to investigate learners' proficiency outcomes in a variety of U.S. K–16 contexts (see, e.g., Hubert, 2013; Tschirner, 2016; Watzinger-Tharp et al., 2018; Winke & Gass 2019; Xu et al., 2015). For example, the data in the study of Watzinger-Tharp et al. (2018) came from the ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL), mandated by the state of Utah to assess dual language immersion students' proficiency in the foreign language. In 2014 Michigan State University and the Universities of Minnesota and Utah launched a large-scale foreign language proficiency assessment project through the Language Flagship Proficiency Initiatives funded by the National Security Education Program. This project administered OPIc for speaking, ACTFL's Reading Proficiency Test (RPT) for reading, and ACTFL's Listening Proficiency Test (LPT) in these three large research-based universities at several curricular levels in a number of different languages over the course of three academic years (Winke & Gass, 2019).

In addition to ACTFL's suite of proficiency assessments, other studies based their data on Avant Assessment's STAMP tests (Burkhauser et al., 2016; Davin et al., 2014; Moeller & Theiler, 2014; Schmitt, 2016; Xu et al., 2015). For example, Xu et al. (2015) selected schools in the same suburban district and compared STAMP scores of fourth and fifth graders from an elementary Mandarin immersion program with students from two high schools at the fourth and fifth levels (Advanced Placement) of Mandarin language study. The findings revealed that the elementary immersion students performed slightly better than the high school students in the interpretive mode of reading but lagged slightly behind the high school students in the presentational modes of writing and speaking. They hypothesized that the open-ended nature of the writing and speaking items in the STAMP assessments may confer an advantage on the older students, who have higher levels of cognitive ability and likely have more experience in crafting answers to prompts.

Proficiency and performance tests developed by school districts have provided another important data source for researchers to investigate students' learning in a foreign language. For example, Vyn et al. (2019) analyzed a school district's language proficiency data in conjunction with the data from the survey and observations to investigate the effects of language teaching practices on student proficiency development. Contrary to an intuition that teachers' target language usage in lower-level courses is constrained by the low

proficiency of beginning students, their study showed that target language usage in beginning classes is critical because its impact on proficiency outcome is more prominent than in higher-level classes. Dzedzic (2012) drew on the data from a school district's assessment to compare proficiency outcomes from classes using TPR Storytelling (TPRS) with classes using traditional grammar-based instruction. The results showed that TPRS students performed significantly better in writing and speaking and there was no significant difference in listening and reading. It is an uncommon study on proficiency outcomes in that Dzedzic was not only the researcher but also the instructor of the classes involved in the study.

The studies reported above are predominantly large-scale and cross-sectional, and their data came from district-created assessments or from the commercial tests distributed by ACTFL or Avant Assessment. In the context of higher education, assessments developed by school districts are not available. In addition, administering commercial tests may not be feasible in some language programs because of the associated expense. Although language teachers might use communicative performance-based testing such as role playing, interviewing, or oral presentation to assess students' oral proficiency, these classroom-based oral performance tasks are rarely standardized. Moreover, in these performance-based speaking tasks, students could adopt avoidance, delaying, and circumlocution strategies, or even well-practiced interview and presentation skills, which all make this kind of assessment less reliable (Van Moere, 2012). Van Moere argued that "a complementary approach to communicative and psycholinguistic testing will undoubtedly lead to stronger and fairer assessments" (p. 340), because psycholinguistic factors such as processing and automaticity are also an integral part of communicative competence. The psycholinguistic testing Van Moere advocated for was elicited imitation (EI).

Use of elicited imitation to measure global oral proficiency in L2

Elicited imitation is a testing method that has been widely used to measure oral proficiency in a second/foreign language and was particularly popular in the 1970s and early 1980s (Yan et al., 2016). In an EI test, the test taker typically listens to a series of stimulus sentences and then repeats the sentences as exactly as possible. EI is perceived as a valid measure of language proficiency because the imitation is reconstructive in nature; that is, EI requires the test taker to process the language stimulus received. Therefore, imitation is hard to achieve by rote, particularly when the EI test includes stimuli of various lengths and complexities and focuses test takers' attention on meaning (Erlam, 2009). Researchers have demonstrated that EI is an effective measure of L2 oral proficiency, particularly when the test is time-pressured (Erlam, 2009; Kim & Nam, 2017).

Language proficiency as measured in EI demands integrative speaking and listening skills because it "requires both successful comprehension and production of the target language" (Gaillard & Tremblay, 2016, p. 422). However, language proficiency is not a construct with a single definition. Real-life functional competency is the kind of proficiency measured in the standardized language tests developed with reference to the proficiency guidelines put forth by ACTFL. ACTFL defines proficiency as "the ability to use language in real world situations in a spontaneous interaction and non-rehearsed context and in a manner acceptable and appropriate to native speakers of the language" (ACTFL 2015, p. 4). Hulstijn (2011) provides another conceptualization of language proficiency, with a distinction between basic language cognition (BLC) and higher language cognition. Basic language cognition involves what all adult native speakers have in common when producing spoken language in any situations of everyday life: the largely implicit, unconscious knowledge of phonology, morphology, and syntax; the largely explicit, conscious knowledge in the lexical domain; and the automaticity with which these types of knowledge can be processed. As the complement or extension of BLC, higher language cognition utterances pertain to both

written and spoken language and may vary widely among native speakers because of such variables as literacy, level of education, and so on. The EIT used in the present study taps Hulstijn's construct of BLC (Wu & Ortega, 2013).

It is noteworthy that the Chinese EIT administered in this study derived from a set of EITs originally designed by Ortega et al. (2002), in which versions of the same EIT format were developed in four different languages—English, German, Japanese, and Spanish—for the purpose of cross-linguistic comparison. Additional parallel versions in the languages of French (Tracy-Ventura et al., 2014) and Korean (Kim et al., 2016) are also currently available.

As described below, this study was conducted in the instructor's own Chinese language classes with a strong proficiency orientation to develop students' implicit linguistic knowledge. Therefore, using the Chinese EIT developed by Wu and Ortega (2013) aligns well with the instructor's interest in investigating the L2 learning outcome of global oral proficiency. This study addresses the following research questions:

- (1) What do the scores from the EIT of L2 Chinese tell about students' global oral proficiency in proficiency-oriented instruction?
- (2) What is the potential practical value of the EIT for a language program with limited resources for standardized testing of proficiency?

Materials and Methods

Setting

This study was conducted in a small Chinese language program with two faculty members in a midsize U.S. public university during the academic year of 19/20. The data were collected from three Chinese classes, CHIN-100, CHIN-201, and CHIN-202, which were all taught by the author of the article. The only standardized proficiency test implemented in the program was OPIc, which was required for Chinese minors, with an exit expectation of Intermediate-Mid. For students with prior Chinese learning experience, placement was decided based on an informal meeting with the student and the courses offered at the semester. In the meeting, the student was asked general questions about Chinese language learning experience and tested for language proficiency by having a conversation in Chinese with the instructor and reading aloud a Chinese passage for reading ability. In other words, no standardized test was used in the program for placement purpose or over the course of language study to periodically track students' proficiency outcomes.

The format of all three Chinese courses was three 50-minute classes per week for 16 weeks, and all were taught in a communicative, proficiency-oriented approach, with at least 90% of instructional time in Chinese. CHIN-100 was a stand-alone conversational class newly created by the instructor for students without any Chinese learning background and centered on developing proficiency in, primarily, listening and speaking and, secondarily, reading. The course did not use any textbook and was designed with reference to the NCSSFL-ACTFL Can-Do Statements (ACTFL, 2017) and the 150 high-frequency words from the vocabulary lists of Hanyu Shuiping Kaoshi, a standardized test of Chinese language proficiency for non-native speakers. CHIN-201 and CHIN-202 were, respectively, the third and the fourth semester of a four-semester sequence of beginning and intermediate Chinese courses. While textbooks were required by the program in CHIN-201 and CHIN-202, none of the practices or activities in the textbooks were used because the instructor did not consider them to be well aligned with communicative language teaching. Grammar was taught using the technique called "pop-up grammar," a quick explanation of grammar points when they naturally arise in context (Lichtman, 2014).

Note that starting from the second half of the spring semester in 2020, both CHIN-202 and CHIN-100 were taught remotely because of the coronavirus pandemic. Still, regardless of the online format, the goal of facilitating learners' language proficiency remained.

Participants

All of the students in the three small Chinese courses agreed to participate in the study. Out of the 13 participants, six came from CHIN-100. Of the remaining seven students, three were enrolled in both CHIN-201 (N = 4) and CHIN-202 (N = 6) in two consecutive semesters. None of the participants reported themselves as a heritage learner of Chinese or having study abroad experience in a Chinese-speaking country. All of the students were native speakers of English with the exception of two students who were native speakers of Thai.

Instrument and data collection

The instructor administered the Chinese EIT developed by Wu and Ortega (2013) to each student individually. This EIT consists of 30 grammatical sentences of increasing length and complexity and contains a wide range of vocabulary and grammatical structures. Comparing the EIT scores from lower-division courses with upper-division courses at a public university in the United States, Wu and Ortega's research showed that this EIT works well along a broad range of oral proficiency and is able to distinguish performances of lower- and higher-proficiency speakers of L2 Chinese. Their research also provided evidence of a significant correlation between performances on the EIT and on an oral narrative task, indicating that this EIT taps the same underlying global linguistic ability as measured by an oral narrative task but is much more practical in scoring than an oral narrative assessment. This EIT also effectively differentiated heritage learners from foreign language learners.

All the participants watched the YouTube video created by Reed Riggs in 2015.¹ The video read aloud each of the 30 Mandarin sentences in Wu and Ortega's EIT with an insertion of a few seconds of silence between hearing the sentence and repeating it for delayed repetition. The entire video lasted eight minutes and 54 seconds. Each participant's performance on the Chinese EIT was recorded for scoring and analysis. Note that while the instructor was also the tester, the teaching materials were independent from the test; consequently, there should not be a concern of teaching to the test.

For the three participants enrolled in both CHIN-201 and CHIN-202 consecutively, the EIT was administered three times, at the beginning and the end of CHIN-201 and at the end of CHIN-202. For the rest of the students enrolled in either CHIN-201 or CHIN-202, the pre- and post-test was administered at the beginning and at the end of their corresponding course. No pre-test was given to the CHIN-100 because all six students, except one, reported not formally studying Chinese in the past, and therefore the EIT was administered to this group of participants only at the end of the course. A total of 23 speech samples were collected.

To provide a deeper understanding of the EIT outcomes, a short, semi-structured interview about the learner's language background and experience was conducted with each participant individually right after the EIT he/she took the last time. Each interview was audio-recorded and transcribed by the researcher for analysis.

Scoring and Analysis

Students' recorded EI responses were scored through using the holistic five-point scoring rubric as provided in Wu and Ortega's (2013) study to quantify the accuracy of repetition, with four points given to verbatim repetition; three to accurate content repetition

with some changes in form; two to responses that included any changes in content and/or in form that affected meaning; one to responses that included repetition of half or less of the stimulus; and zero to silence, one word, or unintelligible repetition. With 30 test items, the total maximum possible score of the EIT is 120.

Due to logistical and funding constraints, each of the participants' speech samples was rated by the instructor without a second rater. Regarding the lack of evidence for inter-rater reliability in this study, EI performance in Wu and Ortega's (2013) validation study was rated solely by Wu after 15% of the responses were rated, and the agreement between Wu and a second rater was found to be 95%. This high percentage of agreement was also found in the Korean version of the EIT (Kim et al., 2016), and neither of the studies explicitly stated a requirement of formal rater's scoring training, suggesting that the scoring rubric is straightforward enough for an instructor to rate alone with reasonable reliability for practical purposes.

Descriptive statistical analysis was conducted to calculate the mean and standard deviation (SD) and identify the highest and the lowest scores in each class's pre- and post-test (only post-test for CHIN-100).

Results

In advance of the discussion of the descriptive statistical results of the EIT, a summary of each participant's language learning experience is presented in Table 1. *[Table 1 near here]* It is evident that the undergraduate language classes consisted of students with diverse language learning experiences. Additionally, most students in the beginner course of CHIN-100 were concurrently learning another language; two students in CHIN-202 (labeled S-202E and S-202F) did not follow the typical sequence of study in this program because CHIN-202 was canceled in the previous year due to low enrollment and therefore these two students took higher-level courses before coming down to complete the required CHIN-202.

The descriptive statistical results from the EIT are depicted in Table 2 to answer the first research question about the test scores of the EIT in a proficiency-oriented approach to second language instruction. *[Table 2 near here]*

For CHIN-100, the mean score on the EIT was 26.17 with a standard deviation of 16.65 and a range of 16–59. As mentioned, CHIN-100 was designed for complete beginners, but one student (labeled S-100F in Table 1) came with the previous Chinese learning experience. After that student's score was removed, the mean became 18.60 with a standard deviation of only 2.70.

For CHIN-201, the mean score increased from 33.75 at the beginning of the semester to 42 at the end of the semester. The standard deviation was 16.50 in the pre-test with a slight increase of 2.94 in the post-test.

For CHIN-202, the mean score increased from 41.33 in the pre-test to 53.50 in the post-test. Note that while the proficiency gain in CHIN-202 was higher than that in CHIN-201, three out of the six students in CHIN-202 were also taking another Chinese course taught by another instructor during the same semester; therefore, the learning hours from another course must be factored into the results. The standard deviation was 15.62 in the pre-test, with a slight decrease of 1.71 in the post-test, mainly because of the increase of the minimum score from 18 to 31.

It is important to note that for the three students who continued Chinese study from CHIN-201 to CHIN-202, the CHIN-201 post-test scores were also treated as the CHIN-202 pre-test. Thus, the student who scored the lowest in the CHIN-201 post-test was the same student who scored the lowest in the CHIN-202 pre-test (labeled S-201/201A in Table 1); and the student who scored the highest in the CHIN-201 post-test was also the same student

who scored the highest in the CHIN-202 pre-test (labeled S-201/202C in Table 1). These two students remained the lowest- and the highest-scoring, respectively, in the CHIN-202 post-test. Note that the score gap between these two students increased from 39 points to 46 points after a semester of proficiency-oriented teaching in CHIN-201, but the gap decreased from 46 points to 37 points after the second semester of proficiency-oriented teaching, with the lowest score jumping from 18 to 31.

Overall, the mean scores in the higher-level courses (the CHIN-202 post-test vs. the CHIN-201 post-test; the CHIN-201 pre-test vs. the CHIN-100 post-test with true beginners only) are higher than those in the lower levels. Moreover, the SD and score range in CHIN-100, without including the student with prior language study, is relatively small compared to those in the second-year language classes. The dispersion measures (i.e., SD and range) suggest that there was a broader range of scores on the EIT in the two intermediate courses of CHIN-201 and CHIN-202 than the beginning course of CHIN-100.

Another interesting finding came from the student S-201/202C, who consistently scored the highest in the CHIN-201 and CHIN-202 courses. While various factors contribute to proficiency development, this student distinguished himself from other students in his advanced-level proficiency (self-reported attainment of just one level below the highest level in the Japanese-Language Proficiency Test) in a language closely related to Chinese, namely Japanese. When asked in the interview if Japanese proficiency helped Chinese language learning, the student said,

I think so. For me, it helps enormously. For me, the best thing is hanzi [Chinese characters]. Also, the pronunciation is similar enough that I can transfer it. For example, education in Japanese is [pronounced as] kyoiku, and in Chinese is [pronounced as] jiaoyu, and the writing is the same. Things like that. There are also a lot of vocabulary not learned in Chinese classes, but I already know, such as guojia zhuyi [meaning nationalism]. They sound very similar in Chinese and Japanese and are written in the same way. When studying Chinese, many words just pop up. This one I know; this one I know.

To this student, the knowledge of Japanese benefited him not only in learning spoken but also written aspects of Chinese. This student self-reported being able to read authentic Japanese novels. His rather advanced proficiency in reading Japanese also reflected in words such as “nationalism” and “education” that he provided as an example to demonstrate how his Japanese knowledge facilitated his study of Chinese. The other three CHIN-100 students, who were all in the same fourth-semester Japanese course, also identified the usefulness of having knowledge in Japanese language writing, specifically in what is known as Kanji, which uses many Chinese characters. However, the words they provided as examples were limited to only a few foundational high-frequency words, such as “girl,” “mountain,” “school,” “water,” and “learning.” Additionally, when asked explicitly if learning Japanese helped Chinese listening comprehension or speaking, these three students responded either “no” or “not really.” The interview data seem to reflect the argument by VanPatten (2017) that while cognates or knowledge of closely related language “may eventually be some help” (p. 67), the benefit of knowing a closely related language may not be evident for a learner who just starts learning a new language.

Discussion and Implications

Implications

As shown in the findings above, this Chinese EIT provided useful information about students’ development in oral proficiency. Several findings correspond to previous studies using the ACTFL OPI or OPIc to examine university students’ speaking proficiency. First,

that the range of scores overlaps across different levels of classes corresponds to Magnan's (1986) finding from the OPI test that "students at the same level of oral proficiency may be enrolled in different levels of study, and students of the same level of study may be at different levels of oral proficiency" (p. 430). Second, the mean score of a higher-level course is higher than the mean score of a lower-level course. This result is similar to the OPIc test finding from Chinese language learners in a university that students make progress on speaking as the levels in the curriculum move up (Polio, 2019). Third, the lowest scores in CHIN-201 pre- and post-test came from the only student in that class without any pre-university learning experience of Chinese. This case from the Chinese EIT corroborates the finding from the OPIc that pre-university language experience with the target language plays an important role in undergraduate proficiency outcomes (Strawbridge et al., 2019). Although these findings are probably not unexpected, of related importance to the present study is that they came from a reliable and valid instrument that can be practically administered by an instructor without substantial funding.

Additionally, the test scores from the EIT of L2 Chinese showed decent proficiency gains in both CHIN-201 and CHIN-202 after one semester of study in each course, meaning students' global oral proficiency kept progressing even when none of the practices and activities from the textbooks were used, nor any form-only activities, commonly called "drills," "mechanical practice," or "pattern practice" (Wong & VanPatten, 2003), were conducted. The proficiency gains were particularly noticeable for the same one student with the lowest score in CHIN-201 (from 13 to 18) and CHIN-202 (from 18 to 31) after two consecutive semesters of proficiency-based teaching. This student was also the only student in CHIN-201 without any pre-university learning experience of Chinese. However, even after taking 10-credit first-year university Chinese courses taught in a traditional textbook-centered manner, this student's CHIN-201 pre-test score was still lower than the lowest score in CHIN-100, and the student's CHIN-201 post-test score was still slightly lower than the average score of CHIN-100 ($M = 18.6$). Fall et al. (2007) stated that active involvement in scoring and analysing students' speech samples from proficiency assessment empowers teachers to be better equipped to make informed instructional decisions. Although this study did not conduct a qualitative analysis of speaking samples and the data were not sufficient to claim that proficiency-oriented language teaching works better than a traditional textbook-centered, drill-based approach in developing global oral proficiency, this test result from the EIT as analysed by me as an instructor did give me more confidence to continue moving toward proficiency-oriented teaching.

Reflecting on this process of assessing student oral proficiency outcomes from a standpoint as a university foreign language educator, the following discussion focuses on addressing the second research question: What is the potential practical value of the EIT for a language program with limited resources for standardized testing of proficiency? As noted earlier, the Chinese EIT has parallel tests in several other languages; thus, the practical value suggested here might be applicable to those parallel tests.

First, for programs where proficiency testing is available only at the exit point, as was the case in the present study, this research suggests that this EIT could be practically administered to periodically track students' oral proficiency development throughout the course of study, such as after the first, second, and third years of study as checkpoints, before taking the exit test. As Norris (2006) emphasized, the fundamental value of assessment does not end at getting the results but should function as a means for a washback effect to help programs do a better job. The empirical evidence from the EIT would enable more immediate action to better prepare students to achieve the expected level of oral proficiency before taking the exit test.

The second potential value of using this EIT is to provide supplementary information for programs without any proficiency assessment for placement, especially given that this EIT is a sensitive measure to distinguish speakers of higher and lower proficiency. Of the 13 total participants, five had school-based pre-university Chinese language learning experience. Three out of the five students had learned Chinese for four years in U.S. public high schools, but each was placed into a different course level: one (labeled S-201D in Table 1) starting with first-semester Chinese, one (labeled S-201/202B in Table 1) with second-semester Chinese, and one (labeled S-202D in Table 1) with fourth-semester Chinese. As mentioned before, the Chinese program in the present study had no formal placement test for students with previous Chinese language learning experience; the placement was based on an informal individual meeting with the student and the availability of courses at the time. According to these three students' performances on the EIT, the placement seemed to be level-appropriate, because none of them scored the lowest or the highest in pre- or post-tests among their peers in class. However, the placement might also be not level-appropriate, as demonstrated in the EIT result of the student labeled S-100F in Table 1, because the accuracy of "teacher intuitions" regarding the proficiency levels of students varies depending on factors such as the training in proficiency rating and the teacher's instructional approach (Glisan & Foltz, 1998). This student took CHIN-100 for the purpose of fulfilling graduation credits and concurrently continued his Chinese study in the second-semester Chinese course when taking the Chinese EIT. A possible explanation for initially placing this student into the first-semester Chinese was because of his intermittent study of the target language in earlier years (before high school), which might make it challenging to make a more informed decision for appropriate placement. Given that initial placement into the university language program has a significant and lasting influence on students' proficiency development and continued enrolment in the program (Strawbridge et al., 2019), standardized testing for placement purposes is imperative, and this EIT may feasibly serve that purpose.

Third, the EIT might be valuable for instructors who need or want a reliable, easy-to-use, and not too time-consuming proficiency assessment to show stakeholders evidence or for themselves to reflect on how effective their teaching practices or curriculum are with regard to developing students' global oral proficiency. For example, the large score ranges and high standard deviations in the second-year courses kept me more cautious about whether my high-percentage use of the target language was comprehensible enough for less proficient students. Additionally, the test scores from the EIT allowed me to examine to what extent my newly designed course, as was the case with the CHIN-100 course, successfully achieved the goal of developing student's oral proficiency. Thus, teachers could use first-hand local knowledge of teaching and learning in their own classes along with EIT results to make informed decisions about or stronger arguments for their curriculum designs and instructional practices.

Limitations and conclusion

In the article developing and validating the EIT used in the present study, Wu and Ortega (2013) called for future research to investigate the usefulness of this proficiency assessment in Chinese programs. This study responded to that call and demonstrated how valuable it could be in enabling an instructor to measure students' oral proficiency in a valid, reliable, and feasible manner beyond classroom achievement tests. The hope is that this research would facilitate a dialogue between second language assessment researchers and classroom teachers to address foreign language educators' practical needs in proficiency testing. Given that this research was conducted by an instructor, it is also hoped that this research sets an example to encourage language educators to start, if they are not already doing so, to use empirical evidence from a valid and reliable proficiency measurement tool to

reflect on, improve, and guide their instructional practices. Although this study was conducted at the university level, it is very likely that this EIT could be administered to Chinese language learners across a wide range of ages, because this EIT only requires the test taker to repeat 30 grammatical sentences in the target language and can be conveniently completed in 10 minutes. Moreover, as mentioned before, this EIT has the ability to distinguish between learners spanning the oral proficiency spectrum and between heritage learners and foreign language classroom learners. Therefore, it would be worthwhile for future research to explore the extent to which this EIT is useful for Chinese language programs in different contexts, such as high school, immersion school, or even heritage language school.

Lastly, it is important to point out the limitations of this research. As Wu and Ortega (2013) cautioned, this EIT does not tap language use in real-life, authentic contexts, and therefore it alone might not be satisfying for university Chinese language programs that prioritize that aspect of oral proficiency and want to know specifically where students' speaking levels are on the ACTFL proficiency scale. Future research can investigate the relationship between the score from this EIT and the rating determined by the ACTFL OPI/OPIc. Additionally, this measure alone is not sufficient for programs where literacy in the target language is also a consideration when placing students into courses. Moreover, there could be a concern regarding practice effect, because some students in this study took the same EIT as many as three times across two semesters. While most of the participants who took this test more than once reported not remembering the content of test items, a small number of them did recall a few phrases after hearing or repeating them. Comparable forms of the EIT in the same target language would greatly alleviate this issue; it would require the expertise of language assessment researchers. Last, this study is limited to a small number of participants; additional research from classes with a larger number of students could expand the potential value of using this EIT.

Reviewing the studies of L2 Chinese that investigated the acquisition and assessment of L2 Chinese speaking abilities, Liao (2018) pointed out that research on this topic is scarce. As communicative language teaching and proficiency-oriented instruction have become more prevalent, oral proficiency development in L2 Chinese deserves more attention not only from researchers but also from classroom teachers. This study demonstrated how valuable it is for a Chinese language teacher to have reliable information from a valid, efficient, and cost-saving oral proficiency test. Implementing the Chinese EIT developed by Wu and Ortega (2013) in different teaching contexts would greatly expand our understanding of how this proficiency measurement tool could be used for a variety of practical and pedagogical purposes.

References

- ACTFL. (2015). *ACTFL Performance Descriptors for Language Learners*. ACTFL
https://www.actfl.org/sites/default/files/publications/ACTFLPerformance_Descriptors.pdf.
- ACTFL. (2017). NCSSFL-ACTFL Can-Do Statements.
<https://www.actfl.org/resources/ncssfl-actfl-can-do-statements>.
- Bernhardt, E., Molitoris, J., Romeo K., Lin, N., & Valderrama, P. (2015). Designing and sustaining a foreign language writing proficiency assessment program at the postsecondary level. *Foreign Language Annals*, 48(3), 329-349.
<https://doi.org/10.1111/flan.12153>
- Burkhauser, S., Steele, J. L., Li, J., Slater, R. O., Bacon, M., & Miller, T. (2016). Partner-language learning trajectories in dual-language immersion: Evidence from an urban district. *Foreign Language Annals*, 49(3), 415-433. <https://doi.org/10.1111/flan.12218>
- Davin, K. J., Rempert, T. A. & Hammerand, A. A. (2014). Converting Data to Knowledge: One District's Experience Using Large-Scale Proficiency Assessment. *Foreign Language Annals*, 47(2), 241-260. <https://doi.org/10.1111/flan.12081>
- Dziedzic, J. (2012). A comparison of TPRS and traditional instruction, both with SSR. *The International Journal of Foreign Language Teaching*, 7(2), 4-6.
<http://www.ijflt.com/images/ijflt/articles-spring-2012/IJFLT.Dziedzic.3.12.pdf>
- Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 65-93). Multilingual Matters. <https://doi.org/10.21832/9781847691767-005>
- Fall, T., Adair-Hauck, B., & Glisan, E. (2007). Assessing students' oral proficiency: A case for online testing. *Foreign Language Annals*, 40(3), 377-406.
<https://doi.org/10.1111/j.1944-9720.2007.tb02865.x>
- Gaillard, S., & Tremblay, A. (2016). Linguistic proficiency assessment in second language acquisition research: The elicited imitation task. *Language Learning*, 66(2), 419-447.
<https://doi.org/10.1111/lang.12157>
- Glisan, E. W., & Foltz, D. A. (1998). Assessing students' oral proficiency in an outcome-based curriculum: Student performance and teacher intuitions. *Modern Language Journal*, 82(1), 1-18. <https://doi.org/10.1111/j.1540-4781.1998.tb02587.x>
- Goertler, S., Kraemer, A., & Schenker, T. (2016). Setting evidence-based language goals. *Foreign Language Annals*, 49(3), 434-454. <https://doi.org/10.1111/flan.12214>
- Hubert, M. D. (2013). The development of speaking and writing proficiencies in the Spanish language classroom: A case study. *Foreign Language Annals*, 46(1), 88-95.
<https://doi.org/10.1111/flan.12010>
- Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8(3): 229-249. <https://doi.org/10.1080/15434303.2011.565844>
- Kim, J. E., & Nam, H. (2017). Measures of implicit knowledge revisited. *Studies in Second Language Acquisition*, 39(3), 431-457. <https://doi.org/10.1017/S0272263115000510>
- Kim, Y., Tracy-Ventura, N., & Jung Y. (2016). A measure of proficiency or short-term memory? Validation of an elicited imitation test for SLA research. *Modern Language Journal*, 100(3): 655-673. <https://doi.org/10.1111/modl.12346>
- Liao, J. (2018). Acquisition and assessment of L2 Chinese speaking. In C. Ke (Ed.), *The Routledge Handbook of Chinese Second Language Acquisition* (pp. 234-260). Routledge.

- Lichtman, K. (2014). Teaching language through storytelling. *The Language Educator*, 9(5), 46-47.
- Magnan, S. S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals*, 19(5), 429-438. <https://doi.org/10.1111/j.1944-9720.1986.tb01031.x>
- Malone, M., & Tschirner, E. (2012, June). *The impact of the ACTFL proficiency guidelines on assessment and assessing foreign language skills in the U.S.* ACTFL-CEFR Symposium 2012: Making Frameworks Useful. Graz, Austria. https://www.researchgate.net/publication/303018227_The_impact_of_the_ACTFL_Proficiency_Guidelines_on_assessment_and_assessing_foreign_language_skills_in_the_US
- Martin, C. L. (2014). Introduction: Past context, present focus, future directions: Shifting focus from intermediate skills in classroom training to advanced/superior and beyond. In T. Brown & J. Bown (Eds.), *To Advanced Proficiency and Beyond: Theory and methods for developing superior second language ability* (pp. xiii-xxiv). Georgetown University.
- Moeller, A. J., & Theiler, J. (2014). Spoken Spanish language development at the high school level: A mixed-methods study. *Foreign Language Annals*, 47(2), 210-240. <https://doi.org/10.1111/flan.12085>
- Norris, J. M. (2006). The why (and how) of assessing student learning outcomes in college foreign language programs. *Modern Language Journal*, 90(4), 576-583. https://doi.org/10.1111/j.1540-4781.2006.00466_2.x
- Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*. Toronto, Canada: 3-6 October 2002.
- Polio, C. (2019). Linking proficiency test scores to classroom instruction. In P. Winke & S. M. Gass (Eds.), *Foreign language proficiency in higher education* (pp. 287-308). Springer International.
- Schmitt, E. (2016). Seat time versus proficiency: assessment of language development in undergraduate students." In N. Mills, & J. Norris (Eds.), *AAUSC 2014 volume-issues in language program direction: Innovation and accountability in language program evaluation* (pp. 110-130). Cengage.
- Soneson, D., & Tarone, E. E. (2019). Picking up the PACE: Proficiency assessment for curricular enhancement." In P. Winke, & S. Gass (Eds.), *Foreign language proficiency in higher education* (pp. 45-70). Springer International.
- Strawbridge, T., Soneson, D., & Griffith, C. (2019). Lasting effects of pre-university language exposure on undergraduate proficiency. *Foreign Language Annals*, 52(4), 776-797. <https://doi.org/10.1111/flan.12430>
- Tracy-Ventura, N., McManus, K., Norris, J. M., & Ortega, L. (2014). "Repeat as much as you can": Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, H. Hilton, & A. Edmonds (Eds.), *Measuring L2 Proficiency: Perspectives from SLA* (pp. 143-166). Multilingual Matters.
- Tschirner, E. (2016). Listening and reading proficiency levels of college students. *Foreign Language Annals*, 49(2), 201-223. <https://doi.org/10.1111/flan.12198>
- Vyn, R., Wesely, P. M., & Neubauer, D. (2019). Exploring the effects of foreign language instructional practices on student proficiency development. *Foreign Language Annals*, 52(1), 45-65. <https://doi.org/10.1111/flan.12382>
- Watzinger-Tharp, J., Rubio, F., & Tharp, D. S. (2018). Linguistic performance of dual language immersion students. *Foreign Language Annals*, 51(3), 575-595. <https://doi.org/10.1111/flan.12354>

- Winke, P., & Gass, S. M. (2018). Individual differences in advanced proficiency. In P. A. Malovrh & A. G. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition* (pp. 157-178). Wiley Blackwell.
- Winke, P., & Gass, S. M. (Eds.). (2019). *Foreign language proficiency in higher education*. Springer International. <https://doi.org/10.1007/978-3-030-01006-5>
- Wong, W., & VanPatten, B. (2003). The evidence is IN: Drills are OUT. *Foreign Language Annals*, 36(3), 403-423. <https://doi.org/10.1111/j.1944-9720.2003.tb02123.x>
- Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46(4), 680-704. <https://doi.org/10.1111/flan.12063>
- Xu, X., Padilla, A. M., & Silva, D. M. (2015). Learner performance in Mandarin immersion and high school world language programs: A comparison. *Foreign Language Annals*, 48(1), 26-38. <https://doi.org/10.1111/flan.12123>
- Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, 33(4), 497-528. <https://doi.org/10.1177/0265532215594643>
- Yuan, F., & Li, S. (Eds.). (2019). *Classroom research on Chinese as a second language*. Routledge.

Table 1

Summary of each student's language learning profile and initial placement into the Chinese language program

	Language Background
S-100A	Three years of Spanish in high school; taking fourth-semester Japanese concurrently
S-100B	Three years of Spanish in high school; taking fourth-semester Japanese concurrently
S-100C	Four years of German in high school; taking German and French concurrently
S-100D	Three years of Japanese in high school; taking fourth-semester French and fourth-semester Japanese concurrently
S-100E	Six years of Spanish throughout middle and high schools; not taking other language classes
S-100F	Native speaker of Thai; learned some Chinese in both elementary and middle school but not in high school; came to the U.S. for college; initially placed into first-semester Chinese; taking second-semester Chinese concurrently
S-201/202A	Six years of French throughout middle and high schools
S-201/202B	Four years of high school study of Chinese; initially placed into second-semester Chinese in the program
S-201/202C	Native speaker of Thai; came to U.S. for high school; one semester of learning Chinese in middle school; started learning Japanese from middle school; started with first-semester Chinese in the program; taking a 300-level Chinese course while taking CHIN-202
S-201D	Four years of high school study of Chinese; initially placed into first-semester Chinese in the program
S-202D	Four years of high school study of Chinese with an exit attainment of 3 in AP; initially placed into fourth-semester Chinese in the program
S-202E	Four years of German in high school; sixth semester learning Chinese and also taking another Chinese course; started with first-semester Chinese in the program
S-202F	Six years of studying French since high school; self-learned Thai, Korean, and Japanese, but not as proficient as Chinese; sixth semester learning Chinese and also taking another Chinese course; started with first-semester Chinese in the program

Abbreviations: Each student was labeled with the course enrolled, followed by a letter after the course; CHIN201/202 means the student was enrolled in both courses consecutively.

Table 2

Descriptive statistics for the Chinese EIT in each class

	Minimum	Maximum	Mean	SD
CHIN-100 (N = 6)	17	59	26.17	16.65
CHIN-100 (N = 5, true beginners only)*	17	23	18.60	2.70
CHIN-201 Pre-test (N = 4)	13	52	33.75	16.50
CHIN-201 Post-test (N = 4)	18	64	42	19.44
CHIN-202 Pre-test (N = 6)	18	64	41.33	15.62
CHIN-202 Post-test (N = 6)	31	68	53.50	13.91

* This group was created after the score from the not true beginner was removed.

ⁱ The video is available at <https://www.youtube.com/watch?v=Gdt3I0X1bE8>.