Understanding the Complexities of Chinese Word Acquisition within an Online Learning Platform

Xiwen Lu[1][a], Korinn S. Ostrow[2][b] and Neil T. Heffernan[2][c]
[1]German, Russian, and Asian Languages and Literature, Brandeis University, 415 South Street, Waltham, MA 02453, U.S.A.
[2]Computer Science Department, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, U.S.A.

Keywords: Chinese (Mandarin), Foreign/Second Language Learning/Acquisition, Sound, Meaning, Character, Computer-Assisted Language Learning (CALL), Randomized Controlled Trial, ASSISTments Testbed.

Abstract: Because Chinese reading and writing systems are not phonetic, Mandarin Chinese learners must construct six-way mental connections in order to learn new words, linking characters, meanings, and sounds. Little research has focused on the difficulties inherent to each specific component involved in this process, especially within digital learning environments. The present work examines Chinese word acquisition within ASSISTments, an online learning platform traditionally known for mathematics education. Students were randomly assigned to one of three conditions in which researchers manipulated a learning assignment to exclude one of three bi-directional connections thought to be required for Chinese language acquisition (i.e., sound-meaning and meaning-sound). Researchers then examined whether students' performance differed significantly when the learning assignment lacked sound-character, character-meaning, or meaning-sound connection pairs, and whether certain problem types were more difficult for students than others. Assessment of problems by component type (i.e., characters, meanings, and sounds) revealed support for the relative ease of problems that provided sounds, with students exhibiting higher accuracy with fewer attempts and less need for system feedback when sounds were included. However, analysis revealed no significant differences in word acquisition by condition, as evidenced by next-day post-test scores or pre- to post-test gain scores. Implications and suggestions for future work are discussed.

1 INTRODUCTION

Mandarin Chinese is one of the most difficult languages for a native English speaker to learn. In 1982, the Foreign Service Institute (FSI) of the U.S. Department of State published a ranking that compared the approximate amount of time required for native English-speaking students to achieve "General Professional Proficiency in Speaking" and "General Professional Proficiency in Reading" in various foreign languages (Liskin-Gasparro, 1982). The report listed Chinese as one of the five most difficult languages to learn, requiring 2,200 class hours to achieve speaking and reading proficiency; by comparison, French and Spanish were both listed as requiring less than 600 hours (Liskin-Gasparro, 1982; Wolff, 1989). Chinese takes substantially longer to master than any of the European languages traditionally taught in American public schools (e.g., French, Spanish, German, etc.) due to its lack of common vocabulary roots, its novel tonal and writing systems, and its distinctly different syntactic structure.

De Francis (1984) summarized this issue by stating that "the most difficult and time-consuming aspects of learning Chinese are character recognition and handwriting." Because Chinese reading and writing systems are not phonetic, learners are required to construct a six-way mental connection in order to learn each new word. When learning a new

[a] https://orcid.org/0000-0001-9243-9069
[b] https://orcid.org/0000-0001-7149-1802
[c] https://orcid.org/0000-0002-3280-288X

Lu, X., Ostrow, K. and Heffernan, N.
Understanding the Complexities of Chinese Word Acquisition within an Online Learning Platform.
DOI: 10.5220/0007585903210329
In *Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019)*, pages 321-329
ISBN: 978-989-758-367-4
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

word, students must learn the word's specific character(s), associate the character(s) with its meaning, and recall the word's sound(s) in order to communicate with the word. Connections between character, meaning, and sound must be bi-directional for successful use of the word in writing, reading, and conversation. In contrast, learners of phonetic European languages usually only need to construct two-way mental connections to be able to write, read, and converse. Native speakers of traditionally phonetic languages struggle when learning Mandarin because it is not possible to "spell" Chinese characters and, often, there is no obvious association between a character and its sound (i.e., it is not possible to "sound it out"). Even native Chinese speakers may blank when called upon to write the character for a relatively common word due to this lack of intuitive connections.

Warschauer and Healey (1998) argued that the development of information technology provided foreign language instructors and learners with new possibilities for practicing language acquisition. For instance, the popular language-learning app Duolingo (2017) offers gamified, self-paced courses for native English speakers to learn 27 languages (NPR/TED Staff, 2014). As users progress through each lesson in their course, the app uses their responses to develop and verify translations of websites and articles on the Internet. However, Duolingo did not offer a course in Mandarin until 2017 (Hagiwara, 2017), likely due to the complexities involved in teaching and learning Chinese as a second language. The app ChineseSkill follows a gamified format similar to that of Duolingo, but focuses strictly on Mandarin (ChineseSkill Co., Ltd, 2017). Applications like these broaden the reach of the Chinese language to learners who may have otherwise been intimidated by the time and commitment it requires to gain fluency.

Research has also shown that reading and writing Chinese characters are two separate information acquisition processes with different influencing factors (Jiang, 2007). The use of Pinyin, a Romanization system for Mandarin, helps to link these processes by transforming characters into phonetic words. Through Pinyin, applications like Duolingo and ChineseSkill, as well as other digital resources, can allow learners to read or submit the phonetic versions of Chinese characters. Zhu et al., (2009) indicated that as a digital input method, Pinyin can strengthen character recognition through the consolidation of pronunciation capabilities. While tablets and other touch devices may allow learners to draw characters, Pinyin bridges the availability of Chinese learning acquisition to broader digital environments. Learners must still memorize characters for the sake of recognition in reading, and in order to connect the character and its Pinyin equivalent, but producing characters and applying the proper stroke order is no longer a necessity.

Despite an understanding that Chinese language acquisition requires characters, meanings, sounds, and often Pinyin, little research has been done on the difficulties inherent to each specific component of the process. Even less work has focused on how Chinese language acquisition has adapted to the digital world. As a Chinese language instructor at a major institution in New England, the first author observed that students typically begin word memorization by practicing connections between sound and meaning, considering the character/meaning connection as a secondary task. Her observation was supported by literature on Chinese language acquisition. Tan and Perfetti (1999) suggested that phonology is an obligatory component of word identification in Chinese reading. Perfetti and Liu (2006) then supported this idea, stating that phonology is automatically activated in reading words, regardless of whether activation occurs before or after the moment of lexical access and regardless of whether it is instrumental in retrieving the word's meaning. Essentially, Chinese characters activate pronunciation, even when the reader's goal is to determine the character's meaning.

Based on past research and considering the six connections required for Chinese word acquisition, it is possible to speculate that connections between meaning and character are more difficult because sound must also be accessed, even if unintentionally. It becomes difficult to discern if and how these components of word acquisition can be teased apart, and whether providing particular types of connections more frequently than others has the potential to produce more robust learning.

As such, we conducted the present study as a randomized controlled trial with three conditions to compare the consequences of removing each bi-directional connection pair involved in Chinese word acquisition: sound focused practice, meaning focused practice, and character focused practice. The present study sought to answer the following research questions:

1. Does student performance, as measured at post-test, differ significantly when a learning assignment for novel words lacks bi-directional sound-meaning, meaning-character, or character-sound connection pairs?

2. Which problem types pose the most difficulty for word acquisition? Are problems that lack sound-based connections more difficult? Do students make significantly more attempts or use hint feedback significantly more when completing these types of problems?

We hypothesized that evidence of student's word acquisition would differ significantly when bi-directional acquisition connections were reduced or removed from the learning experience (i.e., when their learning assignments lacked sound-meaning, meaning-character, or character-sound connection pairs). We also hypothesized that problems providing sound (e.g., sound-meaning, meaning-sound, sound-character, character-sound) would be easier for students and would lead to greater evidence of word acquisition.

2 METHOD

2.1 Participants

Participants included 60 students enrolled in an Intermediate Level Chinese class at a major university in New England during the Fall 2016 semester. Students participated in the course for credit and had been enrolled at the intermediate level following their experience in a preliminary Chinese course or based on their performance on a placement exam. All 60 students were assigned a pre-test and learning assignment on Day 1 of the study. Of these students, two could not access the study's internet-based video content and were removed from the analytic sample. In addition, as random assignment was conducted virtually at the start of the learning assignment, three students were not assigned to condition because they did not participate in the learning assignment and were therefore excluded from the analytic sample. The remaining 55 students were randomly assigned to one of three experimental conditions: 1) a sound focused practice, 2) a meaning focused practice, or 3) a character focused practice. Student-level randomization was conducted by the learning platform used to deliver study materials. Randomization did not result in a particularly normal distribution across conditions because group sizes were somewhat small. However, attrition rates did not differ significantly between conditions allowing the authors to proceeded with the analyses discussed herein. Three students failed to complete the Day 1 learning assignment, and 51 students were assigned to the post-test on Day 2. Of them, 50 students completed the post-test, with 46 having first completed the Day 1 pre-test and learning assignment. Thus, the present work uses an analytic sample of n = 46.

2.2 Setting

All study materials were delivered as part of a graded classwork assignment, and subjects participated during their regular class period. The present work was the first of its kind to be conducted using the ASSISTments Testbed, infrastructure that leverages ASSISTments, an online learning platform typically used for middle school mathematics. The platform is offered as a free service of Worcester Polytechnic Institute (WPI), with the goal of providing students with instructional *assist*ance while offering teachers reports for formative assess*ment*, thereby establishing its moniker (Heffernan and Heffernan, 2014). The Testbed is unique in that it allows researchers to use ASSISTments to conduct randomized controlled trials by manipulating premade content that is then made available to a subject pool of more than 50,000 student users. The platform can also be used to develop materials and conduct research in other domains, and its collections of certified materials in Physics, Chemistry, and Language are growing.

Studies conducted in the ASSISTments Testbed are covered primarily by WPI's IRB and researchers from other institutions can seek exemptions from their own IRB to work with the de-identified student data provided by the system. Under the Testbed's IRB regulations, as long as experimental conditions fall within the boundaries of "normal instructional practice," participants do not need to be informed of or provide consent for their participation in the experiment, but may be debriefed after the fact.

2.3 Materials

The first author, a lecturer in Intermediate Level Chinese but not the active teacher of the participants in this study, worked collaboratively with the sitting lecturer to select ten novel words from the course textbook to assign as learning targets. A pre-test was used to assess participants' knowledge of these words prior to beginning the learning assignment. Students that knew any or all words were still required to complete the learning assignment in class, but they were not included in the analytic sample as our focus was on *novel* word acquisition. Further description of the ten words is available in our supplementary materials (Lu, 2017).

The learning assignment consisted of sixty possible problems. Six problems were developed for each of the ten novel words, corresponding to the six (bi-directional) connections that must be constructed between components of sound, meaning, and character. For each word, problems prompted students to provide solutions associating 1) sound to meaning, 2) meaning to sound, 3) sound to character, 4) character to sound, 5) meaning to character, and 6) character to meaning. Problems featuring sound components utilized brief YouTube videos to deliver audio. All problems can be referenced in our supplementary materials (Lu, 2017).

2.4 Procedures

The experimental design spanned two consecutive class meetings and students were allowed to work at their own pace. On the first day, students created ASSISTments accounts and were provided an explanation of how to proceed with their pre-test and learning assignment. Before starting the study, students first completed a data collection problem assessing their ability to access YouTube videos. Responses were used to verify that students would be able to receive the sound component of problems or feedback. If students could not access video content they were routed into an alternative (but similar) assignment and were excluded from the analytic sample. Students with access to video were randomly assigned to one of three conditions that examined the removal of each bi-directional pair of connections involved in Chinese language acquisition.

Regardless of condition, the learning assignment began with a ten question pre-test to assess knowledge of each novel word. Each pre-test problem followed the format: "Please write down the English meaning of the word '孩子, hai2zi0'. If you don't know this word, please enter the word 'no'." Following the pre-test all participants began a learning assignment with 40 problems. Participants in Condition 1 received four types of problems featuring sound-based word acquisition connections for each of the 10 target words (e.g., sound to meaning, meaning to sound, sound to character, and character to sound). Similarly, participants in Condition 2 received four types of problems featuring meaning-based connections for each of the 10 target words, and participants in Condition 3 received four types of problems featuring character-based connections for each of the 10 target words.

For each problem, participants could ask the system for a single hint if they were unable to provide the correct answer. Each hint was developed to provide the problem's missing language component. For instance, if the problem asked the student to convert a *character* to its *meaning*, the hint would provide the word's *sound* using a YouTube video. If a student was still unable to reach the solution after acknowledging all three components of the word, they were able to request the correct answer from the system in order to move on to the next problem.

On Day 2, students began class by logging into ASSISTments and taking a post-test with problems that mirrored those on the pre-test. As students were not aware that their Day 1 learning assignment was part of an experiment, problems for the ten novel words were interleaved with problems featuring other words from a recent lecture to make the post-test more comprehensive and to promote the need for word recall beyond rote memory. Students were told that their scores on Day 2 material would count as a daily quiz grade.

2.5 Analyses

Data collection included student's pre-test and post-test responses, as well as their performance on the Day 1 learning assignment. All measures were logged by ASSISTments, including students' problem-level accuracy, response time, attempt count, hint usage, and answer requests. All data was generated by the reporting infrastructure of the ASSISTments Testbed. Raw csv files were cleaned and entered into IBM's SPSS for analysis.

Students' accuracy on pre-test, learning assignment, and post-test problems were used to compare learning gains by condition using an Analysis of Variance (ANOVA). Descriptive statistics of students' performance at each of these three time points are presented in Table 1 by condition, with pre- to post-test gain scores calculated for convenience.

Analysis of problem type difficulty focused on detailed measures of students' performance on the Day 1 learning assignment including accuracy, attempt count, hint usage, and answer requests. Student-level averages for these measures were calculated to control for the number of problems students experienced during the learning assignment. Multiple ANOVAs were then conducted using each of these four average measures as dependent variables to examine how problem types differed by problem type and condition. Descriptive statistics and ANOVA results for the four dependent variables are presented in Table 2 by problem type and condition.

Table 1: Descriptive statistics of performance exhibited across conditions.

	n	Pre-test	Assignment	Post-test	Gain (Pre-Post)
C1 – Sound Focused Practice	10	0.32 (0.15)	0.69 (0.11)	0.78 (0.17)	0.43 (0.25)
C2 – Meaning Focused Practice	22	0.41 (0.20)	0.80 (0.08)	0.81 (0.13)	0.40 (0.27)
C3 – Character Focused Practice	14	0.42 (0.26)	0.76 (0.16)	0.83 (0.19)	0.47 (0.25)

Note. C = Condition. Descriptive statistics presented as: Mean (*SD*).

Table 2: Descriptive statistics and ANOVA results for DVs by Problem Type and Condition.

Problem Type/Condition	**Accuracy**	**Hint Count**	**Attempt Count**	**Answer Requests**
Providing Sound	$F(2, 52) = 8.29$**	$F(2, 52) = 20.91$***	$F(2, 52) = 6.32$**	--
C1 – Sound Focused Practice	0.76 (0.14)	0.08 (0.05)	1.38 (0.28)	0.00 (0.00)
C2 – Meaning Focused Practice	0.90 (0.10)	0.00 (0.00)	1.14 (0.16)	0.00 (0.00)
C3 – Character Focused Practice	0.90 (0.11)	0.01 (0.05)	1.16 (0.20)	0.00 (0.00)
Total	0.87 (0.12)	0.02 (0.05)	1.20 (0.23)	0.00 (0.00)
Providing Meaning	$F(2, 52) = 57.80$***	$F(2, 52) = 31.60$***	$F(2, 52) = 24.56$***	$F(2, 52) = 17.16$***
C1 – Sound Focused Practice	0.32 (0.25)	1.02 (0.58)	3.53 (1.62)	0.39 (0.31)
C2 – Meaning Focused Practice	0.65 (0.14)	0.52 (0.33)	2.18 (0.85)	0.21 (0.17)
C3 – Character Focused Practice	0.95 (0.08)	0.00 (0.00)	1.08 (0.13)	0.00 (0.00)
Total	0.67 (0.28)	0.47 (0.52)	2.14 (1.32)	0.19 (0.24)
Providing Character	$F(2, 53) = 48.98$***	$F(2, 53) = 20.33$***	$F(2, 53) = 18.03$***	$F(2, 53) = 14.75$***
C1 – Sound Focused Practice	0.93 (0.08)	0.08 (0.13)	1.19 (0.27)	0.02 (0.04)
C2 – Meaning Focused Practice	0.98 (0.04)	0.00 (0.00)	1.02 (0.04)	0.00 (0.00)
C3 – Character Focused Practice	0.63 (0.19)	0.44 (0.38)	2.43 (1.36)	0.17 (0.18)
Total	0.85 (0.20)	0.17 (0.30)	1.54 (1.02)	0.06 (0.13)

Note. C = Condition. Descriptive statistics presented as: Mean (SD).
*** $p < .001$, ** $p < .01$

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

3 RESULTS

We hypothesized that evidence of student's word acquisition would differ significantly when bi-directional acquisition connections were removed or reduced from the learning experience (i.e., when their learning assignments lacked sound-meaning, meaning-character, or character-sound connection pairs). We also hypothesized that problems providing sound (e.g., sound-meaning, meaning-sound, sound-character, character-sound) would be easier for students and would lead to greater evidence of word acquisition.

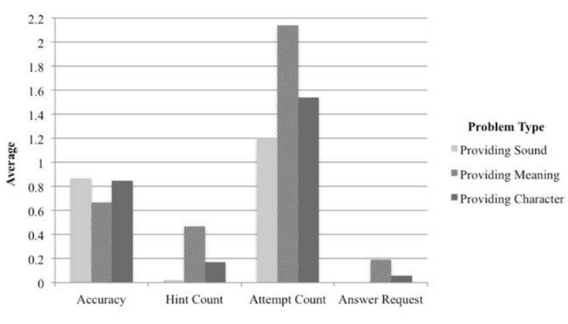

Figure 1: Average accuracy, hint count, attempt count, and amount of answer requests across problem types.

3.1 Learning Gains

In order to assess whether student performance, as measured at post-test, differed significantly when word assignments lacked sound-meaning, meaning-character, or character-sound connection pairs, we analysed the performance of 46 students that completed all Day 1 and Day 2 study materials. ANOVA results revealed that conditions were not significantly different at pre-test, $F(2, 53) = 0.87$, $p > .05$, $\eta^2 = .03$, despite a lower average for those assigned to receive sound focused problems. Average scores on the 40 problem learning assignment were significantly different across conditions, $F(2, 53) = 3.38$, $p < .05$, $\eta^2 = .11$, driven by a significant difference between sound focused practice and meaning focused practice. Post hoc tests revealed that students were significantly more accurate on meaning focused practice problems, $p = .04$, 95% CI [-.21, -.01], Cohen's $d = -.26$. Despite these differences between assignment scores, no significant differences were observed at post-test, $F(2, 43) = 0.29$, $p > .05$, $\eta^2 = .01$. Additionally, pre- to post-test gains were not significantly different between conditions, $F(2, 42) = 0.31$, $p > .05$, partial $\eta^2 = .01$.

3.2 Learning Difficulty

In order to assess whether word assignments that lacked sound-based connections were more difficult, we further analysed the performance of the 56 students who completed the Day 1 learning assignment. Data was sorted at the problem-level to examine the effect of problem types (i.e., problems providing sounds, problems providing meanings, and problems providing characters). Using this organization structure, students had experienced either 10 or 20 problems of each type, depending on their assigned condition. Four dependent variables were considered, including student's average accuracy, attempt count, hint usage, and answer requests. Descriptive statistics and ANOVA results for the four dependent variables are presented in Table 2 by problem type and condition. Descriptive statistics are also compared visually in Figure 1.

3.2.1 Accuracy

Significant differences were observed between students' average accuracy on differing problem types, $F(2,163) = 14.49$, $p < 0.001$, $\eta^2 = .15$. Specifically, significant differences were observed between problems providing sounds and those providing meanings, $p < .001$, Cohen's $d = .93$, as well as between problems providing meanings and those providing characters, $p < .001$, Cohen's $d = .74$. Problems that provided meanings resulted in the lowest accuracy on average ($M = 0.67$, $SD = 0.28$). In contrast, problems that provided sounds or characters resulted in higher accuracy on average, ($M = 0.67$, $SD = 0.28$ and $M = 0.67$, $SD = 0.28$, respectively). Significant differences were also observed between experimental conditions with regard to problems providing sounds, meanings, and characters (all $p < .01$), as shown in Table 2.

3.2.2 Hint Count

Next, significant differences were observed between the average number of hints requested by students on differing problem types, $F(2,163) = 23.58$, $p < 0.001$, $\eta^2 = .22$. Significant differences were again observed between problems providing sounds and those providing meanings, $p < .001$, Cohen's $d = 1.21$, as well as between problems providing meanings and those providing characters, $p < .001$, Cohen's $d = 0.71$. Students required the most hints on average on problems that provided meanings ($M = 0.47$, $SD = 0.52$), and the fewest hints on average on problems that provided sounds ($M = 0.02$, $SD = 0.05$). Students required a moderate amount of hints on average on problems that provided characters ($M = 0.17$, $SD = 0.30$). Significant differences were also observed between experimental conditions with regard to problems providing sounds, meanings, and characters (all $p < .001$), as shown in Table 2.

3.2.3 Attempt Count

Significant differences were also observed between the average number of attempts made by students on differing problem types, $F(2,163) = 13.11$, $p < 0.001$, $\eta^2 = .14$. Significant differences were again observed between problems providing sounds and those providing meanings, $p < .001$, Cohen's $d = .99$, as well as between problems providing meanings and those providing characters, $p = .002$, Cohen's $d = 0.51$. Students made the most attempts on average on problems that provided meanings ($M = 2.14$, $SD = 1.32$), and the fewest attempts on average on problems that provided sounds ($M = 1.20$, $SD = 0.23$). Students made a moderate number of attempts on average on problems that provided characters ($M = 1.54$, $SD = 1.02$). Significant differences were also observed between experimental conditions with regard to problems providing sounds, meanings, and characters (all $p < .01$), as shown in Table 2.

3.2.4 Answer Requests

Finally, significant differences were observed between the average number of answers requested by students on differing problem types, $F(2,163) = 20.54$, $p < 0.001$, $\eta^2 = .20$. Interestingly, regardless of condition, students did not request answers at all when solving problems that provided sounds ($M = 0.00$, $SD = 0.00$, all conditions). Significant differences were observed between problems providing sounds and those providing meanings, $p < .001$, as well as between problems providing meanings and those providing characters, $p < .001$, Cohen's $d = .67$. Students requested answers most frequently on average when working on problems that provided meanings ($M = 0.19$, $SD = 0.24$), but less frequently on average when working on problems that provided characters ($M = 0.06$, $SD = 0.13$). Significant differences were also observed between experimental conditions with regard to problems providing meanings and characters (both $p < .001$), as shown in Table 2.

4 DISCUSSION

Research suggests that Mandarin Chinese is one of the most difficult languages for native English speakers to acquire, requiring almost four times as much class time to reach the same level of speaking and reading proficiency as French or Spanish (Liskin-Gasparro, 1982; Wolff, 1989). Because Chinese reading and writing systems are not phonetic, learners must construct six-way mental connections between sounds, meaning, and characters in order to learn new words. However, little work has focused on the relative importance of each type of connection. Online learning applications have allowed learners to broach language acquisition in new and unique ways, while collecting powerful data that researchers can use to investigate the complexities of word acquisition. The present work leveraged on online learning environment to examine how the components of Chinese word acquisition (sounds, meanings, and characters) influence students' learning outcomes.

The present work first sought to understand if removing a bi-directional connection (i.e., sound to meaning and meaning to sound) would significantly alter students' learning as measured by a delayed post-test. Three learning assignments were constructed to test this hypothesis, and students were randomly assigned to 1) sound focused practice, 2 meaning focused practice, or 3) character focused practice. Differences between conditions were not significant at post-test or when performance was viewed in terms of pre- to post-test gains. However, students' performance within their learning assignments did differ significantly by condition. Thus, despite the emphasis that past research has placed on sound components within Chinese word acquisition, the present work offered no evidence that the removal of sound disproportionately hindered word acquisition.

Based on past research (Perfetti and Liu, 2006; Perfetti et al., 1992; Tan and Perfetti, 1999) the present work also sought to examine whether problem types lacking sound connections would be more difficult to students, as measured by their average accuracy, attempt count, and need for system-provided hint or answer feedback. Restructuring the data by problem type, analysis of 56 students revealed significant differences in within-assignment difficulty as measured by students' average accuracy, attempt count, hint usage, and answer requests across problems within the learning assignment. For each dependent variable, significant differences were observed between problems providing sounds and those providing meanings, as well as between problems providing meanings and those providing characters.

Problems that provided a word's sound (sound to meaning, sound to character) were least difficult for students; they had higher average accuracy on these problems while requiring fewer attempts and asking for hints and/or answers less frequently on average. In contrast, problems that provided a word's meaning (meaning to sound, meaning to character) were most difficult for students; they had lower average accuracy on these problems while requiring more attempts and asking for hints and/or answer more frequently on average. While this finding did not directly replicate the Universal Phonological Principle described by Perfetti et al., (1992), it reflected the principle from an alternative perspective. Based on this finding, teachers and developers of digital language content should expect that practices providing a word's sound will be easier for students, while those providing a word's meaning will be more difficult. As such, it may be beneficial to start assignments and lessons with practices that provide sound-based connections, allocating time to the practice of meaning-based connections as secondary instruction.

It is important to note that the present work was not without limitation. As student-level random assignment was conducted by the ASSISTments platform, distribution across the three experimental conditions was not well balanced. As the unbalanced distribution was caused by chance, a larger sample size may have resolved this issue. A larger sample size may have also revealed a greater difference in learning gains between conditions, as group sizes less than n = 30 suggested reduced power. Non-parametric tests could alternatively be considered in future work with small sample sizes.

Given the observation that problems providing a word's meaning posed the greatest difficulty for students while problems providing a word's sound were met with the greatest ease, future work should consider how altering practice strictly by component (i.e., by removing problems that provide a word's sound and expect students to return a meaning or character) rather than by a bi-directional connection pair (i.e., removing sound to meaning and meaning to sound) impacts evidence of students' word acquisition. Although this shift would likely result in significantly different evidence of word acquisition at post-test, it was not employed in the present work to maintain "normal instructional practice" within an authentic classwork assignment.

Further, participants in the present study experienced each of the four problem types (by condition) only once per target word. Although this resulted in 40 problems spanning the ten target words, future work should examine how adding additional practice might ultimately enhance learning. Evidence of word acquisition and learning gains may be stronger with additional practice.

The present work may also have been limited by a ceiling effect, which future work could seek to confirm or refute. Future work should also consider long-term word retention, as evidence for more robust word acquisition may function differently than that observed over just two days. Future work should also further examine *why* problem types guided by acquisition component (sound, meaning, or character) pose different levels of difficulty to students.

We are left with many questions to be tackled in future work. What is it that makes a particular type of problem more or less difficult to answer? Do problems that provide a word's meaning strain learners to recall it's sound or character? Is the added difficulty inherently beneficial for later word retention? How can teachers and learning platforms better assist students learning Chinese as a second language with these types of problems?

5 CONTRIBUTION

Prior work has failed to focus on the relative importance of the connections between the sound, meaning, and character components required for successful Chinese language acquisition. The present work teased these components apart within the context of an online learning assignment, discovering that problem difficulty levels vary significantly by component type, but that the removal of particular bi-directional connections between components did not significantly impact novel word acquisition as measured at post-test. Results suggested that problems that provide word meaning and expect students to return a sound or character are most difficult, while problems that provide sounds and expect students to return a meaning or character are least difficult. This finding suggests that instructors of Chinese as a foreign language, and those building digital learning content for Mandarin, should begin practices with sound-based connections and spend extra time on meaning-based connections later in practice. This finding has the potential to enhance the way Chinese language is taught in foreign language classrooms and in digital learning environments, reducing difficulty for students and, perhaps, enhancing their motivation to continue pursuing the Chinese language.

ACKNOWLEDGMENTS

The authors acknowledge funding from multiple NSF grants (ACI-1440753, DRL-1252297, DRL-1109483, DRL-1316736 & DRL-1031398), the U.S. Department of Education (IES R305A120125 & R305C100024 and GAANN), the ONR, and the Gates Foundation.

REFERENCES

ChineseSkill Co., Ltd. 2017. *ChineseSkill - Learn Chinese/Mandarin Language for free.* Retrieved from: http://www.chinese-skill.com/cs.html.

De Francis, J. 1984. *The Chinese Language: Fact and Fantasy*. Honolulu, HI: University of Hawaii Press.

Duolingo. 2017. *Language Courses for English Speakers.* Retrieved from: https://www.duolingo.com/courses.

Hagiwara, M. 2017. *Duolingo now supports Chinese, but it probably won't help you become fluent*. Retrieved from:https://www.theverge.com/2017/11/16/16598626

Heffernan, N. T. & Heffernan, C. L. 2014. The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. *International Journal of Artificial Intelligence in Education. 24*(4): 470-497. doi:10.1007/s40593-014-0024-x.

Jiang, X. 2007. An experimental study on the effect of the method of "teaching the learner to recognize characters more than Writing". *Chinese Teaching in the World, 2007*(2): 91-97.

Liskin-Gasparro, J. 1982. *ETS oral proficiency test manual.* Princeton, NJ: Educational Testing Service.

Lu, X. 2017. Experiment data. Retrieved from: tiny.cc/LuOstrowHeffernanCSEDU19

NPR/TED Staff. 2014. Translating the Web with Millions: Luis Von Ahn Answers Your Questions. *TED Radio Hour*. Retrieved from: http://www.npr.org/2014/06/10/319071368.

Perfetti, C. A. & Liu. Y. 2006. Reading Chinese characters: Orthography, phonology, meaning, and the lexical constituency model. In P. Li, L. H., Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of East Asian Psycholinguistics*. 1(Chinese): 225-236. New York: Cambridge University Press.

Perfetti, C. A., Zhang, S., & Berent, I. 1992. Reading in English and Chinese: Evidence for a "universal" phonological principle. In R. Frost & L. Katz (Eds.), *Advances in Psychology*. 94 (Orthography, phonology,

morphology, and meaning): 227-248. Oxford, England: North-Holland. doi:10.1016/S0166-4115(08)62798-3.
Tan, L. & Perfetti, C. A. 1999. Phonological activation in visual identification of Chinese two-character words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 25(2): 382-393. doi:10.1037/0278-7393.25.2.382
Warschauer, M., & Healey, D. 1998. Computers and language learning: An overview. *Language Teaching*. 31(2): 57-71. doi: 10.1017/S0261444800012970
Wolff, D. 1989. Teaching language in context. Proficiency-oriented instruction: Omaggio, Alice C., Boston: Heinle and Heinle Publishers, Inc., 1986, 479 pp. *System*. 17(2): 286-288. doi:10.1016/0346-251X(89)90047-X
Zhu, Z., Liu, L., Ding, G. & Peng, D. 2009. The influence of Pinyin typewriting experience on orthographic and phonological processing of Chinese characters. *Acta Psychologica Sinica*. 41(09): 785-792. doi: 10.3724/SP.J.1041.2009.00785.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS