MSU Working Papers in Second Language

Studies Volume 9, Number 1

2017

Online Chinese Reading Behavior with Pinyin and Zhuyin Transcriptions Introduction and Literature Review

Rachel Lin

TESOL, Michigan State University
wilso657@msu.edu

Abstract

The purpose of this study is to analyze eye movements of second language learners of Mandarin Chinese during online reading to determine patterns of attention. I intend to replicate and expand part of Stickler and Shi's (2014) original eye tracking experiments on Chinese second language learners completing an online reading activity with Chinese characters and Pinyin transcription scaffolding. Since Chinese uses characters, reading can prove a daunting task for learners, especially those at the beginning levels (Wang, 2014). I have personally experienced this frustration myself and observed similar difficulties for other students. Although, many characters do contain some phonetic information, it is usually easier for students to learn to read with the help of a phonetic transcription system such as Pinyin, which uses the Roman alphabet "to approximate the pronunciation of Chinese characters" (ibid, p. 53). Often textbooks and reading materials will annotate some or all of the included Chinese characters in such a transcription system, at least for beginner levels, for the benefit of first or second language learners (See Figures 1 and 2).

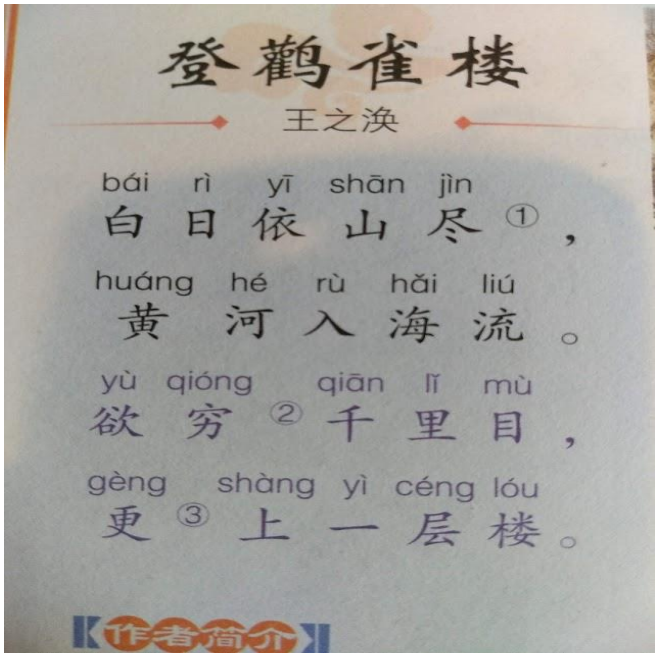


Figure 1. Example of the Chinese poem 登鸛雀樓 by 王之渙 with Pinyin transcription (from a children's book from China).

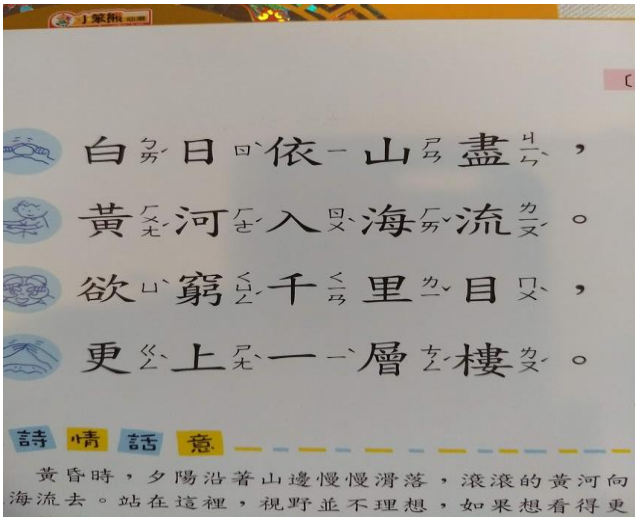


Figure 2.
Example of the same Chinese poem with Zhuyin transcription (from a children's book from Taiwan).

Pinyin is widely used in mainland China, while Zhuyin, another transcription system, is used in Taiwan. Instead of borrowing the Roman alphabet, Zhuyin more closely

resembles parts of Chinese characters. For example, the character 你 is transcribed as nǐ in Pinyin and ㄋㄧˇ in Zhuyin (See Table 1 for more examples).

Table 1.
Comparison of Pinyin and Zhuyin transcriptions.

Chinese Character (Simplified / Traditional)	Pinyin	Zhuyin	IPA	English Translation
苹果 / 蘋果	píng guǒ	ㄆㄧㄥˊ ㄍㄨㄛˇ	pʰǐŋ kuò	apple
香蕉 / 香蕉	xiāng jiāo	ㄒㄧㄤ ㄐㄧㄠ	ɕiáŋ tɕiáu	banana
芭乐 / 芭樂	bā lè	ㄅㄚ ㄌㄜˋ	pá lɤ^	guava
凤梨 / 鳳梨	fèng lí	ㄈㄥˋ ㄌㄧˊ	fêŋ lǐ	pineapple
葡萄 / 葡萄	pú táo	ㄆㄨˊ ㄊㄠˊ	pʰǔ tʰǎu	grapes

Stickler and Shi's results show that lower level learners relied more on Pinyin than higher level learners while reading Chinese characters online. I would like to expand this research by looking for any differences in behavior when characters are supplemented by Zhuyin annotation instead of Pinyin. The results of this experiment may provide more insight into any advantages or disadvantages of using either transcription system while studying Chinese as a second or foreign language.

I will use eye tracking data in this study because it "provides objective and quantitative evidence of the user's visual and (overt) attentional processes" (Duchowski, 2002, p. 455) and has been used extensively in second language acquisition research (see Winke, Godfroid, & Gass, 2013 for a review). For example, Godfroid, Boers, and Housen (2013) used eye tracking to research the attention of unknown words in leisure reading and whether or not this attention contributed to incidental learning of new vocabulary. Two types of eye movements recorded in eye tracking are saccades, or movement from one area to another, and fixations, which are pauses in between saccades. Early measures of eye movements are taken when a subject reads a target passage for the first time, and may provide insight into early stages of reading, which includes word recognition, for example. Revisiting the same text at a later time, such as with backward saccades, is said to provide data pertaining to later stages of the reading process or interruptions in such process (Winke et al., 2013).

In addition to attention, eye tracking methods have been used to research cognitive processes (see Bax & Weir, 2012). Just and Carpenter (1976, 1980) assert that the eye and mind are linked. Therefore, eye behavior denotes cognitive behavior, and fixation duration is equal to the amount of time that a word is being processed (Just & Carpenter, 1980). In addition, cognitive processing dictates eye movement behavior (Godfroid et al., 2013). Word processing cannot occur during saccades because "sensitivity to visual input is reduced" (Rayner, 1998, p. 373). In other words, the eyes move too fast during a saccade in order to perceive any information. Therefore, examining the nature of a reader's eye movements can illuminate their processing of the text (Frenck-Mestre, 2005). During reading, an increase of fixation durations and backwards saccades in relation to the control variable are correlated with interruptions or difficulties in cognitive processing (Winke et al., 2013). Such "processing difficulties are endemic in L2 learners and may be compounded by first language (L1)-L2 differences" (p. 207).

To illustrate, Bassetti's (2006) results show that "second language orthographic input affects the mental representations of L2 phonology in instructed beginner L2 learners" (p. 95). In this study native readers of Roman alphabetic languages who were studying Mandarin Chinese and familiar with Pinyin were found to omit the main vowel in three Chinese rimes ([iou], [uei], [uən]) in their L2 phonological representation, because these main vowels were not represented in the Pinyin (<iu>, <ui>, <un>, respectively). The participants' lack of main vowel pronunciation in these rimes corresponded to previous findings, and Bassetti's reasoning for this is that "the Pinyin orthographic input is reinterpreted according to L1 phonology-orthography correspondences, and interacts with the phonological input in shaping the phonological representations of Chinese syllables in beginner learners" (p. 95).

Complications in second language acquisition due to grapheme-phoneme correspondences were also found in Mathieu (2016), wherein the Arabic word-initial /ħ/ and /χ/ fricatives were contrasted and presented to monolingual native English speakers using three unfamiliar scripts in a word-learning experiment (Arabic, Cyrillic, and a blend of Roman and Cyrillic). There was also a learning condition without any orthography. The results showed that participants in the no orthography condition performed better than those with any other script. There were no significant differences across script conditions, but L1 grapheme-phoneme correspondences may have influenced the results. The Arabic script was found to be an inhibitory factor in L2 phonological acquisition, compounded with the unfamiliar contrast. This was further exacerbated with participants reading the script form left to right as in their L1. The Cyrillic script condition was also inhibitory because some Cyrillic letters appear similar to Roman letters, and may have activated the participants' L1 grapheme-phoneme correspondences. The Roman/Cyrillic blended script used word-initial non-Roman-like

letters in Cyrillic with the remaining part of each word in the Roman alphabet. Since there was no significant difference between this blended condition and the no orthography condition, the researcher maintained that the familiar letters in the words activated L1 grapheme-phoneme correspondences and interfered with the L2 phonological acquisition.

Hayes-Harb and Cheng (2016) compared Pinyin and Zhuyin conditions in three experiments to examine the influence of orthography on L2 word learning based on whether the writing system and grapheme-phoneme correspondences are shared between the L1 and the L2. In the first experiment native English speakers placed in one of the two conditions were exposed to audio and written forms (in Pinyin or Zhuyin) of Chinese words and had to determine whether the word they heard matched the written form. These Chinese words were labeled as either "congruent," meaning that their Pinyin spellings were also possible English spellings (for example, <nai> and [nai]), or "incongruent," meaning the Pinyin form included a familiar grapheme but represented an unfamiliar phoneme in English (for example <xiu> and [ɕiou]). The second experiment had the participants associating the aurally presented words with pictures for meaning. The results showed that the Zhuyin group outperformed the Pinyin group in both tasks due to the interference of incongruent items for the Pinyin condition. A third experiment proved that both groups did not differ in their ability to perceive differences in the target Chinese sounds, and, thus, orthography hindered the Pinyin group's efforts. (For further reading on the influences of L1 grapheme-phoneme correspondences on L2 acquisition see Han & Kim, 2017; Hayes-Harb, Nicol, & Barker, 2010; Showalter & Hayes-Harb, 2015, 2013.)

Teachers do not agree on the duration of when Pinyin should be used in teaching Chinese. Since Pinyin graphemes correspond to the Roman alphabet, native English speakers need to "suppress native language grapheme-phoneme correspondences in favor of new ones" (Hayes-Harb & Cheng, 2016, p. 1). As a result, some teachers believe that Pinyin should be phased out of the classroom as soon as possible, while others see it as an essential scaffold for beginning learners (Everson, 2008). The results of this study may yield further insight into beginning learners' use of Pinyin and Zhuyin that could further inform Chinese practitioners in their pedagogical decisions.

My research questions are:

1. Where is learners' attention focused in online Chinese reading tasks?
2. What learning strategies are used in these tasks?
3. Do the Pinyin and Zhuyin phonetic systems influence the learners' attention in these tasks?

Methods

Participants

Stickler and Shi recruited ten adult learners of Chinese as a second language at the lower levels of proficiency, and further divided them into levels ranging from poor to excellent. I propose to include participants from all four years of the Michigan State University undergraduate Chinese program, who will be familiar with Pinyin. In addition, I will recruit students from the first four years of the National Taiwan Normal University (NTNU) Mandarin Training Center program, as they will most likely have learned Zhuyin. These participants' native language should utilize the Roman alphabet, such as English, French, etc. Although these two programs may not be equivalent, especially since one is in an immersive environment, I will categorize the participants into proficiency levels based on a reading pretest, which will be a mix of questions from a range of levels from the Hànyǔ Shuǐpíng Kǎoshì (HSK) and Test of Chinese as a Foreign Language (TOCFL) tests. Both of these tests are national standardized tests of Chinese proficiency for non-native speakers in China and Taiwan, respectively.

Materials & Procedure

Following Stickler and Shi's methodology, the participants will fill out a background questionnaire to provide basic demographic information such as age, gender, prior language study, familiarity with technology, etc. In the original study the synchronous learning platform *Elluminate* was used. However, it was acquired by Blackboard so I will use *Blackboard Collaborate* to recreate the reading activity used by Stickler and Shi. Using the whiteboard function in this program, students will move independently through a series of screens containing a short paragraph in Chinese characters, either a Pinyin or Zhuyin transcription below that, and three comprehension questions in English (see Figure 3 for Stickler and Shi's original activity). The Chinese characters will either be simplified or traditional, based on what each participant has learned, and the phonetic transcriptions will be located in a similar manner to Figures 1 and 2, with Zhuyin to the right of each character or Pinyin directly below each character.

Figure 3. Example of gazeplot for Stickler and Shi's original reading activity (p. 60).

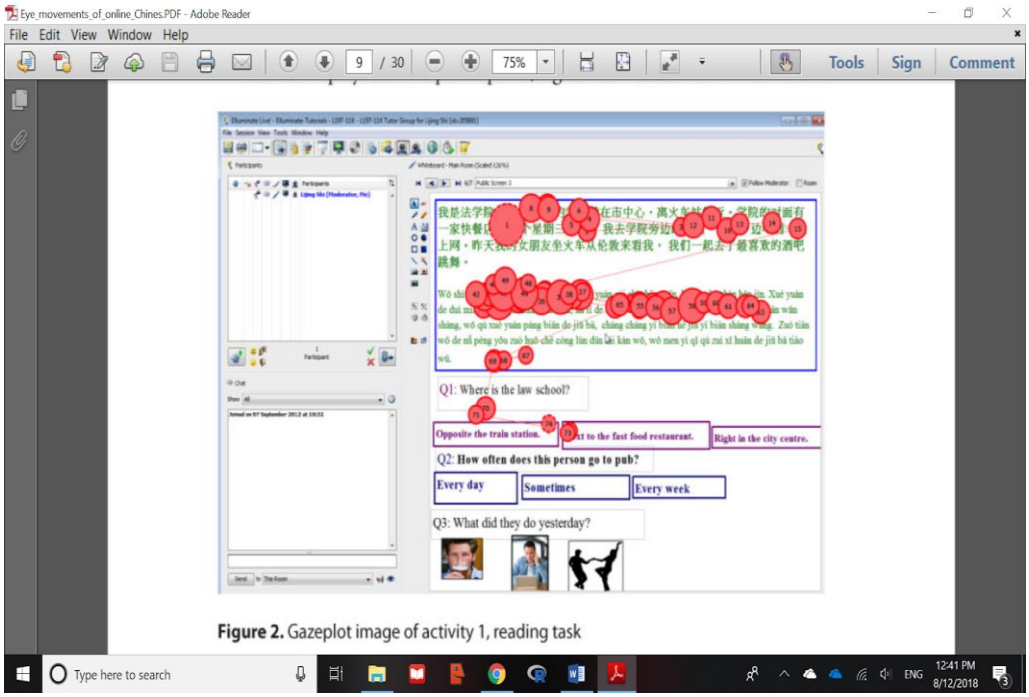


Figure 2. Gazeplot image of activity 1, reading task

I will use Tobii Pro to record the during-task eye tracking data. After the reading task, the participants will watch a replay of their gaze plot and I will follow the stimulated recall method (Gass & Mackey, 2000) to ask them to elaborate on reasons for their behavior. Stimulated recall is a type of introspective method, which assumes that 1) "it is possible to observe internal processes in much the same way as one can observe external real-world events" (p. 11), and 2) one is capable of verbalizing one's own internal processes. In stimulated recall a visual reminder of an event or task is presented to the participant in order to aid him or her in recounting their cognitive processes during said event. I would like to use this method to ask participants to watch a recording of their eye-movements during the task and verbalize where they were focusing their attention and why. For example, if a participant consistently looks at the Pinyin or Zhuyin transcription and not the characters I would ask him or her for the reason behind that behavior and whether any cognitive strategies were being used. Asking participants to recall their thoughts and strategies for this task will provide further insight into possible explanations for their behavior and may add to a richer interpretation of the results.

Analysis

To analyze the data, first I will create a heat map using the eye tracking results. Since this map will designate the main areas of focus for each participant, I will then use the results to create areas of interest (AoIs). Stickler and Shi created AoIs for the Chinese text and the Pinyin, but, due to the difference in layout for the characters and transcriptions in this study, I expect to create AoIs for the characters, Pinyin, and Zhuyin.

Then I will analyze the fixation duration. This information will indicate participants' main focus, whether on the Chinese characters, Pinyin, Zhuyin, or comprehension questions. I will calculate the total number of fixations on specific areas, which should indicate areas of difficulty. Finally, I will investigate the stimulated recall data for common themes and unique explanations of behavior. This analysis procedure closely follows Stickler and Shi's study.

Expected Results

I expect my data will also confirm Stickler and Shi's results that lower proficiency Chinese language learners rely mostly on Pinyin while reading online. However, since Zhuyin does not use the Roman alphabet and more closely resembles Chinese characters, Zhuyin learners may be less likely to rely on phonetic transcription while reading. Therefore, I estimate that the participants seeing Pinyin will rely more on the phonetic transcription than participants under the Zhuyin condition. Stickler and Shi did not analyze data from the comprehension questions, but in this study I can compare both groups' performances on the comprehension questions and include questions in the stimulated recall to ascertain whether or not the Pinyin or Zhuyin transcriptions influenced their answers. However, results from the comprehension questions cannot be solely attributed to these transcription systems because of the differences in each group's background in learning Chinese.

Limitations and Implications for Future Research

Potential limitations for this study are that the quality of education of the two Chinese programs at MSU and NTNU may not correspond with each other. In addition, students in Taiwan have the opportunity for language immersion in their environment, whereas students in Michigan are learning in a foreign language context. One last possible issue could be whether or not the participants have the basic technological literacy to be able to complete the task, but the initial questionnaire could alert the researcher if this will be a problem. A worthy follow up study could include Chinese native speakers at a range of levels who are familiar with either Pinyin or Zhuyin to compare their eye-tracking behavior with the participants in this study. Future research should also explore more longitudinal methods and include other non-native learner populations such as children, teens, and older learners.

References

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy*, *9*, 95–114.

Bax, S., & Weir, C. (2012). Investigating learners' cognitive processes during a computer-based CAE Reading test. *Cambridge Research Notes, Cambridge ESOL*, *47*(February 2012), 3–14.

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Methods, Research, Instruments, and Computers*, *1*, 1–15.

Everson, M. E. (2008). Issues in Chinese literacy learning and implications for teacher development. In P. Duff & E. Lester (Eds.), *Issues in Chinese language education and teacher development* (pp. 70–78). University of British Columbia: UBC Centre for Research in Chinese Language and Literacy Education.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, *27*, 175–198.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Erlbaum.

Godfroid, A., Boers, F., & Housen, A. (2013). Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3).

Han, J.-I., & Kim, J.-Y. (2017). The influence of orthography on the production of alphabetic, second-language allophones by speakers of a non-alphabetic language. *Journal of Psycholinguistic Research*, *46*(4), 963–982.

Hayes-Harb, R., & Cheng, H. W. (2016). The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native english speakers. *Frontiers in Psychology*, *7*(JUN). https://doi.org/10.3389/fpsyg.2016.00785

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367–381.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, *32*(2), 145–170.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Showalter, C., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, *29*(2), 185–200.

Showalter, C., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, *36*, 23–42.

Stickler, U., & Shi, L. (2014). Eye movements of online Chinese learners. *CALICO Journal*, *32*(1), 52–81.

Wang, J. (2014). Reading strategies and reading comprehension retention with and without a pop-up dictionary by beginning learners of Chinese. *CALICO Journal*, *31*(2), 244–260.

Winke, P., Godfroid, A., & Gass, S. (2013). Eye-movement recordings in second language acquisition research. *Studies in Second Language Acquisition*, *35*(2), 205–212.

The effects of Output-Based Focus on Form on Japanese EFL Learners' Implicit Knowledge Development through a Text-Reconstruction Story-Retelling Task

Kiyotaka Suga
TESOL, Michigan State University
sugakiyo@msu.edu

Abstract

This research proposal of a quasi-experimental study aims at exploring the effects of a text-reconstruction output task called Story-Retelling (SR) on Japanese English as a foreign language (EFL) learners' implicit knowledge development on three types of English relative clauses (RCs). Thirty university students are randomly assigned to two experimental groups and a control group. Each experimental group receives two different form-focused instruction (FFI): the SR task and explicit grammar explanation of the target-linguistic form. Responding to a measuring issue of previous effect-of-instruction studies, an Oral Elicited Imitation (OEI) test and an untimed fill-in-the-blank test will be administered to accurately measure learners' implicit and explicit knowledge development respectively. Differential effects of the instructional treatments depending on the degree of typological markedness and processing difficulty of the three RC types are also explored. The findings of the study will provide empirical explanations whether and how an output practice conducted through the SR task contributes to L2 learners' implicit knowledge development. Considering the relative easiness of the practical application of the SR task in L2 teaching, the findings also provide L2 teachers with pedagogical implications on how to effectively conduct focus on form (FonF) instruction in the L2 classroom.

Introduction

One of the most engaging questions in second language (L2) instruction is how to introduce form-focused instruction (FFI) in the classroom context to effectively promote L2 development (Izumi, 2013b; Spada & Lightbown, 2008). Previous Instructed Second Language Acquisition (ISLA) studies have shown that not only receiving 'comprehensible input,' as proposed by Krashen (1982), but also output produced by L2 learners, especially in meaningful contexts, plays crucial roles to facilitate the processes of language learning and then develop learners' linguistic knowledge (e.g., de Bot, 1996; Izumi, 2002a, 2003a; Izumi & Bigelow, 2000; Izumi & Izumi, 2004; Izumi, Bigelow, Fujiwara, & Fearnow, 2000; Kowal & Swain, 1994; Morgan-Short & Bowden, 2006; Muranoi, 2007b; Swain, 1985; 1995, 1998, 2005; Toth, 2006; Uggen, 2012; among others). In spite of the accumulation of these studies on output, what has not been fully investigated is what kind of output practice contributes to the development of L2 learners' proficiency, which enables them to use the language in a spontaneous communication (Morgan-Short & Bowden, 2006; Muranoi, 2007a). Muranoi (2007a) argues that reconstructing a text that learners have comprehended can be predicted to effectively promote the four functions of output (i.e, the noticing, hypothesis testing, metalinguistic, and automaticity functions), and then eventually contribute to the development of L2 proficiency (Swain, 2005).

Along with the importance of output in language acquisition, the type of linguistic knowledge developed through certain types of FFI is also a key issue in L2 teaching as well as L2 research. To help learners use their target language in a real communicative situation, it is crucial to develop their implicit knowledge of the language, which can be used automatically without conscious monitoring of their own language use (Ellis, 2015). Although the development of implicit knowledge is a key in L2 acquisition, Doughty (2003, 2004) pointed out that the outcome of instruction in previous ISLA studies was not precisely measured through the use of valid measurement which can assess L2 learners' implicit knowledge development (except Erlam, Loewen, & Philp, 2009). Based on the results and the limitations of the previous ISLA studies, the present study investigates whether a text-reconstruction output task introduced in the framework of focus on form (FonF) contributes to the development of Japanese EFL learners' implicit knowledge development by using a more valid measure.

Literature Review

FonF in L2 Instruction

FonF is an instructional approach that aims at drawing learner's attention to certain linguistic forms within a meaningful communication (Long, 1991; Doughty & Williams, 1998). This instructional approach was introduced to strike a balance between two opposite instructional approaches: traditional grammar-based teaching and extreme meaning-oriented communicative language teaching (CLT). Traditional grammar-based language teaching, which is often introduced by the combination of detailed rule-explanations and audio-lingual pattern drills, tends to focus only on certain target linguistic forms in less-meaningful, decontextualized manner. On the other hand, the main focus of the strong version of CLT is placed exclusively on the meaning that is conveyed in communication; hence, it is likely that learners pay no attention to the target-linguistic forms. Previous ISLA studies have reported that both of these opposite instructional approaches failed to develop learners' well-balanced communicative competence (Brown, 2000; Izumi, 2002b; Long, 1998; Richards & Rogers, 2014). Thus, dealing with the limitations of these two instructional approaches in the L2 classroom while capturing each strength has been a long-standing challenge in L2 teaching. To overcome this challenge, FonF, which directs learners' attention to certain target-linguistic forms within meaning-focused L2 tasks, has a great potential (Izumi, 2002b).

Measuring Effect of FonF on Learners' Implicit Knowledge Development

One of the important questions in FonF research is what kind of knowledge, either explicit or implicit knowledge, is developed through a particular type of FonF instruction (Izumi, 2013b). While explicit knowledge is the type of knowledge that is consciously available to learners when they have enough time for controlled processing (Ellis, 2005; Ellis, Loewen, Elder, Erlam, Philp & Reiders, 2009), implicit knowledge is not consciously available but can be used intuitively and automatically in real communication without conscious awareness (Ellis, 2005; Loewen, 2015). Because of the obvious advantage of implicit knowledge in real communication, Ellis (2015) claimed that the prime goal of L2 instruction should be developing learners' implicit knowledge.

However, it is still debated whether grammar should be taught explicitly or implicitly to achieve the ultimate goal. Norris and Ortega's (2000) meta-analysis showed that explicit type of instructions led to higher effect size than implicit type of instructions. However, Doughty (2003, 2004) pointed out that most of the studies which were treated in Norris and Ortega's (2000) were biased because the research period of these studies was very short and the measurement tasks were decontextualized, which is advantageous for measuring learners' explicit knowledge. This methodological issue needs to be further addressed by empirical studies that employ a type of measurement that can accurately assess L2 learners' implicit knowledge development.

Ellis (2005) investigated the reliability and validity of the following five measurements for explicit and implicit knowledge: an oral elicited imitation (OEI) test, an oral narrative test, a timed grammaticality judgement test (TGJT), an untimed grammaticality judgement test (UGJT), and a metalinguistic knowledge test. The results of the study revealed that the first three tests were found to be reliable and valid measurements for implicit knowledge while the latter two tests were found to be better at measuring explicit knowledge. Particularly, he concluded that the OEI test can solidly measure learners' implicit knowledge. Other similar studies also supported the conclusion of Ellis' (2005) study (e.g., Bowles, 2011; Erlam, 2006; Kim & Nam, 2017; Spada, Shiu, & Tomita, 2015). Based on the findings from a range of psychometric validation studies, Ellis (2015) proposed that the following four criteria are the key-factors to create more valid measuring-tools of implicit and explicit knowledge: (1) degree of awareness (intuition or rule knowledge), (2) time availability, (3) focus of attention (meaning or form), and (4) degree of utility of knowledge of metalanguage.

L2 Knowledge Development through Output-Based FonF Tasks

In addition to the measurement issue of FonF research, effects of diverse types of FonF studies need to be investigated further (Izumi, 2013b), especially output-based FonF instructions that is conducted in meaningful L2 classroom contexts (Morgan-Short & Bowden, 2006; Erlam, et al., 2009). Muranoi (2007a) claims that "having learners reconstruct a text (story) that they have comprehended is one of the most effective instructional techniques that elicit learner output and eventually promote L2 learning" (p. 67). Based on the four primary functions of output: noticing, hypothesis-testing, metalinguistic and automaticity functions (see de Bot, 1996; Izumi,

2003a; Swain, 2005 for more comprehensive reviews of the four functions of output), Muranoi (2006) introduced Story-Retelling (SR) task, which is a type of text-reconstruction task in which learners are asked to reconstruct a text that they have comprehended using a concept map. The concept map is a type of lexical representation which indirectly guides learners to use specific target-linguistic forms. Contrary to the dictogloss task (see Kowal & Swain, 1994; Swain, 1998; Swain & Lapkin, 1995), the first comprehension phase of the SR task does not involve listening and learners' self-notetaking. Since L2 learners tend to primary focus on lexical elements while engaging in production tasks (Swain & Lapkin, 1995; Hanaoka, 2007; Hanaoka & Izumi, 2012; Uggen, 2012), providing the lexical items using a concept map in the SR task enables learners to direct their attention to the target-grammatical forms more systematically, and guides them to use these forms during the retelling than does the dictogloss task, which often fails to direct learners' attention to specific target-grammatical forms.

One study that investigated the effects of this type of text-reconstruction task is Muranoi (2007b), which examined the effects of text-summary writing task termed *focus on form through guided summarizing* (FFGS). In FFGS, learners are asked to summarize a reading-text using a concept map. In this task, the learners are directed to summarize the text twice. Between the first and second summarizing performances, an interval reflection-time was given to promote learner-noticing by comparing their first summarizing-performance with the original text. The results of the study revealed that FFGS had a positive effect on learners' accuracy development in the use of English perfect passive and the effects of the task held over two months. He attributed these positive effects of FFGS to the facilitative roles of output on learners' cognitive processes of L2 acquisition, such as induced-noticing, comparing, and hypothesis formulation and testing (see Gass, 1988, 1997 for detailed discussions on the cognitive processes of L2 acquisition). However, the study reported that the positive effects of FFGS was found only on the learners who were psycholinguistically ready to the target form.

Another study that explored the impacts of output through a text-reconstruction task is Izumi (2002a), which investigated whether a text-reconstruction task, an input enhancement technique or the combination of these two treatments promotes adult ESL learners' noticing and the acquisition of English relativization. In this study, the researcher compared the effects of instruction through four different measuring tasks: a sentence combination task, a picture-cued

sentence completion task, an interpretation task, and a grammaticality judgment task. The study found that pushed output coupled with subsequent input-processing promoted learners' noticing and led to the acquisition of the target linguistic form (English relativization).

Based on the implications drawn from the previous text-reconstruction studies, the SR task can be predicted to promote noticing and contribute to the development of learners'L2 knowledge. However, paucity of empirical research investigated whether and how the SR task promote learners' L2 knowledge development. The only study that tested the effects of the SR task is Suga (2016). This study is a small-scale research that explored the effectiveness of the SR task in an L2 classroom context. In this study, seven Japanese high school EFL learners went through four 90-minute instructional-treatment-sessions over one month. The effects were measured through two types of untimed oral- and written-fill-in-the-blank tests. The results of the study showed some positive effects on the learners' accuracy development on learners who received relatively high scores in the pretest. However, the instruction did not seem to have positive effects on the learners who received very low pretest-scores. One interesting finding from the follow-up interviews was that both in oral- and written-tests, the learners who improved their performances in the post-test phases worked on the tests implicitly. One student commented that even though she could not verbalize the reason of her answer, she was able to work on the tests relatively smoothly, relying on her intuition. However, the biggest limitation of the study was the test-types used to measure the learners' performances. In the study, although both oral- and written-responses were elicited, the effects of the instructional treatment were measured only through an untimed fill-in-the-blank tests, both of which cannot precisely measure learners' implicit knowledge (Doughty, 2003, 2004). Thus, even though some learners' comments implied the development of their implicit knowledge, it was difficult to specify whether the SR task truly contributed to the development learners' implicit knowledge. Another methodological limitation of this study was that the factors that truly contributed to the learning gains were not specified because of the small sample size, which made it difficult to employ a group comparison design. These results and methodological limitations of the study generated a need for further research with a more rigorous research design.

The only output-study that employed precise measurements to assess explicit and implicit knowledge development is Erlam, Loewen, and Philp (2009), which compared the effects of

meaning-oriented output instruction and processing instruction using an OEI test and an Untimed GJT. The results showed that both output instruction and processing instruction had positive effects on both implicit and explicit knowledge. This study revealed the facilitative effects of meaningful output on L2 learners' implicit knowledge development by the use of a more accurate measurement tools.

As reviewed in this section, producing output facilitates learners' noticing on the target-linguistic forms (e.g., Hanaoka, 2007; Russell, 2014; Song & Suh, 2008; Uggen, 2012) and may have some positive effects on the development of L2 knowledge. However, the effectiveness of output-based instruction is greatly influenced by various factors, such as learners' developmental readiness (Muranoi, 2007b), task designs (Izumi & Izumi, 2004), and the degree of difficulty and complexity of the target forms (Izumi, 2002a, Izumi & Bigelow, 2000).

Although it is quite challenging to determine the degree of difficulty of target linguistic forms, there are several ways to make predictions on it. One of the most researched and supported predictions on the degree of linguistic-difficulty is the Noun Phrase Accessibility Hierarchy (NPAH), which hypothesized the universal order of linguistic difficulty of relativization from most accessible to least accessible based on the typological relationship of relative clauses (RCs) (Keenan & Comrie, 1977). The order predicted by the NPAH is: subject-type (SU), direct-object-type (DO), object-of-preposition-type (OPREP), genitive (GEN), and object-of-comparison (OCOMP). Another prediction that helps researchers to determine the degree of difficulties is Kuno's (1974) Perceptual Difficulty Hypothesis (PDH). Since effects of the matrix-position of the embedded clause was not considered in the NPAH, the PDH focused on learners' working memory capacity and claimed that any RC-types embedded in the subject position is more difficult to process than the RCs embedded in the object position, requiring more short-term memory capacity. One implicational study is Izumi (2003b), which investigated how the typological markedness and processing difficulty of English RCs interact task-types (i.e., a sentence combination, interpretation, and grammaticality judgement task). The results of the study revealed that L2 learners' performances on different English RC types are largely influenced by the combination of the place of embedded RC clause and task-demands (production or comprehension). The findings of Izumi (2003b) allow detailed examination of learning effect in relation to typological complexity, processing difficulty of the target-linguistic

form (i.e., English relative clauses), and learners' developmental stages. However, few previous studies investigated effects of output on learners' implicit knowledge development in relation to the typological markedness and complexity of the target linguistic forms.

Motivated by the findings and limitations of the previous studies, the present study explores whether the SR task conducted in the framework of FonF has positive effects on Japanese EFL learners' implicit knowledge development using a type of test that can accurately measure learners' implicit knowledge development.

Research Questions and Research Hypotheses

1. Does Story-retelling (SR) develop Japanese EFL learners' implicit knowledge on the target linguistic forms: RC (SU), RC (DO), and RC (OPREP)?
2. Do the two types of instructions (SR and explicit grammar instruction) have differential effects on the development of two types of different knowledge (implicit knowledge and explicit knowledge)?
3. If SR develops Japanese EFL learners' implicit knowledge on the target linguistic forms, will the effect hold over the post-test period (two weeks)?
4. Does SR have differential effects on Japanese EFL learners' implicit knowledge development depending on the complexity of the target linguistic forms: RC (SU), RC (DO), and RC (OPREP)?

Hypothesis 1. As shown in the previous output studies, the SR task have positive effects on Japanese EFL learners' performances on both the OEI test and the untimed fill-in-the-blank test. Also, the effects facilitated by the SR task hold over until the two-week delayed post-test (Muranoi, 2007b).

Hypothesis 2. Explicit grammar explanation contributes to Japanese EFL learners' performances only on the Untimed fill-in-the-blank test in the immediate post-test, but the positive effects of the explicit grammar explanation do not hold over until the delayed post-test. Even though the provision of explicit knowledge through rule-explanations helps learners to understand specific grammatical features (Erlam, 2013), the effects of explicit instruction usually disappear very quickly (Doughty, 2004).

Hypothesis 3. The SR contributes to Japanese EFL learners' implicit knowledge development on all the three types of English RCs in both immediate and delayed post-tests, while explicit grammar instruction has significant positive effects on their explicit-knowledge-development only on the simpler and less-marked RC types: RC (SU) and RC (DO), but not on RC (OPREP) in the immediate post-test. Dekeyser (2005) argues that implicit type of instruction is advantageous for complex grammatical rules while explicit instruction is good for simple rules. Although it is highly complicated procedures to decide which grammatical forms are more complex and which forms are not (see Ellis, 2006; Dekeyser, 2005, 2016; Spada & Tomita, 2010), the present study decided the degree of difficulty of the three RC-types based on the typological markedness and processing difficulties of the RC-types reported by Izumi (2003b).

Study Methodology

In this study, two types of form-focused instructions (i.e., the SR task and explicit grammar explanation) are implemented, and the effects of the instructions are explored through a pretest, immediate post-test, and delayed post-test (see Figure 1 for the summary of the research design). To explore whether SR and explicit grammar explanation have differential effects on two types of different knowledge (i.e., implicit knowledge and explicit knowledge), an oral elicited imitation (OEI) test and an untimed fill-in-the-blank test are used. An exit-questionnaire is also given after delayed post-test to supplement the test-results.

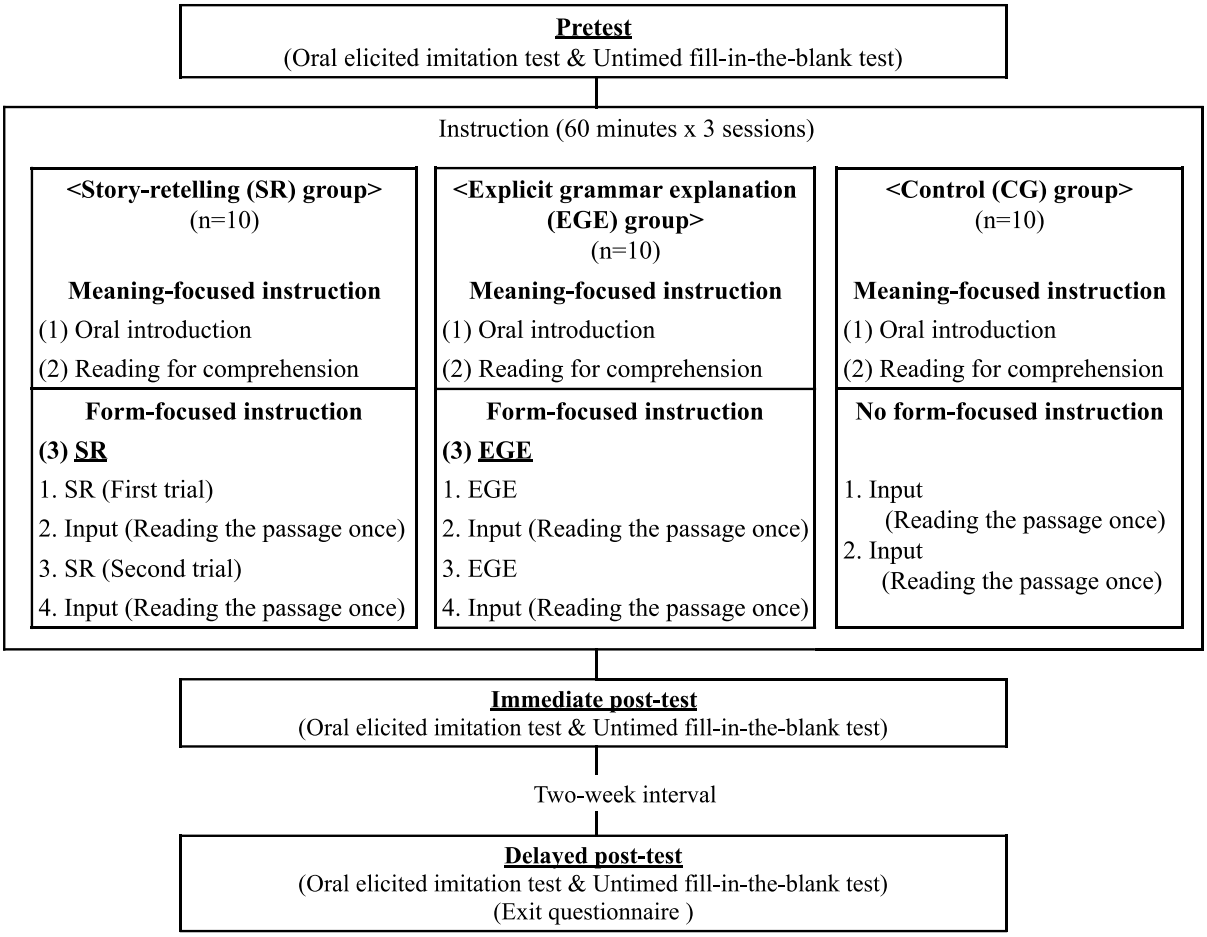


Figure 1. Research Design

Participants

The participants of the study are Japanese EFL university students majoring foreign language studies. To create three groups: the SR group, the explicit grammar explanation (EGE) group, and the control group, 30 students (10 students for each group) will be recruited from a university in Tokyo. Proficiency level of the participants are intermediate, ranging between 800 to 600 on the Test of English for International Communication (TOEIC). In this university, all the participants are required to take the TOEIC test regularly. Based on their TOEIC scores, the participants will be randomly assigned to the three groups.

Since the participants are university students, all of them have received formal instruction on the targeted linguistic forms (i.e., three types of English relative clauses) at least once in their

previous formal education. The proficiency factor and the amount of existing knowledge are very important to conduct an effect-of-instruction study. Previous ISLA studies have shown that partially acquired knowledge of the target linguistic form is a prerequisite condition to develop L2 leaners' implicit knowledge through meaningful production practices (Ellis & Shintani, 2014; Williams & Evans, 1998). Thus, the target population may be good candidates to measure the impact of the SR task on learners' implicit knowledge development.

Target Forms

The impact of the instructional-interventions is tested on the acquisition of three types of English restrictive RCs: RC (SU), RC (DO), and RC (OPREP). The main reason that these three types of English RCs are chosen as the target-linguistic forms of the study is that English RCs have been studied extensively in previous SLA studies (e.g., Ammar & Lightbown, 2005; Doughty, 1991; Hamilton, 1994; Izumi, 2002a, 2003b; Izumi & Izumi, 2004; Keenan & Cormrie, 1977; Kuno, 1974, Spada & Tomita, 2010), which provide rich source of information to analyze the impact of the SR task on learners' implicit knowledge development in relation to their psychological readiness, the formal complexity and each processing difficulty of the target-linguistic forms.

Instructional Treatments

In the instructional treatments, three training sessions will be conducted for three days. All the instructions are conducted by the present researcher. During the three 60-minute treatment sessions, three short passages consisting of about 300 words for each passage, each of which includes the same number of target linguistic forms, are introduced; each session uses one of these passages. In each instructional treatment, the SR group first receive background knowledge for comprehension and vocabulary glosses, both of which are purely meaning-oriented and any kinds of form-focused instruction related to the target-linguistic forms are not given. After reading the passage once and then fully comprehend the text, the participants of the SR group are instructed to engage in the SR task twice using a concept map (a schematic representation of keywords and phrases), which indirectly guides learners to use the target-linguistic forms in the task (see Figure 2). During the first and second trials of the SR task, the

learners are not allowed to access the original text. All the SR performances are audio-recorded to further examine whether each participant used the target-linguistic forms and what kind of cognitive processes happened during the SR performances. After performing the SR task each time, the participants are asked to read the original passage once. The EGE group also goes through the same procedures until the comprehension phase finishes. After the comprehension phase, the EGE group receives metalinguistic explanations on the target-linguistics forms in their first language (Japanese) using each RC clause in the text. Since the SR group engages in SR twice in the task phase, the explicit instruction group is also provided the explanation twice. First, they receive an oral-explanation. Then, they read the summary of the explicit explanation given in Japanese as the second explanation. The control group only read the passage twice in the task phase without engaging in any additional form-focused task.

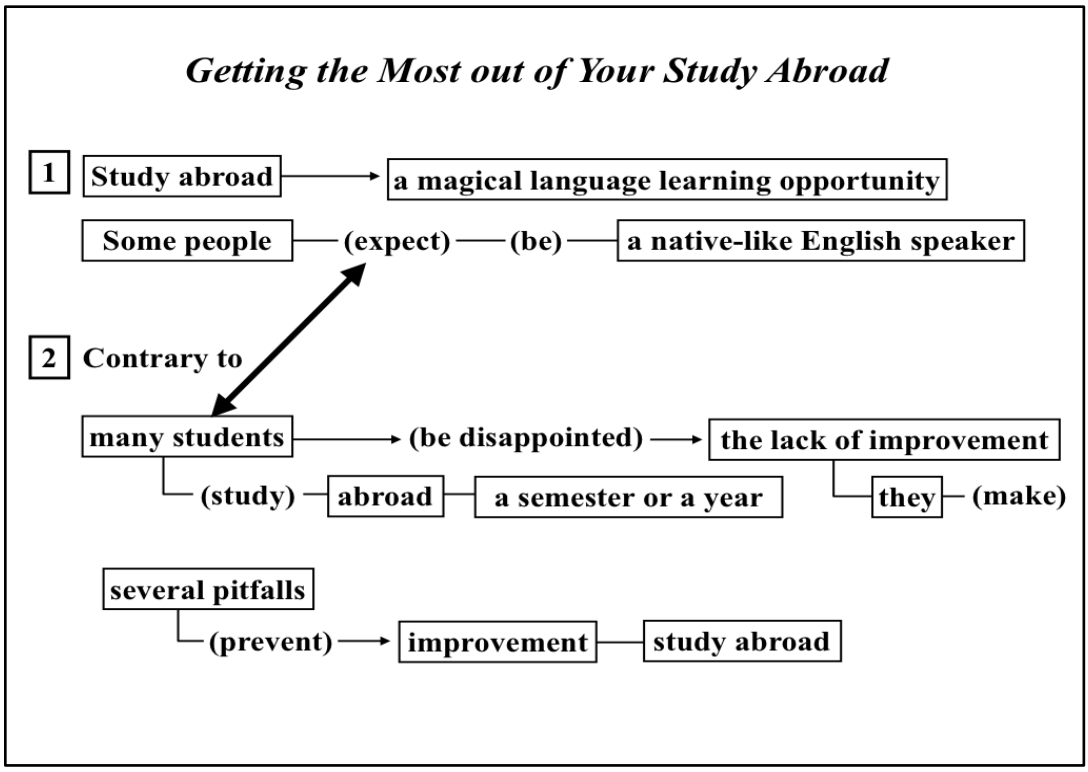


Figure 2. An example of the concept map used in the SR task

Testing Instruments

The effects of the treatment are measured in three testing phases (pretest, immediate post-test, and delayed post-test) using two different types of test: an OEI test, whose requirement of online production and the time pressure forces the participants to use their implicit knowledge (Ellis, 2005; Erlam, 2006, 2009) and an untimed fill-in-the-blank test, which allows learners to consciously access to their explicit knowledge. The OEI test consists of 36 statements, in which twelve statements are created for each type of RCs: SU, DO, and OPREP. For each twelve statements, eight statements are novel statements for the participants and four statements are quoted from the reading texts used in the instructional treatments. Informed by Erlam (2006) and Erlam et al. (2009), the OEI test of this study uses a belief questionnaire, which directs learners' focus on the meaning of each statement. First, learners listen to a statement and then are asked to decide whether the statement is true, false, or whether they are not sure. After judging the content of each statement, learners are asked to quickly repeat the statement as accurately as possible (see Erlam et al., 2016 for detailed explanation of OEI test). The untimed fill-in-the-blank test is also created with 36 items and for each RC type, twelve items are created. Similarly, eight items are created with novel sentences and four sentences are taken from the text used in the treatment sessions. One version of each test is used in the three testing phases, but the order of the statements is changed each time. The time-gap between the immediate post-test and delayed post-test is two weeks.

Scoring and Analysis

The results from these two types of tests are analyzed quantitatively using both descriptive and inferential statistics. In order to analyze the homogeneity of the three groups, one-way analyses of variance (ANOVAs) will be performed based on the participants' TOEIC scores and their scores on the OEI and untimed fill-in-the-blank tests. All data from the pretest and the two post-tests are submitted to mixed-design analyses of variance (ANOVAs) using one between-subjects (Instruction) and one within-subjects (Time) factors. The between-subject factor (Instruction) has three levels: SR group, explicit explanation group and control group, and the within-subject factor (Time) also has three levels: pretest, immediate post-test, and delayed post-test. The participants' exit questionnaires will also be analyzed qualitatively to supplement

the analysis of quantitative results.

Possible Limitations

This study is limited by several methodological problems. First, the study cannot be generalized due to the small number of participants. Each group has 10 students, which is a quite small number of students compared to many EFL classrooms. Second, the time between the immediate post-test and delayed post-test are two weeks, which might not be reasonable to claim that the results of the delayed post-test reflect on long-term effects of the instructional treatments. Because the development of implicit knowledge takes time, how L2 learners' noticing in the subsequent input-processing opportunities and how their implicit knowledge develops over time needs to be further investigated in the future research with a longitudinal research design. Third, although two different tests are used to measure the impacts of the SR on implicit and explicit knowledge respectively, the validity of these test is still being debated among SLA researchers (e.g., Ellis, 2005; Erlam, 2006; Suzuki & Dekeyzer, 2015; Spada, Shiu, & Tomita, 2015). Although the OEI test has been supported as a valid measure of implicit knowledge, recent studies started to challenge the findings of the previous studies by claiming that the OEI test does not actually measuring implicit knowledge but rather measuring another construct, which is automatized explicit knowledge (Suzuki and Dekeyser, 2015). However, the focus of the study is not to identify whether the knowledge acquired through the SR task is implicit knowledge or automatized explicit knowledge. Dekeyzer (2003, 2017) states that automatized explicit knowledge and implicit knowledge are functionally indistinguishable in communicative interaction. Thus, even if the knowledge that the OEI test is measuring is not exactly implicit knowledge but highly proceduralized explicit knowledge through the use of careful laboratory experiment, learners can still use the knowledge in their spontaneous communication. This is what most L2 instruction aims at.

Conclusion and Pedagogical Implications

The purpose of the study is to investigate whether and how a text-reconstruction output task promotes the development of L2 learners' implicit knowledge in relation to the degree of complexity and processing difficulty of the target-linguistic forms. Despite the methodological

limitations discussed in the previous section, the findings of the study will provide empirical explanations on effects of output practice on L2 learners' implicit knowledge development with the use of a more valid measure of implicit knowledge. By comparing the effects of the SR task with the provision of explicit grammar explanation, the results of the study may also provide L2 practitioners with pedagogical implications by showing both strengths and weaknesses of each type of instruction in the two different performance-requirements. If the results of the study reveal effectiveness of introducing the SR task on learners' implicit knowledge development, it is valuable to consider including this task as one of the viable options to conduct output-based FonF instruction that eventually promotes learners' well-balanced communicative competence.

References

Ammar, A., & Lightbown, P. M. (2005). Teaching marked linguistic structures- More about the acquisition of relative clauses by Arab learners of English. In A. Hausen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 167-198). Berlin: Mouton de Gruyter.

Brown, H. D. (2000). *Principles of language learning and teaching*. White Plains, NY: Longman.

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning,* 46, 3, 529-555.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), The handbook of second language acquisition (pp. 312–348). Oxford, UK: Blackwell.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language learning*, *55*(S1), 1-25.

DeKeyser, R. (2016). Of moving targets and chameleons-Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition, 38(2),* 353-363.

Dekeyser, R. (2017). Knowledge and skill in ISLA. In S, Loewen, & M, Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 15-32). New York: Taylor & Francis.

Doughty, C. (1991). Second language instruction does make a difference. *Studies in Second Language Acquisition, 13*, 431–469.

Doughty, C. (2003). Instructed SLA: Constraints, Compensation, and Enhancement. In C. J. Doughty & M. Long (Eds.), *The handbook of second language acquisition*, (pp.256-310), Oxford: Blackwell.

Doughty, C. (2004). Effects of Instruction on Learning a Second Language: A Critique of Instructed SLA Research. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form meaning connections in second language acquisition* (pp. 181-202). Mahwah, US: Taylor and Francis.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*, *27*(2), 141-172.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied linguistics*, *27*(3), 431-463.

Ellis, R., Loewen, S. D., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. (Volume 42). Mulilingual Matters.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464–491.

Erlam, R. (2013). Explicit knowledge and grammar explanation in second language instruction. In C. Chapelle (Eds.), *The encyclopedia of applied linguistics.* Oxford, England: Wiley Blackwell. doi:10.1002/9781405198431.wbeal.0404

Erlam, R., and Akakura, M. (2016). New developments in the use of elicited imitation. In A. Mackey, & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 94-111). Taylor and Francis.

Erlam, R., Loewen, S., & Philp, J. (2009). The roles of output-based and input-based instruction in the acquisition of L2 implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching*, 241-261.

Gass, S. (1988). Integrating research areas: A framework for second language studies1. *Applied linguistics*, *9*, 2, 198-217.

Gass, S. (1997). *Input and interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hamilton, R. (1994). Is implicational generalization unidirectional and maximal? Evidence from relativization instruction in a second language. Language Learning, 44, 123-157.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research*, *11*(4), 459-479.

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21(4)*, 332-347.

Izumi, S. (2002a). Output, input enhancement, and the Noticing Hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition 24*: 541-77.

Izumi, S. (2002b). EFL education in japan from the perspective from the second language acquisition research, *Sophia Linguistica, 49,* 3-14.

Izumi, S. (2003a). Comprehension and production processes in second language learning. *Applied Linguistics,* 24, 2, 168-196.

Izumi, S. (2003b). Processing difficulty in comprehension and production of relative clauses by learners of English as a second language. *Language Learning* 53, 285-323.

Izumi, S. (2013a). Noticing and L2 Development: Theoretical, Empirical, and Pedagogical Issues. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka, (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 37-50). Honolulu: University of Hawai'i, National Foreign Language Resource Center.

Izumi, S. (2013b). Focus on form (FonF). In P. Robinson (Eds.), *The Routledge encyclopedia of second language acquisition*. Routledge.

Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition?. *Tesol Quarterly*, *34*(2), 239-278.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21*, 421–452.

Izumi, S., and Izumi, Y. (2004). Investigating the effects of oral output on the learning of relative clauses in English: Issues in the psycholinguistic requirements for effective output tasks. *The Canadian Modern Language Review, 60(5),* 587-609.

Keenan, E., and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8*, 63-99.

Kuno, S. (1974). The position of relative clauses and conjunctions. *Linguistic Inquiry, 5*, 117-136.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam/Philadelphia: John Benjamins.

Long, M. H. (1998). Focus on Form in Task-based Language Teaching. *University of Hawai'i Working Papers in English as a Second Language, 16 (2),* 35-49.

Long, M., and Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.) *Focus on form in classroom second language acquisition*

(pp. 15-41). New York: Cambridge University Press.

Loewen, S. (2015). *Introduction to instructed second language acquisition.* New York: Rutledge.

Morgan-Short, K., & Bowden, H. W. (2006). Processing instruction and meaningful output-based instruction: effects on second language development. *Studies in Second Language Acquisition*, *28*(1), 31-65.

Muranoi, H. (2006). *Dainigengosyutokuriron kara mita koukatekina eigo gakusyuho shidoho [Effective English learning and teaching methods from the perspective of second language acquisition].* Tokyo: Taisyukan.

Muranoi, H. (2007a). Output practice in the L2 classroom. In R. Dekeyser (Eds.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 51-84). New York: Cambridge University Press.

Muranoi, H. (2007b). Focus on form through guided summarizing and EFL learners' interlanguage development. *Journal of the English institute, 33,* 15-59. Took Gakuin Uniersity.

Pinemann, M. (1998). *Languae processing and second language development: Processability theory*. Amsterdam: Benjamins.

Russell, V. (2014). A closer look at the output hypothesis- The effect of pushed output on noticing and inductive learning of the Spanish future tense. Foreign Language Annals, 47(1), 25-47.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

Spada, N. and Tomita, Y. (2010). Interaction between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60,* 2, 263-308.

Spada, N., Shiu, J. L. J., and Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*(3), 723-751.

Suga, K. (2016). *The effects of story-retelling in PCPP instruction: Output practice in meaning oriented focus on form.* Unpublished M.A. thesis, Sophia University, Tokyo.

Suzuki, Y., and DeKeyser, R. M. (2015). Comparing elicited imitation and word monitoring as

measures of implicit knowledge. *Language Learning*, *65*, 860–895.

Swain, M. (1985). Communicative competence: Some role of comprehensible input and comprehensible outputting tis development, In S, Gass and C. Medden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, Mass.: Newbury House.

Swain, M. (1995). There functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125-144). Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-481). New York: Cambridge University Press.

Swain, M. and Lapkin, S. (1995). Problems in output and cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371-91.

Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. *Language learning*, *56*(2), 319-385.

Uggen, M. S. (2012). Reinvestigating the noticing function of output. *Language learning*, *62*(2), 506-540.

Williams, J., and Evans, J. (1998). What kind of focus and on which forms?. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139-155). Cambridge University Press.

Processin g Variability in Intentional and Incidental Word Learning: a Follow-up of Solovyeva and Dekeyser (2018)

Bronson Hui
Second Language Studies, Michigan State University

huibrons@msu.edu

Abstract

Inspired by Solovyeva and DeKeyser (2018), I utilized Coefficient of Variation (CV) (Segalowitz & Segalowitz, 1993), a measure traditionally used in detecting automatization and recently extended to index addition of new linguistic representations, to capture the trajectory in processing variability in both intentional and incidental word learning. This paper reports two studies involving (1) an intentional word learning experiment and (2) a re-analysis of published eye-tracking data from an incidental vocabulary learning study (i.e., Elgort et al., 2017). In the word learning experiment, native English speakers (N = 35) studied Swahili-English word pairs (k = 16) before performing ten testing blocks of animacy judgment tasks on the Swahili words. This design differed from that of previous studies which often focused on products of learning (i.e., typically, adopting a pre- / post-test or cross-sectional design) (e.g., Leow, 2015). By computing a CV value from the reaction time (RT) data for each participant in each testing block, I captured the learners' development in processing stability during the process of learning. Results replicated an initial increase in CV as new representations were established, indicating less stable processing of the Swahili word meanings in the beginning. At the same time, CV peaked at about the 6th block before decreasing, which was then consistent with automatization. This inverted U-shaped development was confirmed by the quadratic term of Block in the mixed-effects model. In the second study, I computed CVs for each participant from Elgort et al.'s eye-tracking experiment where learners gained lexical, orthographic familiarity through repeated exposures in natural reading. These CVs were based on individual participants' first fixations and gaze durations on the target words at each of the first 12 occurrences in the reading text. CV based on gaze durations significantly decreased over time only when the analysis was restricted to target words. CV based on first fixations did not change in all

analyses. Taken together, these results suggested relative stability in CV in word learning in natural reading contexts. Discussion highlighted the importance to investigate the process of learning, differences in processing demands of incidental and intentional word learning, and the signature of CV across different stages of vocabulary acquisition.

Learners need to process linguistic information automatically to become fluent, accurate language users (DeKeyser, 2015; Segalowitz, 2010). On this account, Segalowitz and Segalowitz (1993) introduced the measure of coefficient of variation (CV) to the field of second language acquisition (SLA) to measure processing automaticity. A CV value reveals the variability of an individual's reaction times (RTs) obtained from experimental, linguistic tasks, such that a small CV value would index more stable processing, one of the three major elements in automatic processing (Segalowitz, 2010). To achieve automaticity, according to skill learning theory, learners go through a developmental process, known as automatization (see DeKeyser, 2015 for a review). According to this view, learning a language is similar to learning a general skill in that learners first acquire declarative knowledge (e.g., grammar rules); then proceduralize the knowledge (put the rules into actual language use); and finally automatize the knowledge through practice (become automatic in processing). Over the years, traditional predictions have been that when automaticity develops, along with proficiency, processing should become more stable as reflected by a decreasing value of CV (i.e., reaction times become less variable) (Hulstijn, Gelderen, & Schoonen, 2009). This prediction was, however, not always supported in the research base (e.g., Hulstijn et al., 2009; Lim & Godfroid, 2015). At least some of the discrepancy may be attributed to the differences in the developmental stages of participants between studies. For example, lower-level learners' processing could be dominated by the acquisition of declarative and procedural knowledge; hence automatization is less likely to be observed (Hulstijn et al., 2009; Lim & Godfroid, 2015). On this account, the prerequisite to first acquire declarative knowledge in automatization highlights the need for researchers to capture learners' development throughout a fuller trajectory.

Related to this need to capture language development in entirety, CV has also been extended to investigating addition of new linguistic representations, stretching further from the narrow reading of automatization in skill learning theory that CV was originally proposed (Solovyeva & Dekeyser, 2018). Shedding light on the signature of CV during the initial stage of word learning, Solovyeva and Dekeyser (2018) found an increase in CV (less stable processing) as learners first engaged in learning novel words. The authors reasoned that when improvements in processing stability (a decrease in CV value) resulted from more efficient processing routes, as skill learning theory proposed, addition of new representations, then, was responsible for the deterioration of processing stability (an increase in CV). Their findings, therefore, casted doubts upon the traditional, universal prediction of a decreasing CV as proficiency develops. Importantly, their work pointed to a possibility that CV could potentially be an indicator of addition of new representations. On this account, a logical extension of this strand of work is to investigate how CV changes, if at all, in different word learning paradigms (i.e., intentional and incidental word learning) and when involving different aspects of word knowledge (e.g., meaning and orthographic forms). Inspired by Solovyeva and Dekeyser (2018), this present study attempted to capture the potential change in CV during a fuller trajectory of word learning processing. This paper reports two studies: (1) an intentional word learning experiment, looking at intentional learning of word meaning and (2) re-analyses of published data from an eye-tracking experiment on contextual word learning, investigating incidental learning of orthographic forms.

Literature Review

Coefficient of Variation as a measure of automaticity

Segalowitz and Segalowitz (1993) were the first to propose Coefficient of Variation (CV) as a measure of processing stability, a key characteristic in processing automaticity (Segalowitz, 2010). Given the important role of processing automaticity in skill acquisition theory (see DeKeyser, 2015 for an overview), this measure needs to be valid, reliable in order verify (or falsify) theoretical claims about learners' processing trajectory (e.g., learners become more automatic in processing as proficiency develops). CV, is derived from an individual's response time (RT) distribution in a behavioral task such as a lexical decision task. Specifically, it is calculated as the standard deviation divided by the mean RT of a participant. As such, CV reveals the variability of the same individual performing the same task repeatedly while controlling processing speed. A small CV value indexes more stable processing because of the smaller variability of the learner's RTs. Due to the role of both speed and stability in automatic processing, CV data have often been analyzed together with RT data, generating three general predictions as automaticity develops (Hulstijn et al., 2009): first, a decrease in mean RT (i.e., improving processing speed), second, a decrease in CV (i.e., improving processing stability), and third, a positive correlation between RT and CV.

Relationship between CV and Proficiency

Hulstijn et al. (2009) argued that not many studies, if any at all, have confirmed all three predictions, providing "the complete, hard within subject evidence" (p. 564) for automatization. One issue is concerned about whether the learner must have strong linguistic knowledge as a prerequisite for automatization. For example, Segalowitz and Segalowitz (1993), in their seminal paper, reported different results for French learners of English across skill levels. In Experiment 2, participants performed a visual lexical decision task with 90 repetition items (15 words over six times). RTs obtained from the repetition items correlated with CVs both at the beginning and at the end of the experiment, but only for the more skillful learners at the outset. RTs of those initially slow correlated with their CVs only at the last presentation. No changes in CV values were reported. Further, Segalowitz, Segalowitz, and Wood (1998) reported correlations between RTs and CVs for both initially fast and slow participants, but a decrease in CV values was only found in the initially fast group.

While these two studies showed differences in automatization of participants of different skill levels, other studies revealed effects of materials difficulty. Segalowitz, Watson, and Segalowitz (1995) reported data of a Turkish learner of English (aged 26) who performed a series of lexical decision tasks on 120 base words across four frequency bands over three weeks. Results showed a downward trend in CVs only in the lowest frequency words. Similarly, Akamatsu (2008) trained Japanese learners of English to recognize 150 English words over seven weeks. The pre- and post-training lexical decisions showed a decrease in CV and correlation between RTs and CVs only for low-frequency words. The inconsistencies in the research base highlighted the importance of taking the learners' current developmental stage in relation to the experimental materials into consideration.

This issue of learner's proficiency being a confounding factor is more apparent when comparing across studies. Lim and Godfroid (2015) conceptually replicated Hulstijn et al.

(2009), and these two studies found quite different results. First, in Hulstijn and colleagues' study, participants were Dutch high school learners of English who performed lexical decision and retrieval tasks, as well as sentence verification and production tasks over two years. Contrary to expectations, decrease in CVs was found only in two out of eight tasks, and RTs did not always correlate with CVs, neither in the expected direction (positively) nor in a significant manner. In contrast, Lim and Godfroid found a decrease in CVs from intermediate to advanced learners to native speakers in sentence processing. Though some correlations were not significant, the authors concluded that their findings were largely consistent with automatization. According to Lim and Godfroid, the Dutch learners in Hulstijn et al. "were still in a relatively early stage of their language learning trajectory, in which declarative knowledge accumulation (vocabulary and grammar learning) dominated" (p. 1266).

This comment, together with the discrepancy in the literature, highlighted the theoretical prerequisite of strong linguistic knowledge in automatization, and more importantly, a compelling need to capture how CVs change in the course of a fuller developmental trajectory, from no-knowledge, through acquisition of declarative and procedural knowledge, to automatization. Inclusion of the early stages in the investigation is crucial because, first of all, they are the bases upon which later automatization develops, at least according to skill learning theory. Second, a predicted downward CV is meaningful only when all other directions of change have been attested (Solovyeva & Dekeyser, 2018).

Processing Stability in the Earliest Stages of Word Learning

Elgort (2011) investigated the early stage of word learning by investigating the effectiveness of deliberate novel word learning using CV as a learning outcome measure of automaticity. Advanced L2-English learners were presented 48 pseudowords in the lab and given a set of word cards for study at home for a week. Participants returned to take part in three priming experiments testing their acquisition of lexical-formal and semantic knowledge of these target words. Results showed participants' higher degree of automaticity in processing these target words than low-frequency non-target words and other nonwords (which they had had no exposure to). Although the author was interested in deliberate learning of novel words, the main experiments were not conducted until the end of the week after learners had studied the target words intensively. In other words, the design of the study did not capture some of the very first lexical processing in the development, measuring only products of learning (Leow, 2015).

Another study that investigated specially at the initial word learning processing in relation to CV changes was Solovyeva and DeKeyser (2018). One important contribution of this study was extending the use of CV to capture how processing patterns changed when new representations were being established. The authors hypothesized that if a decrease in CV resulted from elimination of component processes (reorganization or restructuring), addition of new representations should result in an increase in CV. The study involved 73 native English-speaking university students. Word learning was operationalized in two aspects: first, learning new form-meaning pairings of 40 Swahili words, and second, learning the lexical-formal knowledge of 20 nonwords. In one single experiment, primed lexical decision tasks were conducted before and after participants' deliberate study of the 40

Swahili-English translations. Again, the design was set up to detect products of learning, not to investigate the actual learning process (Leow, 2015). In the pre-training priming task, the learner was first presented a Swahili word as the prime, followed by its English translation as the target. The participant needed to make a lexical decision on the English target. Note that at this point of testing, the learner did not know they were going to learn the Swahili primes deliberately. But this pre-training testing offered a baseline measure of processing the Swahili words from no-knowledge. In the post-training priming task, half of the prime-target pairs were changed, which created a (semantically) mis-matched condition. The unchanged, original prime-target pairs formed the (semantically) matched condition. These two conditions provided the basis for the semantic priming task. Any facilitation of the Swahili prime in the lexical decisions on the English translation equivalent in the matched condition served as evidence of learning of the form-meaning association of the Swahili word. Importantly, these lexical decisions afforded RT, and hence CV measures for analysis. Results showed an increase in CVs and a decrease in RTs comparing the pre- and post-training sessions. These changes meant that, as a result of training, the learner processed the words faster, but in a less stable manner (as demonstrated in the greater variability in RTs), confirming the authors' initial hypothesis.

In this study, the second aspect of word learning (i.e., learning of lexical-formal knowledge) was embedded in the same experiment, but in the nonword trials. Since participants encountered these nonwords a total of four times in the tasks, the authors used the RT data to nonwords (hence the CV data) to investigate potential learning of lexical-formal knowledge of these nonwords as a result of multiple exposures. Results also showed an increase in CVs and decrease in RTs. However, as also acknowledged by the authors, a major limitation of this part of the study was the conflict between learners becoming more familiar with the nonwords, and this familiarity serving as the measure of learning while participants still had to treat them as nonwords in the experiment (requiring a NO response). The conceptual tension could influence whether this growing familiarity could be considered as word learning. To mitigate this limitation, the authors also reanalyzed published data from two studies (Bartolotti & Marian, 2014, Brown & Gaskell, 2014) to add further support to their claim: that increased variability as measured by CV may index addition of representations during its initial stage. As Solovyeva and DeKeyser have shed light on the initial end of the learning spectrum, it is also important to note that the semantic priming task in this study was a test of products of learning, as opposed to the process of learning (Leow, 2015). This distinction between process and products of learning is highly crucial, especially in this strand of work looking at automatization and addition of new representations. If processing stability decreases as a result of establishing new representations, researchers should be more interested in capturing this change as new representations are *being* established, less so after they are already established. As such, there is a need for a study to capture the impact of establishing new representations on processing stability, and its ramification after that, covering the entire development trajectory, from no-knowledge, through acquisition of declarative and procedural knowledge, to automatization. With the available evidence in the research base (e.g., Lim & Godfroid, 2015; Solovyeva & DeKeyser, 2018), one logical prediction is an initial increase in CV, revealing addition of new

representations, followed by automatization of such knowledge as revealed by a decrease in CV in the later trajectory.

Apart from the need to capture the entire developmental trajectory, another limitation of this strand of work has been its reliance on behavioral tasks such as lexical decision or semantic categorization tasks, involving motor responses of the participants. While data elicited from these tasks are able to index processing, their application to the learning of lexical formal word knowledge has been less straightforward, as pointed out by the previous review of Solovyeva and DeKeyser (2018). Perhaps because of this challenge, a lot of the work done in this area has focused on intentional learning of word meaning (e.g., Elgort, 2011). At the same time, it has also been argued that vocabulary is often learned incidentally; that is as a result of other cognitive processes involving comprehension (Gass, 1999). On this account, it is important to understand how processing changes also in the incidental learning paradigm which often involves the learning of lexical formal word knowledge (Elgort, Brysbaert, Stevens, & Van Assche, 2017; Godfroid, Ahn, Choi, Ballard, Cui, Johnston, Lee, Sarkar, & Yoon, 2017). If CV could index addition of new representations, researchers should attest the extent to which such an increase is universally applicable across different word learning paradigms (intentional and incidental) and across different aspects of word knowledge. First, word learning in the incidental learning paradigm often involves a shallower depth of processing of the target words because the primary task for the learner is to comprehend the text. Engagement with unfamiliar words, then, depends on factors such as the overall context and more global understanding of the text. This difference in depth of processing in incidental learning is then in sharp contrast with the intentional paradigm where learners are typically told to learn the materials by heart. On this account, could such shallower processing still have an impact on processing stability as measured by CV? Perhaps related to this shallower processing, the more robust findings in the investigation of incidental word learning have been the learning of orthographic forms (Elgort et al., 2017; Godfroid et al., 2017). If this is the case, can CV also index addition of new orthographic forms in the mental lexicon?

Despite the difficulty in using behavioral tasks to investigate the incidental learning paradigm, one increasingly useful way adopted by researchers interested in incidental vocabulary learning is eye tracking (e.g., Elgort et al., 2017; Godfroid et al., 2017; Godfroid, Boers, & Housen, 2013; Mohamed, 2017; Pellicer-Sánchez, 2015). Eye tracking records eye movements to capture participants' attentional processing, based on the eye-mind link assumption (Reichle, Pollatsek, & Rayner, 2012). This methodology has also been argued to be one of the closest operationalizations of natural reading with great ecological validity because of its noninvasive nature (Godfroid & Winke, 2015). Typically, researchers interested in incidental word learning asked participants to read an experimental text in which target words were embedded. Without notice, the learners then took part in vocabulary tests which assessed their learning during reading. This learning was then associated with their reading behavior as captured by eye tracking. Godfroid et al. (2017), for example, found total reading times statistically significantly predicting scores in the off-line, explicit meaning vocabulary posttest. Also, the researchers found a nonlinear decrease in reading times when learners encountered unfamiliar lexical items repeatedly. This decrease in reading times served as evidence of learning of lexical formal word knowledge. Similarly, Elgort et al.

(2017) reported the learners' quick and reliable development of orthographic processing after approximately eight exposures in reading. Taken together, reading times have been used to index processing of new words in natural reading. On this account, then, computation of CVs based on reading times could reveal learners' processing variability when they see the same target words repeatedly, shedding light on CV's signature in incidental word learning, involving often lexical formal word knowledge.

The Present Study

Given this literature review, the present study had two aims: first, to capture the change in processing stability during the entire developmental trajectory of novel word learning, from absolute no-knowledge, through acquisition of declarative and procedural knowledge, to automatizing this newly established knowledge; second, to extend the research base of using CV to index new formation of representations to incidental vocabulary learning of orthographic forms by applying the measure of CV in analyzing eye-tracking data.

Research Questions

These research questions guided the present study:

RQ1: What is the pattern of change in CV over time in the case of:

(a) intentional learning of novel word meanings?
(b) learning of orthographic word forms through natural reading?

To address the two aspects of the research question, this paper reports two studies: first, addressing RQ1(a), an intentional word learning experiment where participants deliberately learned the meaning of target words; second, addressing RQ1(b), re-analyses of data from a published eye tracking study investigating incidental learning of orthographic word knowledge through reading (Elgort et al., 2017).

Study I: Intentional Word Learning Experiment

Methodology

Participants. The initial sample was 40 native English-speaking undergraduate students at a Mid-Western university in the US. Five participants were excluded due to unsatisfactory performance (see Data Analysis). The final sample consisted of 35 participants (mean age = 20, 74% female). All of them were not enrolled in any formal foreign language class at the time of participation and reported no knowledge of any African languages. They received extra credit for a class for their participation. Ethical clearance was obtained from the Institutional Review Board (IRB) in according to the University's regulations governing research involving human participants.

Materials. Sixteen English-Swahili translation equivalent pairs (e.g., boy-mvulana) were used. Highly frequent, concrete, and imaginable nouns were selected based on the SUBTLEXUS data base for written frequency and concreteness and imageability ratings on the MRC psycholinguistic database (Coltheart, 1981). The mean frequency in English of the target words was 166 (to provide a reference, frequency of *bed* was 187). Both means for the

concreteness and imaginability were larger than 600 (out of 700). These selection criteria were meant to make it possible for the participants to learn from scratch and proceed to automatization in one single laboratory session because frequent, concrete, imaginable words were easier to learn (e.g., Walker & Hulme, 1999). Half of the stimuli had an animate reference (e.g., *boy*). The descriptive statistics of the characteristics of these pairs are presented in Table 1 (see also Appendix A for all stimuli).

Table 1. Descriptive statistics of characteristics of the experiment word pairs

	Mean (SD)	Range
English word length	4.19 (1.22)	3-7
English written frequency (SUBTLEXUS)	166.32 (183.66)	22-557
English concreteness[1] (100 - 700)	613.31 (8.85)	600-635
English imagability[1] (100 - 700)	623.25 (10.43)	602-638
Swahili word length	5.19 (1.72)	2-8

[1]Concreteness and imaginability ratings were based on the MRC psycholinguistic database (Coltheart, 1981), such ratings were not available for four words

Procedure. After consent, participants proceeded to the experiment. The experiment was conducted in a quiet lab and run by SuperLab 5, a piece stimulus presentation software. Overall, there was one learning block and ten testing blocks. The entire procedure took about 20 minutes.

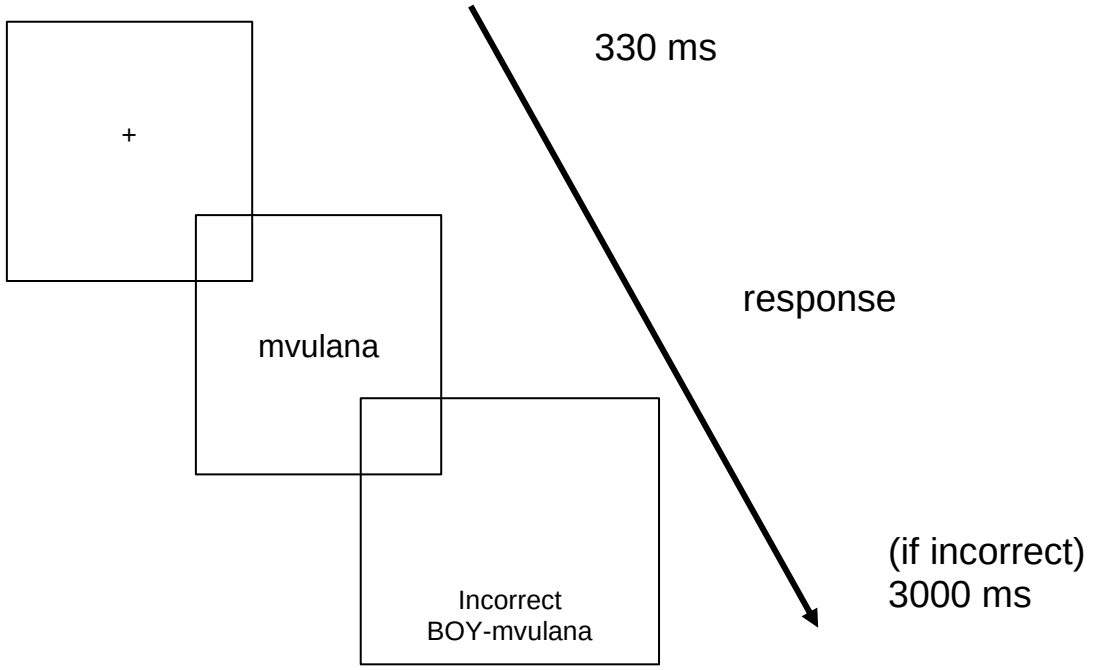


Figure 1. Order of presentation in a testing trial

In the learning block, participants were presented with all Swahili words with their English translation equivalent for eight seconds, one at a time. The experiment proceeded automatically. This block was immediately followed by ten testing blocks, in each of which the participant performed animacy judgement tasks on all 16 Swahili words presented in the learning phase. Participants judged if the words were living things and hit the corresponding button on the keyboard (i.e., "l" for living things and "a" for non-living things) as quickly and accurately as possible. Each trial started with a fixation cross (+) presented for 330 ms, followed by the target word presented in the lower case (e.g., mvulana). The target word remained on the screen until a response by the participant. Visual feedback (*Incorrect* and the Swahili-English pair) was given for 3000 ms only when an incorrect judgement was made. No feedback was given for the correct trials. Figure 1 summarizes the procedure for each trial in the testing blocks. Each word pair was tested once in each block. The order of trials was pseudo-randomized in SuperLab 5 in each block.

Data Analysis. To ensure engagement in the task, participants who did not consistently reach an 80% accuracy in the 7th block and beyond were excluded from the final analysis. As a result, five participants were excluded. Only correct responses were analyzed. RT data trimming involved removing negative RTs (0%) and those slower than 3000 ms (5.5%). At the participant level, RTs were further trimmed by winsorization from the top and the bottom of the distribution. This procedure replaced extreme values with values at the trim and 1 minus trim quantiles (Wilcox, 2005) and was conducted using the winsor() function in R's psych package (Revelle, 2018). The trim quantile was specified as 0.1, meaning that RTs faster than the 5th and slower than the 95th percentiles were replaced by values at the the 5th and 95th percentiles respectively. A CV value was computed for each participant at each testing block, based on their RTs to the target words. The resulting dataset were then used to build multi-level, mixed effects regression models, using lme4 package in R (Bates, Mächler,

Bolker, & Walker, 2015). The outcome variable was specified as CV. The primary predictor was Block (e.g., Testing Blocks I, II, and so on). Based on visual inspection of the observed data, the quadratic term of Block (i.e., Block ^ 2) was added as an additional predictor in order to capture any potential curvilinearity in the data. Random effects were included to account for dependence due to participant. Specifically, model building started with maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). In the present case, it included random intercepts by participant, and random slopes of all fixed-effect factors by participant (i.e., Block by participant and the quadratic term of Block by participant). Backward model selection involving removal of random-effects factors was based on the Akaike criterion (AIC) values, an indicator of model fit adjusted for model complexity. A model with the lower AIC value was always selected unless there was significantly better fit as revealed by an ANOVA F-tests. Model selection involving removal of fixed-effects factors was based on p-values. With alpha set at .05, the non-significant factor with the highest p-value was removed until there was significantly better fit as revealed by an ANOVA F-tests. Both selection procedures resulted in a final model for interpretation.

Results

Table 2 shows the descriptive statistics of accuracy, reaction time, and CV across the ten testing blocks of animacy tests. In general, accuracy increased, indicating that learning took place across the blocks. Reaction times decreased, showing faster processing by the participants over time. For CV, there was an initial increase trend followed by a plateau and then a potential decrease. The final mixed-effects model indicated both Block and its quadratic term were significant (see Table 3 for model summary). The statistical significant curvilinearity (i.e., the quadratic term of Block) confirmed that the initial increase levelled off and was potentially followed by a decrease. Since the random effect structure allowed the intercept (i.e., theoretical, initial processing stability) and the slope (i.e., developmental trajectory of CV) to vary by participant, I then plotted the observed CV values and those predicted by the model to inspect variability in the data and general fit at the participant level (see Figure 2). In general, it appeared that models fitted the observed values reasonably well for most participants. In Figure 3, I plotted the distributions and the means of the observed and predicted CV values across blocks. Although the quadratic term of Block was significant in the model, there was a need to investigate the nature of the curvilinearity, which could potentially represent a simple level-off or a significant decrease. To this end, I restricted the analysis to the second half of the trajectory (i.e., from Block 6 and beyond, see the vertical dotted line in Figure 3). I build an additional model which included Block as the only predictor and the same random-effects structure as the final model described above (i.e., random intercepts and random slopes of Block by participant). This additional model confirmed a significant decrease ($p = .038$) from Block 6 to Block 10 (see model summary in Table 3). Summarizing the results, the data suggested that CV increased initially during the first stages of intentional word learning, followed by a peak at around Block 6 and then a decrease afterwards.

Discussion of these results is reserved after the presentation of the second study of the present paper: reanalyses of published eye-tracking data from an incidental word learning

experiment. These reanalyses would address RQ 1(b) in respect to the change of CV in the learning of orthographic forms in natural reading contexts.

Table 2. Descriptive Statistics of the Intentional Word Learning Experiment

Block	1	2	3	4	5	6	7	8	9	10
RT (Mean, ms)	1485	1330	1253	1169	1084	1025	999	981	953	936
RT (SD, ms)	405	334	309	242	223	214	250	252	227	220
RT(CI, ms)	139	115	106	83	77	73	86	87	78	76
CV (mean)	0.195	0.262	0.293	0.324	0.315	0.335	0.326	0.322	0.319	0.301
CV (SD)	0.078	0.078	0.079	0.082	0.086	0.080	0.092	0.104	0.116	0.104
CV (CI)	0.027	0.027	0.027	0.028	0.029	0.028	0.032	0.036	0.040	0.036
Accuracy (Mean)	0.667	0.720	0.796	0.864	0.920	0.938	0.948	0.964	0.966	0.980
Accuracy (SD)	0.472	0.449	0.404	0.343	0.272	0.242	0.222	0.185	0.181	0.139
Accuracy (CI)	0.039	0.037	0.033	0.028	0.023	0.020	0.018	0.015	0.015	0.012

Table 3. Model Summaries for Intentional Word Learning Experiment

	Overall (Blocks 1 – 10)			Second Half (Blocks 6 – 10)		
	Beta	95% CI	p	Beta	95% CI	p
Fixed Effects						
Block	0.50	[0.32, 0.66]	<.001	-0.0074	[-0.015, -0.00070]	.038
Second-Order Block (Block ^ 2)	-0.53	[-0.66, 0.40]	<.001			
Observations		350			175	
R^2 (Marginal) / R^2 (Conditional)		0.15 / 0.59			0.011 / 0.57	
AIC		-826			-359	

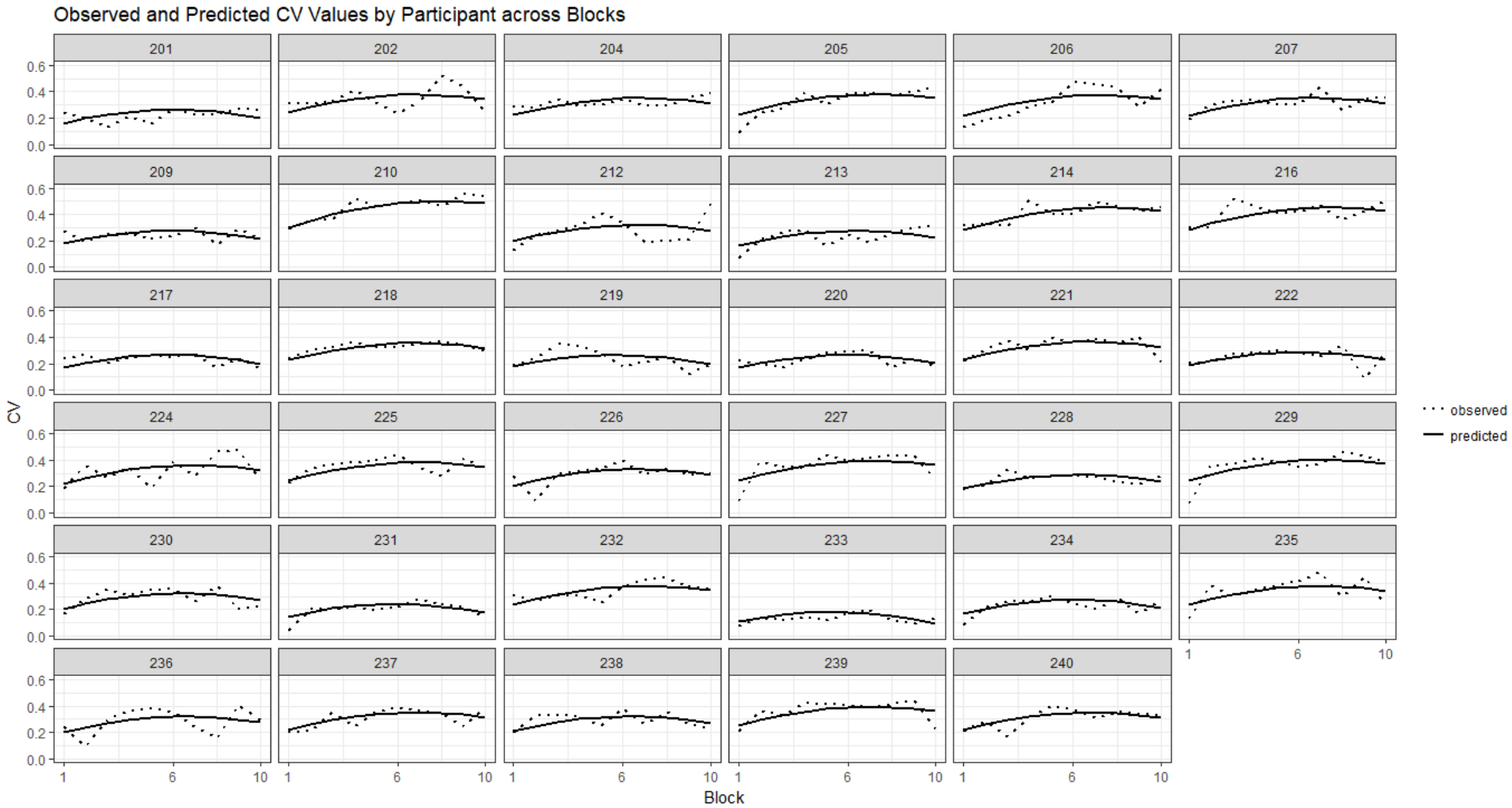


*Figure 2***.** Observed and Predicted CV Values by Participant across Blocks

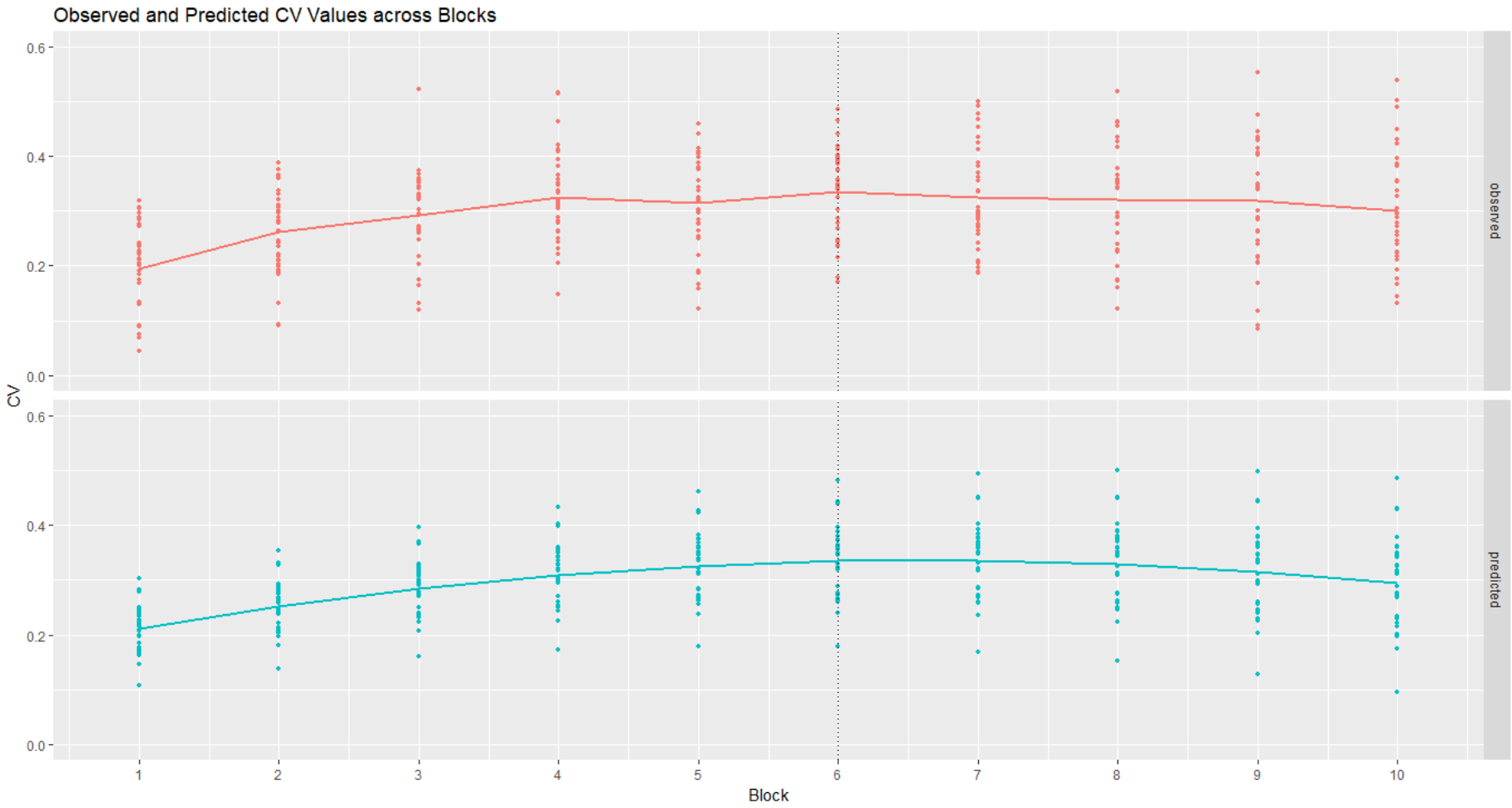


Figure 3. Observed and Predicted CV Values across Blocks

Study II: Reanalyses of Incidental Learning Eye-tracking Data

Description of the study by Elgort et al. (2017)

Using eye tracking, the authors investigated vocabulary learning in natural reading contexts through the patterns of change in eye movements on target words appearing in a text multiple times. Participants were 34 Dutch learners of English who read an expository text embedded with low-frequency target words (k = 14) and high-frequency control words (k = 9). Data analysis in their study focused on the first eight occurrences of each target word. The most relevant findings was a decrease in time spent of the target words in such measures as first fixations and gaze durations, total reading times, and number of fixations. These decreases were interpreted as evidence that learners had gained orthographic familiarity "relatively quickly and reliably" (p. 2). Building upon these findings, the present analysis intended to investigate if CV could index this gain in familiarity or the learning of orthographic forms in natural reading contexts.

Data Collection and Analysis

Eye-movement data collected in Elgort et al. were generously made available by the authors. Data analysis of the eye-movement data in the present study was based on the first 12 occurrences in the text (as opposed to eight in Elgort et al.). This decision was made in accordance with the major purpose of the present study: to capture the fuller developmental trajectory; as such, inclusion of four more occurrences was thought to be able to capture further along the developmental trajectory. Four target words which occurred fewer than 12 times in the original study were removed from the analysis. The analysis also included only two early measures of eye movements: first fixations and gaze durations which index such early processes in reading as word recognition or lexical access (Godfroid & Winkie, 2015). This decision was in line with the aim of the part of the study: investigating the learning of orthographic word forms (as opposed to meaning). The data were not further trimmed or cleaned because it was already done by Elgort et al.

The statistical procedure included, first, computation of a coefficient of variation value for each participant at each order of occurrence of the words (e.g., the first time they saw the words). This CV value was based on their eye-movement measures across all ten target and nine control words. This CV value was specified as the outcome variable for all statistical models. R package lme4 (Bates et al., 2015) was used to fit the data to mixed-effects models. Order of Occurrence (e.g., first / second time reading the words), Word Type (i.e., target vs control words), and their interaction term were included as the primary predictors of interests. Based on initial inspection of the observed data, I decided not to include polynomial terms of the Order of Occurrence. Radom effects structures and model selection were the same as the modeling in the intentional word learning experiment described above in Study I of the present paper.

Results

Tables 4 and 5 present descriptive statistics of both CV measures based on first fixations and gaze durations across occurrences of the target and control words. At

descriptive levels, there was seemingly a decrease in CV based on gaze durations on target words across occurrences; but this potential decrease needed to considered relative to the control words. CV measures based on first fixations appeared to be comparatively stable across occurrences. In the mixed-effects model where CV based on gaze durations was specified as the outcome variable, only word type made it to the best-fitting model, indicating frequency effects (see Table 6 for model summary). In other words, Order of Occurrence (p = .23 in the rival model) and the critical interaction between the two (p = .49 in the rival model) were both not significant. The lack of interaction revealed that the readers did not process the target words in a more stable manner through multiple exposures, compared with the control words. This was the case even though Order of Occurrence emerged significant when the analysis was restricted to only target words (p = .035). In short, though there appeared to be a general decrease in CV based on gaze durations at descriptive levels, such a decrease was not robust when compared against control words.

For CV based on first fixations, again, only Word Type made it to the best-fitting model (see Table 6 for model summary). Order of Occurrence (p = .48 in the rival model) and the interaction between the two (p = .20 in the rival model) were not significant. Restricting the analysis to only target words, Order of Occurrence was also not a significant predictor (p = .25 in the rival model). The overall findings from both CV measures suggested a relative stability in the CV data, meaning that the readers did not process the target in a more (or less) stable manner (i.e., similar variability of processing times) as they read the words repeatedly, despite their gains in orthographic familiarity of the words found in the original study by Elgort et al..

Summary of Results

To capture an overall picture of both studies in this paper (i.e., the intentional learning experiment and the reanalysis of eye-tracking data), results are briefly summarized here. In the intentional word learning experiment, participants were first presented with 16 Swahili-English pairs, before they took part in ten testing blocks where they made animacy judgements on the same Swahili words. CV values were computed for each participant in each block. Results showed an initial increase trend was accompanied by a later decrease captured by the quadratic term of Block in the final model as well as the additional model restricting the analysis to the second half of the trajectory (i.e., from Block 6 onwards). The entire trajectory appeared to resemble an inverted U-shape. In the re-analyses of the eye-tracking data obtained from natural reading contexts. Early eye-movement measures at the first 12 occurrences of the target and control words in the reading were used to compute a CV value for each participant at each occurrence. Both CV measures based on first fixations and gaze durations were stable across all 12 occurrences.

Overall Discussion

Building upon previous work on the signature of CV in early word learning (i.e., Solovyeva & DeKeyser, 2018), this study set out to capture any change in CV as a measure of processing stability along the entire trajectory of novel word learning in both intentional and incidental learning paradigms. First and foremost, the focus of this study on the process of learning (as opposed to the products of learning) allowed me not only to replicate the

initial increase in CV in intentional word learning reported in Solovyeva and DeKeyser, but also to capture a quadratic learning curve in the development of processing stability.

The increase in CV reported in Solovyeva and DeKeyser (2018) was in sharp contrast with the general prediction that CV would decrease as proficiency develops (Hulstijn et al., 2009). As Solovyeva and DeKeyser noted, the original idea in Segalowitz and Segalowitz's (1993) argumentation was that language learners process linguistic input in a more stable manner along the trajectory. However, this prediction was not entirely supported by the data in Solovyeva and DeKeyser, and to some degree, the data in this study too. Qualifying their results, Solovyeva and DeKeyser discussed the heterogeneity in automatization in the lexicon. As known words are automatized, processing could become more stable; yet, a decrease in CV may not be observed because, at the same time, novel word learning could act as an "equalizing force" (p. 11). This opposite force resulted from the addition of new form-meaning links which could elicit more variable responses (i.e., less stable processing). The direction of change in CV along the trajectory then depends on strength of these two forces relative to each other. Previous studies adopting a pre- / post-treatment design often failed to capture the relationship between these forces in relation to developmental stages, leading to predictions and (a lack of) confirmations of a decrease in CV.

Moving beyond the binary predictions, the data of this study, however, demonstrated fine-grained curvilinearity in the signature of CV in early stages of intentional word learning. My data suggested that the force resulting from establishing new representations was initially stronger, driving an increase in CV. But it was only true up until the 6th testing block where CV appeared to have peaked. During this initial period, a closer inspection of Table 2 revealed that participants actually processed faster and responded more accurately; yet in a less stable (i.e., more variable) manner. This trajectory patterns well with skill learning theory which proposes that learners need to first acquire declarative and procedural knowledge (as observed in faster and more accurate responses), before proceeding to automatization as observed in the second half of the blocks. Potentially, this could explain some of the inconsistency in the literature when a decrease in CV was not found. As noted by Lim and Godfroid (2015), for example, when declarative and procedural knowledge dominated development, learners had not proceeded to automatization, hence a decrease in CV should not be predicted.

Table 4. Descriptive Statistics of First Fixations in the Eye-Tracking Experiment

Order of Occurrence	1	2	3	4	5	6	7	8	9	10	11	12
First Fixation – Target												
FF (Mean, ms)	252	255	260	259	252	269	238	247	252	245	227	244
FF (SD, ms)	35	37	43	41	53	88	38	38	44	39	33	36
FF (CI, ms)	12	13	15	14	18	31	13	13	15	13	11	13
	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
CV (mean)	43	24	45	28	38	31	33	26	16	19	45	07
	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.0
CV (SD)	24	90	32	94	00	17	12	07	99	06	26	87
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CV (CI)	43	32	46	33	35	41	39	37	34	37	44	30
First Fixation) – Control												
FF (Mean, ms)	231	224	241	228	246	248	250	233	241	221	237	236
FF (SD, ms)	34	34	31	37	46	35	29	38	39	37	37	38
FF (CI, ms)	12	12	11	13	16	12	10	13	14	13	13	13
	0.2	0.3	0.3	0.2	0.3	0.3	0.3	0.2	0.3	0.3	0.3	0.3
CV (mean)	76	31	29	94	32	04	18	86	35	14	24	11
	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0
CV (SD)	83	13	07	99	48	12	63	98	34	09	85	98
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CV (CI)	29	40	37	34	53	39	57	34	47	38	30	34

Table 5. Descriptive Statistics of Gaze Durations in the Eye-Tracking Experiment

Order of Occurrence	1	2	3	4	5	6	7	8	9	10	11	12
Gaze Durations – Target												
GD (Mean, ms)	348	337	338	343	340	313	316	322	310	294	277	286
GD (SD, ms)	71	73	78	62	63	90	51	55	66	56	60	62
GD (CI, ms)	25	26	27	22	22	31	18	19	23	20	21	22
	0.4	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.3	0.3	0.4	0.3
CV (mean)	78	24	30	11	97	95	31	26	97	93	81	75
	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CV (SD)	31	24	51	41	20	18	55	06	31	41	51	47
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CV (CI)	46	43	53	49	42	41	54	37	46	49	53	51
Gaze Durations – Control												
GD (Mean, ms)	272	278	288	270	287	296	309	281	270	249	273	265
GD (SD, ms)	58	50	55	57	45	52	51	80	51	49	55	52
GD (CI, ms)	20	18	19	20	16	18	18	28	18	17	19	18
	0.3	0.4	0.3	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
CV (mean)	69	30	50	02	59	44	73	49	70	67	66	69
	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1
CV (SD)	12	35	10	41	33	27	47	98	44	42	05	15
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CV (CI)	39	47	38	49	47	44	51	34	50	50	37	40

Table 6. Model Summaries for Eye-Tracking Data

	CV based on First Fixations			CV based on Gaze Durations		
	Beta	95% CI	p	Beta	CI	p
Fixed Effects						
Word Type	0.017	[0.0041, 0.030]	.024	0.057	[0.039, 0.076]	<.001
Observations	814			814		
R^2 (Marginal) / R^2 (Conditional)	0.0057 / 0.096			0.045 / 0.11		
AIC	-1277			-976		

Importantly, the significance of the quadratic term of Block captured the in inverted U-shaped trajectory that peaked at the 6th block. This study is probably one of the first to report a decrease after an increase in CV, capturing at least a fuller trajectory. The decrease after Block 6 was then consistent with automatization. At this stage and beyond, skill learning theory would predict that processing becomes more efficient, and restructuring of processing routes could then enhance stability (DeKeyser, 2015). However, I acknowledge the decrease in CV from the peak could have been captured even more clearly, perhaps further towards and beyond its initial value at Block 1, if there had been more numbers of blocks in the experiment. Even more practice could potentially lead to further automatization and decrease in CV. However, a large number of testing blocks was not adopted because my pilot study suggested that participants lost interests very quickly once they could score a high accuracy. I then struck a balance between keeping participants engaged with the task and testing for a potential, further automatization. I acted in favor for the former because the later stages of automatization have been studied repeatedly in the literature. At the same time, the data here did not rule out any further increase along the trajectory; in that case, an experiment with more practice and statistical procedure including the cubic term of time (or testing block) could address this possibility.

Another important finding of this study was that CV was relatively stable in the learning of orthographic forms in natural reading. Although CV was extended by Solovyeva and Dekeyser (2018) beyond a measure of automaticity to an index of forming new knowledge representations, the data here suggested cautions. In particular, CV measures based on first fixations and gaze durations on target words that appeared multiple times in the reading text did not change significantly across the first 12 occurrences. While Elgort and colleagues (2017) concluded in their study that learners developed familiarity of the orthographic forms quickly, the authors also found that the learners were unable to retrieve

meaning when subsequently encountering the same words in semantically neutral contexts. Taken together, the stability of CV might be attributed to insufficient depth of processing despite multiple exposures. It is important to note that learners were told to read and understand the texts. As such, cognitive resources may be diverted to particular novel words only when deeper processing assists comprehension. The data appeared to suggest that the gains in orthographic familiarity alone were not sufficient to cause an observable change in the mental lexicon that would have an impact on processing stability. In order words, CV as an index of establishing new representations is not universally applicable, especially in the present case of the learning of orthographic forms through natural reading. Further research needs to be conducted to identify the exact limits of what CV could index.

Conclusion

Inspired by Solovyeva and DeKeyser (2018), this study investigated the signature of CV in the early stages of vocabulary learning in both intentional and incidental paradigms. With a focus on measuring the process of learning, I captured an inverted-U shaped trajectory in CV in the intentional paradigm. The relative stability of CV in the eye-tracking data highlighted CV as an index of establishing new knowledge representation might not be universally applicable in the learning of orthographic word forms under incidental conditions.

References

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, *29*(2), 175-193.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bartolotti, J., & Marian, V. (2014, January). Wordlikeness and novel word learning. In *Proceedings of the Cognitive Science Society*, *36*(36). Retrieved from www.escholarship.org/content/qt2j8652jh/qt2j8652jh.pdf.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J*ournal of Statistical Software*, *67*(1), 1 - 48. doi:http://dx.doi.org/10.18637/jss.v067.i01

Brown, H., & Gaskell, M. G. (2014). The time-course of talker-specificity and lexical competition effects during word learning. *Language, Cognition and Neuroscience*, *29*(9), 1163-1179.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A, 497-505.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 94-112). New York: Routledge.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*(2), 367-413.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2017). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263117000109.

Gass, S. (1999). Discussion: Incidental vocabulary learning. Studies in Second Language Acquisition,*21*(2), 319-333.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., …, & YOON, H. J. (2017). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728917000219.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483-517.

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.) *Implicit and explicit learning of languages*, 325-348, Amsterdam: John Benjamins.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us?. *Applied Psycholinguistics*, *30*(4), 555-582.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149-174.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York: Routledge.

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, *36*(5), 1247-1282.

Mohamed, A. A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263117000092.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97-130.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using EZ Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological review*, *119*(1), 155.

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.3.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime: User's guide. Psychology Software Incorporated.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369-385.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*(1), 53-67.

Segalowitz, N., Watson, V., & Segalowitz, S. (1995). Vocabulary skill: Single-case assessment of automaticity of word recognition in a timed lexical decision task. *Second Language Research*, *11*(2), 121-136.

Solovyeva, K., & DeKeyser, R. (2018). Response time variability signatures of novel work learning. *Studies in Second Language Acquisition*, *40*(1), 225-239.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256.

Wilcox, R (2005). *Introduction to robust estimation and hypothesis testing*. Elsevier: Amsterdam.

Appendix A
List of stimuli used in the intentional word learning experiment

Item No.	SwWord	EnWord	CNC	IMG	SUBTLwf	EnNLET	SwNLET
1	mvulana	BOY	609	618	530	3	7
2	msichana	GIRL	607	634	557	4	8
3	farasi	HORSE	613	624	93	5	6
4	ua	ROSE	608	623	53	4	2
5	mbwa	DOG	610	636	193	3	4
6	mti	TREE	604	622	65	4	3
7	nyoka	SNAKE	621	627	22	5	5
8	kuku	CHICKEN	614	619	62	7	4
9	gari	CAR	622	638	483	3	4
10	kitanda	BED	635	635	187	3	7
11	dirisha	WINDOW	609	602	86	6	7
12	ufunguo	KEY	612	618	87	3	7
13	jiwe	ROCK	600	612	86	4	4
14	kiti	CHAIR	606	610	49	5	4
15	barafu	ICE	621	635	80	3	6
16	mkate	BREAD	622	619	28	5	5

Notes: SwWord = Word in Swahili; EnWord = Word in English; CNC = Concreteness rating; IMG = Imaginability rating; SUBLwf = English written frequency (SUBTLEXUS); EnNLET = English word length; SwNLET = Swahili word length

Confucius's Analects: An advanced reader of Chinese language and culture

Rachel Lin
TESOL, Michigan State University
wilso657@msu.edu

Introduction

Zu-Yan Chen wrote the textbook *Confucius's Analects: An advanced reader of Chinese language and culture* (讀論語學中文) for learners with at least three years of university Chinese language courses (Chen, 2010, p. xiii). The main focus of study is excerpts from the *Analects*, a collection of Confucius' aphorisms and teachings, written in Classical Chinese, and widely studied by the Chinese population for around two thousand years. These passages are accompanied with translations in modern Chinese and short texts analyzing the underlying ideas. Because this textbook is only in Chinese, with the exception of the preface, some vocabulary definitions or section titles, and the content does not include grammar explanations, the author assumes that the readers are already generally competent in basic Chinese vocabulary and grammar. Based on the fact that the textbook is mostly written in Chinese and on the reviewer's personal experience in learning Chinese as a second language, it does indeed appear to be suitable for Chinese language learners with at least three years of consistent Chinese study. However, it may not be necessary for the textbook users to have had university language classes as long as they have comparable language skills.

Objectives

Although the author does not directly state objectives, based on the main content and exercises in each lesson, one can assume an objective to be: students will be able to identify main ideas in a given text. This objective appears to be met because each lesson presents three texts with vocabulary lists as scaffolding, and tests this objective with exercises that require students to either correct content mistakes in Chinese sentences, answer questions, or rephrase the main ideas. Based on the vocabulary lists and other exercises in each chapter, other possible objectives may be that students will be able to apply vocabulary and Chinese idioms in writing, compose short texts about familiar or unfamiliar topics, analyze opposing sides in a debate, and find

information online about unfamiliar topics. The textbook user has to apply new vocabulary knowledge in all ten exercises per lesson, especially in the first and seventh ones wherein the correct vocabulary word must be chosen out of options to complete sentences. Idioms must also be applied in the sixth exercise in each lesson, where sentences must be rewritten using idioms. Finally, the remaining, possible objectives are met by the tenth exercise in each lesson, which is either a composition on a familiar topic, debate preparation, or topical research assignment.

Content of the Textbook

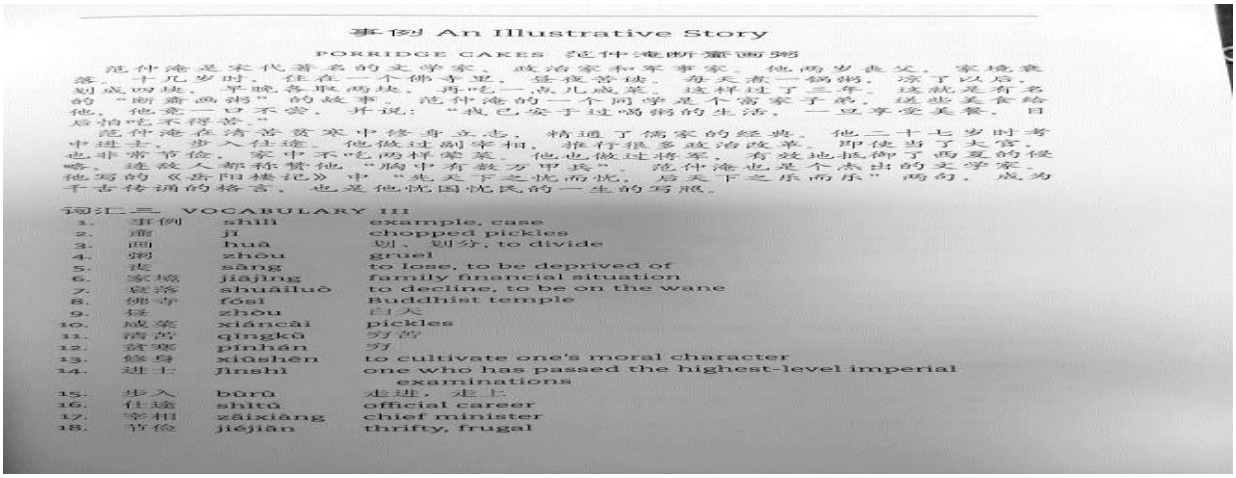
This textbook begins with a preface, a user's guide, and a short passage introducing Confucius and the literary work the *Analects*. The main content of the book is divided into four parts or themes of Confucius' teachings: Knowledge, Morality, Wisdom, and Government. Each of these parts are subdivided into five lessons or chapters with sub-themes (see Table 1 for a brief overview of lesson content). Each lesson begins with concise excerpts from the *Analects*, followed by a translation in modern Chinese (see Figure 1).

Figure 1. Example of presentation of one excerpt from the *Analects* in textbook (p. 4).

After these passages there is a short text interpreting the ideas in those excerpts, and a common story in Chinese culture illustrating those ideas (See Figures 2 and 3, respectively).

Figure 2. Excerpt of text interpreting passages from *Analects* (p. 5).

Figure 3. Example of illustrative story that supports ideas in *Analects* excerpts (p. 7).



事例 An Illustrative Story

PORRIDGE CAKES 范仲淹断齑画粥

范仲淹是宋代著名的文学家、政治家和军事家。他两岁丧父，家境衰落。十几岁时，住在一个佛寺里，昼夜苦读。每天煮一锅粥，凉了以后，划成四块，早晚各取两块，再吃一点儿咸菜。这样过了三年。这就是有名的"断齑画粥"的故事。范仲淹的一个同学是个富家子弟，送些美食给他，他竟一口不尝，并说："我已安于过喝粥的生活，一旦享受美餐，日后抽吃不得苦。"

范仲淹在清苦贫寒中终身立志，精通了儒家的经典。他二十七岁时考中进士，步入仕途。他做过副宰相，推行很多政治改革，即使当了大官，也非常节俭，家中不吃两样荤菜。他也做过将军，有效地抵御了西夏的侵略，连敌人都称赞他"胸中有数万甲兵"。范仲淹也是个杰出的文学家。他写的《岳阳楼记》中"先天下之忧而忧，后天下之乐而乐"两句，成为千古传诵的格言，也是他忧国忧民的一生的写照。

词汇三 VOCABULARY III

1.	事例	shìlì	example, case
2.	齑	jī	chopped pickles
3.	画	huà	划，划分, to divide
4.	粥	zhōu	gruel
5.	丧	sāng	to lose, to be deprived of
6.	家境	jiājìng	family financial situation
7.	衰落	shuāiluò	to decline, to be on the wane
8.	佛寺	fósì	Buddhist temple
9.	昼夜	zhòuyè	白天
10.	咸菜	xiáncài	pickles
11.	清苦	qīngkǔ	穷苦
12.	贫寒	pínhán	穷
13.	修身	xiūshēn	to cultivate one's moral character
14.	进士	jìnshì	one who has passed the highest-level imperial examinations
15.	步入	bùrù	走进，走上
16.	仕途	shìtú	official career
17.	宰相	zǎixiàng	chief minister
18.	节俭	jiéjiǎn	thrifty, frugal

All of these texts are accompanied with separate lists of vocabulary, categorized by the author as vocabulary words, grammatical function words, and proper names of people or places mentioned in the text (See Figure 4).

Figure 4. Excerpts of vocabulary lists presentation.

After the vocabulary lists there are two more sections of input for the textbook user: an idioms list and one last vocabulary list. All of the idioms used in the lesson's texts are presented again with sentence examples (see Figure 5), whereas the final vocabulary list focuses on single Chinese characters featured in the texts and lists various multi-character words that can be derived from them (see Figure 6).

Figure 5. Excerpt from idiom list (p. 8).

Figure 6. Excerpt from Vocabulary Expansion list (p. 9).

Following the textual input each lesson contains ten exercises (See Appendix A). First there are three to eight multiple choice questions, wherein vocabulary must be used to complete sentences taken directly from the texts. Then the reader must correct four to eight sentences from the texts that were altered to be inaccurate in content, and answer around ten short questions about the lesson content. Then there are about five open-ended discussion questions related to the topic of the lesson. The fifth exercise involves the student retelling the illustrative story from the lesson in his or her own words. Following this is an activity where around five sentences are rewritten using idioms from the chapter, but it lacks further directions as to what students should do. Next is another multiple choice vocabulary exercise to complete about five sentences. Each lesson also provides two translation activities between either Old Chinese and Modern Chinese or from Modern Chinese to English and vice versa. Finally, the lessons conclude with one of the following four activities: 1) a composition assignment on a familiar topic, 2) preparing a debate, 3) a composition assignment requiring online research, or 4) writing a story based on online research. See Table 1 for an outline of the content in each lesson.

Table 1.
Content of each lesson

Textual Input	Authentic passages from the *Analects* Translation into Modern Chinese Vocabulary list, function words list, and proper names list Text interpreting the passages Vocabulary list Illustrative Story Vocabulary list, function words list, and proper names list Idioms List Vocabulary Expansion List
Exercises	Multiple Choice Correcting Mistakes Answering Questions Discussion Questions Retelling the Story Rewriting Sentences Using Idioms Word Choice Either 1) translating a passage from Classical Chinese to Modern Chinese or 2) translating a passage from Modern Chinese to English
	Translating a passage from English to Chinese
	One of the following: 1. Composition 2. Debating 3. Topical Research 4. Storytelling

The appendices contain English translations of all of the passages from the *Analects* and compile indices of the vocabulary, proper names, function words, and idioms. Each of these lists provide the characters, Pinyin (alphabetic transcription), definition in either Chinese or English, and where to find it in the lessons.

Theory Behind the Textbook

Chen does not outline his teaching approach or philosophy for this book. However, on further examination there is a strong emphasis on authentic materials, multiple recurrences of vocabulary or idioms in context, and translation. Although many of the texts are not authentic by the definition of being "produced by a real speaker or writer for a real audience and designed to convey a real message of some sort" (Morrow, 1977, as cited in Gilmore, 2007, p. 98), the main focus is on direct excerpts from the *Analects*, which is attributed to Confucius and written for the Chinese population at that time. Therefore, these passages from the *Analects* can be said to be authentic according to Morrow's (1977) definition. Authentic materials are important because textbook language does not necessarily reflect real-world language and may not serve as adequate models for the learner. Moreover, authenticity contributes to learner motivation (Gilmore, 2007). The Chinese used in the *Analects* certainly does not reflect current real-world usage of the language. However, one can argue that learning to read and discuss Confucius' teachings may be motivating for some learners. Another benefit to using these authentic passages is that it connects the learners to the outside world (Thomas, 2014), as the *Analects* are still studied by Chinese peers today. Therefore, another motivating factor may be for Chinese language learners to be able to converse with native speakers on this subject. However, Thomas (2014) also argued that authentic materials should be locally relevant, which may not be the case for all learners using this textbook as everyone has different needs.

New vocabulary is reflected multiple times in each lesson in the textbook through the texts, subsequent vocabulary lists, and exercises. Schmitt (2000) believes that seeing a word in different contexts contributes to a learner's quality of knowledge, and multiple exposures help solidify this knowledge. Sökmen (1997, as cited in Schmitt, 2000) also advocates for multiple exposures to the same vocabulary as well as independent learning strategies involving word lists. Recurrent exposure and highlighting words in vocabulary lists also contribute to noticing and thoughtfully processing for meaning (Nation, 2001). Therefore, the multiple reoccurrences of the same vocabulary may benefit this textbook user. In addition, many exercises in each lesson require the learners to retrieve the meaning of these vocabulary words like in the multiple choice questions, and others elicit creative use like with the composition (Nation, 2001). This textbook then allows for the Chinese learner to process vocabulary in each of Nation's (2001) three stages of noticing, retrieving, and creative use, to which he attributes successful vocabulary acquisition.

Finally, each lesson includes two translation activities, one between Classical Chinese and Modern Chinese, and one between Modern Chinese and English. This is reminiscent of a typical grammar-translation exercise, where the learner is not using the language for communication (Celce-Murcia, 2014). Although translation does not appear to be the main focus

of this textbook, it may be considered part of the author's language learning philosophy as he includes two of these activities per lesson.

Critique

My impression of this textbook is that its authentic material is highly engaging. Although studying the *Analects* may not be useful for all Chinese language learners, this textbook can still fulfill the needs of particular audiences. For example, a Chinese language and culture class could benefit from this textbook, as the students will be using Modern Chinese to discuss ideas from the past. Another example would be graduate students or scholars. From an academic point of view it's still worthwhile to learn about the *Analects* in both its original language and Modern Chinese, especially for someone who wants to understand the Chinese heritage and the influences of Confucianism throughout history. Culture and history are significant parts of language and language learning. Although one may argue that some Confucian values may not apply anymore in modern China, young students still study the *Analects* in their education system. Confucianism is also still influential throughout the world in other Chinese speaking countries. Moreover, it is valuable to trace how this thought system changed throughout history. So despite any arguments to the contrary, this reviewer maintains that the *Analects* is a valuable source of material for many Chinese language learners, depending on their individual needs.

After successfully using this textbook, one might expect to be able to have more common knowledge with Chinese native speakers and gain insight into their culture. The reading materials appear to be of manageable length and difficulty for someone with at least three years of Chinese study, and the exercises, although perhaps not communicative, do require students to use the vocabulary and language presented in the texts. My only qualms with this textbook are the lack of directions for the idiom and vocabulary expansion lists, and the translation exercises. As there are no directions for how to use the two lists I can only guess the author's intention, and inexperienced teachers may simply opt for presenting these lists to the class one by one in an unengaging manner. One may in fact need supplementary materials to make these lists suitable for a classroom lesson and fit the needs of the specific students. I disliked the translation exercises because they are reminiscent of the Grammar Translation Method and do not reflect natural language use unless you are employed as a translator.

Although this textbook is not perfect, I would certainly use it to study Chinese by myself and recommend it for a Chinese class depending on the needs and levels of the students. Not every language learner is interested or has use for ancient literature, but, in the case of Chinese, the *Analects* and Confucianism is deeply entrenched in the culture, so having an insight into this thought system could bridge cultures. That being said, with excerpts in Classical Chinese, it is inevitable for some readers to become frustrated with the mismatches between that and Modern Chinese. That is why it is important that the author included Modern Chinese translations next to the original *Analects* passages. To make this textbook more suitable for a communicative classroom it would certainly need to be supplemented, but it does provide a good base from which to organize a sound Chinese lesson for advanced learners.

References

Celce-Murcia, M. (2014). An overview of language teaching methods and approaches. In Celce-Murcia, M., In Brinton, D., & In Snow, M. A. (Eds.). *Teaching English as a second or foreign language*. Boston: National Geographic Learning.

Chen, Z. (2010). *Confucius's Analects: An advanced reader of Chinese language and culture* (讀論語學中文). Washington, D.C.: Georgetown University Press.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, *40*, 97-118.

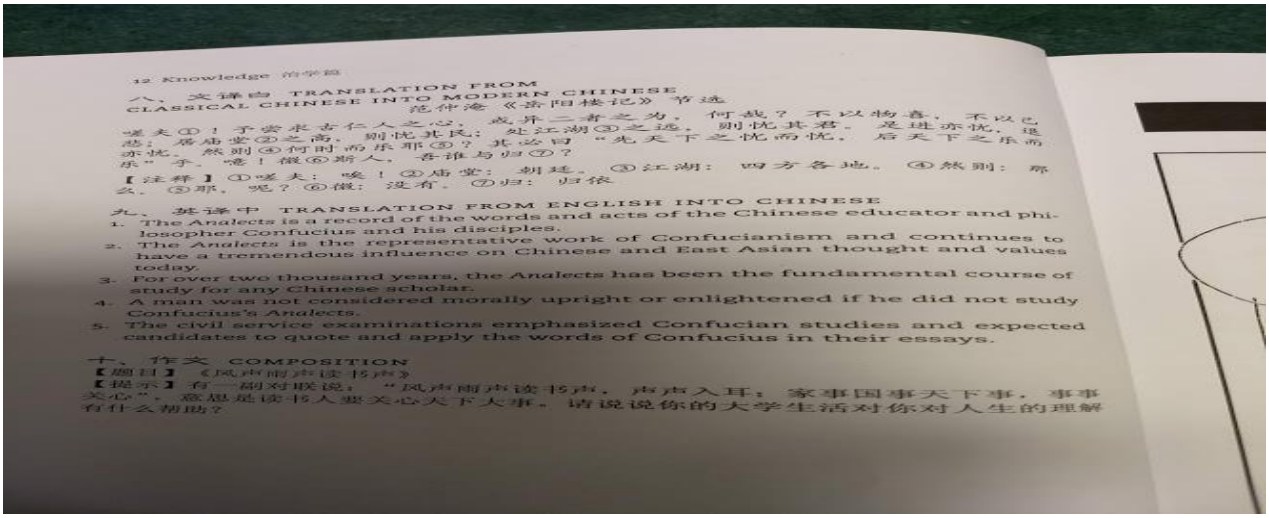
Nation, I.S.P. (2001). *Learning vocabulary in another language*. UK: Cambridge University Press. (Chapter 3).

Schmitt, N. (2000). *Vocabulary in language teaching*. NY: Cambridge, UP. (Chapter 8).

Thomas C. (2014). Meeting EFL learners halfway by using locally relevant authentic materials. *English Teaching Forum*, *52*(3), 14-23.

Appendix A

Examples of Exercises from Lesson One (limited to one question per exercise)



12 Knowledge 尚学篇

八、文译白 TRANSLATION FROM CLASSICAL CHINESE INTO MODERN CHINESE

范仲淹《岳阳楼记》节选

嗟夫①！予尝求古仁人之心，或异二者之为，何哉？不以物喜，不以己悲；居庙堂②之高，则忧其民；处江湖③之远，则忧其君。是进亦忧，退亦忧。然则④何时而乐耶⑤？其必曰"先天下之忧而忧，后天下之乐而乐"乎。噫！微⑥斯人，吾谁与归⑦？

【注释】①嗟夫：唉！②庙堂：朝廷。③江湖：四方各地。④然则：那么。⑤耶，兜？⑥微：没有。⑦归：归依

九、英译中 TRANSLATION FROM ENGLISH INTO CHINESE

1. The *Analects* is a record of the words and acts of the Chinese educator and philosopher Confucius and his disciples.
2. The *Analects* is the representative work of Confucianism and continues to have a tremendous influence on Chinese and East Asian thought and values today.
3. For over two thousand years, the *Analects* has been the fundamental course of study for any Chinese scholar.
4. A man was not considered morally upright or enlightened if he did not study Confucius's *Analects*.
5. The civil service examinations emphasized Confucian studies and expected candidates to quote and apply the words of Confucius in their essays.

十、作文 COMPOSITION

【题目】《风声雨声读书声》

【提示】有一副对联说："风声雨声读书声，声声入耳；家事国事天下事，事事关心"，意思是读书人要关心天下大事。请说说你的大学生活对你对人生的理解有什么帮助？