Studies in Second Language Acquisition **42** (2020), 327–357
doi:10.1017/S0272263119000603

Research Article

PROCESSING VARIABILITY IN INTENTIONAL AND INCIDENTAL WORD LEARNING

AN EXTENSION OF SOLOVYEVA AND DEKEYSER (2018)

Bronson Hui*

Michigan State University

Abstract

I investigated the trajectory of processing variability, as measured by coefficient of variation (CV), using an intentional word learning experiment and reanalyzing published eye-tracking data of an incidental word learning study (Elgort et al., 2018). In the word learning experiment, native English speakers ($N = 35$) studied Swahili-English word pairs ($k = 16$) before performing 10 blocks of animacy judgment tasks. Results replicated the initial CV increase reported in Solovyeva and DeKeyser (2018) and, importantly, captured a roughly inverted U-shaped development in CV. In the reanalysis of eye-tracking data, I computed CVs based on reading times on the target and control words. Results did not reveal a similar inverted U-shaped development over time but suggested more stable processing of the high-frequency control words. Taken together, these results uncovered a fuller trajectory in CV development, differences in processing demands for different aspects of word knowledge, and the potential use of CV with eye-tracking research.

Second language (L2) 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 coefficient of variation (CV) to the field of second language acquisition (SLA) as a measure of processing automaticity. A CV value reveals the variability of an individual's reaction times (RTs)

This piece of research is supported by the Second Language Studies PhD Program at Michigan State University. I would like to thank the two anonymous reviewers for their insightful feedback that has helped strengthened the paper. I would also like to thank Drs. Aline Godfroid, Shawn Loewen, and Patti Spinner for their constructive comments on earlier drafts of this manuscript. I must also thank Lizz Huntley and my peers on LLT 864 Second Language Psycholinguistics for their ideas. Audience members in my talk at SLRF 2018 in Montreal and members in our Daily Writing Slacker group offered very useful comments as well. All errors, however, remain my own. Finally, I am very grateful for the sharing of published data by Dr. Irina Elgort and colleagues.

The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials are available at https://osf.io/2ta7m/.

*Correspondence concerning this study should be addressed to Bronson Hui, Second Language Studies Program, Michigan State University, East Lansing, Michigan. E-mail: huibrons@msu.edu

Copyright © Cambridge University Press 2019

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

obtained from experimental, linguistic tasks, such that a small CV value indexes less variable (more stable) processing. Over the years, one received view has 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 has, however, not always been supported in research (e.g., Hulstijn et al., 2009; cf., Lim & Godfroid, 2015). At least some of the discrepancy may be attributed to the differences in developmental stages participants are in between studies. Because automatization theory poses that acquisition of declarative and procedural knowledge precedes automatization (see DeKeyser, 2015 for a review), lower-proficiency learners who are still accumulating declarative and procedural knowledge may not show signs of automatization: a decreasing CV value (Hulstijn et al., 2009; Lim & Godfroid, 2015).

Further, CV research has also been extended to signaling the addition of new linguistic representations in the lexicon (Solovyeva & DeKeyser, 2018). The authors found an increase in CV values (less stable processing) measured across all/multiple items as learners first engaged in learning novel words. According to the authors, when more efficient processing routes as a result of automatization can lead to improvements in processing stability (a decrease in CV values), one can attribute the increase in CV values found in their study to the addition of new representations. Their findings, therefore, cast doubt upon the accepted prediction of a decreasing CV along the entire developmental trajectory because one can have a higher CV value but may be further along the trajectory. If this finding can be replicated consistently, it challenges the fundamental assumption of a linear relationship between development and CV values. This challenge then has implications for researchers using CV to index learning (McManus & Marsden, 2019; Pili-Moss, Brill-Schuetz, Faretta-Stutenberg, & Morgan-Short, 2019) and lexical proficiency (Elgort, 2011; Elgort & Warren, 2014). For this reason, it is time researchers captured learners' development over a fuller learning trajectory in one single study.

Another limitation of this research is perhaps its reliance on behavioral tasks (e.g., lexical decision tasks) and its focus on intentional word learning. It has also been argued that vocabulary is often learned incidentally; that is, as a by-product of other cognitive processes involving comprehension (Gass, 1999). Therefore, it is also important to understand how processing variability changes in the incidental learning paradigm where results indicating learning of orthographic word forms has been the most robust (e.g., Elgort et al., 2018; Godfroid et al., 2018).

Inspired by Solovyeva and DeKeyser (2018), in the present study, I attempted to capture the potential change in CV during a fuller trajectory of word learning in two paradigms (intentional and incidental word learning), involving different aspects of word knowledge (orthographic forms and form-meaning mappings [Webb, 2005]). I report (a) an intentional word learning experiment, looking at intentional learning of form-meaning mappings and (b) reanalyses 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 the CV as a measure of processing stability, a key characteristic in processing automaticity (Segalowitz, 2010) in

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

the field of SLA. Given the important role of processing automaticity in skill learning theory (see DeKeyser, 2015 for an overview), it is critical that this measure be valid and reliable to verify (or falsify) theoretical claims about learners' processing trajectory (e.g., learners become more automatic in processing as proficiency develops). Validating and understanding the properties of CV have become more urgent since researchers have started using CV to index lexical proficiency and measure learning gains in both vocabulary and grammar research (Elgort, 2011; Elgort & Warren, 2014; McManus & Marsden, 2019; Pili-Moss et al., 2019). CV is calculated 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 for his or her processing speed. A small CV value indexes more stable processing evidenced by a smaller variability in the learner's RTs. Due to the role of both processing 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 RT-CV correlation.

RELATIONSHIP BETWEEN COEFFICIENT OF VARIATION AND DEVELOPMENT

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 in language acquisition. One issue here is concerned with whether the learner must have strong linguistic knowledge as a prerequisite for automatization. One way to address this issue is to include experimental materials of a spectrum of difficulty levels. Assuming comparable proficiency within a participant sample, researchers can expect to observe automatization in a subset of items, which, in theory, their participants are familiar with but have not yet automatized. Segalowitz, Watson, and Segalowitz (1995), for instance, used words across four frequency bands to sample processing of words at different developmental stages in the participants' lexicon. The authors reported data from a Turkish learner of English (aged 26) who performed a series of lexical decision tasks on different sets of words that were comparable in frequency 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 posttraining lexical decisions showed a decrease in CV values and a RT-CV correlation only for low-frequency words. The evidence provided by these studies highlighted the importance of sampling experimental materials across a spectrum of difficulty levels such that the sampling can cover words at various stages of development in the mental lexicon (Jiang, 2000, 2002). In the studies by Segalowitz et al. and Akamatsu, one interpretation could be that the learners had already acquired the declarative and procedural knowledge of those low-frequency words, hence, such knowledge was available for automatization; in contrast, high-frequency words might have already been automatized and hence further decrease in CV should not be expected.

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

However, the issue of whether declarative and procedural knowledge is a prerequisite for automatization could be addressed from the perspective of learners' proficiency. Given a set of experimental materials (e.g., target words), more proficient learners should have already automatized the lexical knowledge attested; in contrast, less proficient learners could still be acquiring declarative and procedural knowledge of the same words (hence automatization should not be expected). For example, Segalowitz and Segalowitz (1993), in their seminal paper, reported different results for French learners of English across skill levels. In their 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 participants correlated with their CVs only at the last presentation. The authors did not report any change in CV values across time. Also, Segalowitz, Segalowitz, and Wood (1998) reported RT-CV correlations for both initially fast and slow participants, but a decrease in CV values was only found in the initially fast group.

One pattern that seemed to have emerged from these earlier studies was that automatization was more often observed in more skillful learners. This pattern does not seem to change when comparing across studies that sampled participants from different proficiency levels and language learning contexts. In particular, 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, a 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 and advanced university learners of English in America, as well as English native speakers. Though some correlations were not significant, the authors concluded that their findings were largely consistent with automatization. According to Lim and Godfroid, the high school 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).

Researchers, however, should exercise caution when interpreting this comment because of the potential circularity involved. When automatization is found, one could argue that the learners have sufficient linguistic knowledge to be automatized. If evidence for automatization is not found, the results could also be attributed to their relatively low proficiency levels. On this account, then, it could be challenging to falsify the theory of automatization. To break the circularity, there is 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 knowledge stages in the investigation is crucial because, first, they are the basis upon which later automatization develops, according to skill learning theory. However, previous studies focused only upon lexical items already known by the learners because the use of CV was originally proposed to measure automatization, a much later stage of development. Second, a predicted downward CV is meaningful only when all other directions of change (an increase or no change) have been attested

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

(Solovyeva & DeKeyser, 2018). Only from that point on could this prediction be falsifiable in the operationalization of measuring automaticity.

PROCESSING VARIABILITY IN THE EARLIEST STAGES OF INTENTIONAL WORD LEARNING

In a study of deliberate novel word learning, Elgort (2011) investigated the early stage of word learning using CV as an outcome measure of automaticity. Advanced L2-English learners were presented with 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 possessed a higher degree of automaticity when processing these target words than low-frequency nontarget words and other nonwords (to which they had had no exposure). Although the author was interested in deliberate learning of novel words, the main testing did not take place 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 the processing of products of learning, not processing *during* the learning process (Godfroid, 2019; Leow, 2015).

Another study that investigated specifically the initial word learning process in relation to CV was Solovyeva and DeKeyser (2018). One important contribution of this study was that the authors extended the use of CV to capturing how processing patterns changed when new representations were established in the lexicon. 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 a larger variability in processing as observed by an increase in CV values. The study involved 73 native English-speaking university students. Word learning was operationalized in two ways: first, learning new form-meaning pairings of 40 Swahili words, and second, acquiring the orthographic word forms 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. In the pretraining priming task, the learner was first presented with 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 in a training block. But this pretraining testing offered a baseline measure of processing the Swahili words with no knowledge of English-Swahili translation equivalences. In the posttraining priming task, half of the prime-target pairs were changed, which created a (semantically) mismatched condition. The unchanged, original prime-target pairs formed the (semantically) matched condition. These two conditions provided the basis for the semantic priming task. If participants responded faster to the English translation equivalent following its corresponding Swahili prime in the matched condition, this facilitation in processing (priming) 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 posttraining testing sessions. These changes meant that, as a result of training, the learner processed the words faster, but in

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

a less stable manner (as demonstrated in the greater variability in RTs), confirming the authors' initial hypothesis. This finding also challenged the accepted prediction of a decreasing CV along with development. Importantly, it also implied that an increase in CV could potentially accompany the establishment of new lexical representations in the lexicon. Although an increase in CV cannot indicate representations being established at the level of individual words, the increase signals a higher within-learner processing variability because a learner is likely to process known and newly learned words differently. For this reason, addition of new representations to the lexicon seems to be accompanied by an increase in CV values. However, caution should be exercised in interpreting the findings because processing variability of the L2 items (the target learning items) was not directly measured.[1] Specifically, the RT data were obtained from decisions to the L1 targets (although primed by newly learned L2 items). One way to address such a limitation is then to measure variability through judgments made directly to the L2 words.

Solovyeva and DeKeyser (2018) embedded the second aspect of word learning (i.e., the implicit learning of orthographic word forms) in the same experiment, but in the nonword trials (where a NO response was expected in the lexical decision task to make it a genuine task for participants). Because participants encountered these nonwords a total of four times in the tasks (once in each condition, pre- and posttraining), the authors used the RT data to nonwords (hence the CV data) to investigate potential learning of orthographic word forms of these nonwords as a result of multiple exposures. Although results here aligned with the previous analysis, a couple of points are worth noting. Because the goal of this part of the study was the learning of orthographic word forms (i.e., devoid of meaning), the learners did not have the opportunity to associate these nonwords with meaning, which is probably not typical for language learners who often encounter new lexical items in context. Second, as also acknowledged by the authors, there was a conflict between learners becoming more familiar with the nonwords and having to treat these items as nonwords (a NO response was required). This tension could have influenced whether this growing familiarity could be considered as word learning and could have been attributed as part of the cause of the more variable responses observed. Another source of variability could have come from the processing of the primes. Indeed, participants responded to these nonwords that followed the primes. When these primes became more familiar to different degrees, as also pointed out by a reviewer of the present study, they could have exerted influence on the processing of these nonwords to various degrees as well. To mitigate some of these limitations, Solovyeva and DeKeyser 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 to the lexicon at both levels of orthographic word forms and form-meaning mappings during the initial stages.

The Solovyeva and DeKeyser study sheds important light on the initial stages of the learning spectrum and is also the first study to report an increase in CV in lexical processing. At the same time, it is also important to note that, similar to the tests in Elgort (2011), the semantic priming task in this study was only a test of processing of learning products and not that during the learning process. The distinction between process and products of learning is crucial, especially in this strand of work looking at automatization and addition of new representations (Godfroid, 2019). If processing stability deteriorates

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

(an increase in CV) as a result of addition of new representations to the lexicon, researchers should be more interested in capturing this change as new representations are *being* established, less so after they are already established. Again, there is a need for researchers to capture the impact of establishing new representations on processing stability, and its ramification after that, covering at least a fuller development trajectory, from no knowledge, through acquisition of declarative and procedural knowledge, to automatization. With the available evidence in the research base so far (e.g., Lim & Godfroid, 2015; Solovyeva & DeKeyser, 2018), one logical prediction is an initial increase in CV, revealing addition of new representations to the lexicon, followed by automatization of such knowledge as demonstrated by a decrease in CV in the later part of the trajectory. In other words, this prediction means a potential inverted U-shaped development in CV values across time.

Apart from the need to capture a fuller developmental trajectory, another limitation of this strand of work has been its reliance on behavioral tasks such as lexical decision or animacy judgment tasks, involving motor responses of the participants. Evidenced by the preceding review of Solovyeva and DeKeyser's (2018) study, using behavioral tasks to examine different word learning paradigms (intentional and incidental word learning) involving different aspects of word knowledge (e.g., orthographic forms and form-meaning mappings) has been less straightforward. Indeed, much of the work done in this area has focused on intentional word learning of form-meaning mappings (e.g., Elgort, 2011). At the same time, incidental learning of vocabulary—that is, as a by-product of other cognitive processes involving comprehension (Gass, 1999)—is a key source of word learning. For this reason, understanding how processing variability changes also in the incidental learning paradigm can shed light on the exact properties of CV in word learning.

INVESTIGATING PROCESSING VARIABILITY IN INCIDENTAL WORD LEARNING

Examining learners' processing during incidental learning of vocabulary, researchers have recently adopted the eye-tracking methodology (e.g., Elgort et al., 2018; Godfroid et al., 2018; Godfroid, Boers, & Housen, 2013; Mohamed, 2017; Pellicer-Sánchez, 2016). Eye tracking is the recording of 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 ask participants to read an experimental text in which target words are embedded. Without prior notice, the learners then take part in vocabulary tests that assess their word learning during reading. This learning is then associated with their reading behavior as captured by eye tracking. Godfroid et al. (2018), for example, found that total reading times predicted explicit posttest scores of vocabulary meaning. Also, the researchers found a nonlinear decrease in reading times when learners encountered unfamiliar lexical items repeatedly. This decrease in reading times served as online evidence of learning of orthographic forms. Similarly, Elgort et al. (2018) reported learners' quick and reliable development of orthographic processing speed up in reading after approximately eight

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

exposures in reading. In particular, reading times on target words dropped quickly, indicating readers' gains in familiarity with the target words.

These studies, however, focused mostly on the association between reading times (processing speed) and learning gains. In skill learning theory, a simple speed up in processing does not always index automatization: that faster processing alone (as opposed to improved processing stability) does not reveal a restructuring of processing routes (Anderson, 1982, 1983; DeKeyser, 2007; Segalowitz, 2010). Therefore, investigating only the relationship between processing speed and learning gains sheds little light on any potential automatization of knowledge in the reading process. Importantly, if a change in processing variability could index addition of new representations in the lexicon, as Solovyeva and DeKeyser (2018) reported, researchers should also be able to observe such changes with eye-tracking data. Specifically, learners should display a change (reduction) in reading time variability over multiple exposures during natural reading.

This direction of work could potentially attest the extent to which an increase in CV is universally applicable across different word learning paradigms (intentional and incidental) and across different aspects of word knowledge (orthographic forms and form-meaning mappings). There are reasons researchers might expect different patterns in CV changes in these two paradigms. In the incidental paradigm, the primary task for the learner is to comprehend the text. Learners pay attention to unfamiliar words only when they consider them important in function of comprehension (Hulstijn, 2001; Montero-Perez, Peters, & Desmet, 2018). Other factors such as the overall context and more global understanding of the text could also play a role. This difference is then in sharp contrast with the intentional paradigm where learners are typically told to memorize the materials and hence are more likely to engage in deep processing.

Perhaps related to this, the more robust findings in the investigation of incidental word learning have been the learning of orthographic forms in meaningful contexts (Elgort et al., 2018; Godfroid et al., 2018). If this is the case, these eye-tracking data can test CV's ability to index addition of new orthographic forms in the mental lexicon as claimed by Solovyeva and DeKeyser (2018). In particular, unlike in Solovyeva and DeKeyser's study, learners in these reading studies processed vocabulary items in a meaningful context. Therefore, these data present a stronger case in which learners did have opportunities to associate the items with meaning (hence lexicalization might be more likely) even though the focus of investigation remains to be the learning of word form. In other words, this investigation could shed light on CV's signature in incidental word learning, involving often orthographic word forms. At the same time, processing during the learning process could be observed through tracking the pattern of change across participants' multiple exposures to the unfamiliar items. Eye tracking, then, appears to offer exciting opportunities in CV research.

THE PRESENT STUDY

In the light of the reviewed studies, the present study had three aims: first, to capture the change in processing variability when learners establish novel form-meaning mappings, from no knowledge, through the acquisition of declarative and procedural knowledge, to automatizing this new knowledge. Second, it aims at investigating if CV, derived from

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

eye-tracking data, could index the addition of representations of orthographic word forms in the incidental word learning paradigm. Finally, it explores the extent to which the use of CV as a processing variability measure can be extended to eye tracking. In achieving these aims, I formulated the following research questions to guide the present study:

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

(a) Intentional learning of form-meaning mappings of novel words?
(b) Incidental learning of orthographic word forms through natural reading?

RQ2: To what extent can CV derived from eye-movement measures index processing variability?

To address the research questions, I report, first, an intentional word learning experiment where participants deliberately learned the meaning of target words. This experiment will address RQ1 (a). Second, I report reanalyses of data from a published eye-tracking study investigating incidental learning of orthographic word forms through reading (Elgort et al., 2018), which will address RQ1 (b) and RQ2.

METHODOLOGY

INTENTIONAL WORD LEARNING EXPERIMENT

Participants

The initial sample was comprised of 40 native English-speaking undergraduate students at a midwestern university in the United States. Five participants were excluded due to unsatisfactory performance (see "Data Analysis"). The final sample consisted of 35 participants (mean age = 20, 74% female). None of them were enrolled in any formal foreign language classes at the time of participation or reported any knowledge of African languages, including Swahili. Participants received extra credit for a class for their participation. Ethical clearance had been obtained from the Institutional Review Board in accordance with the university's regulations governing research involving human participants.

Materials

Sixteen English-Swahili translation equivalents (e.g., *boy-mvulana*) were used. I selected highly frequent, concrete, and imageable nouns based on the SUBTLEX-US database for written frequency (Brysbaert & New, 2009). Concreteness and imageability ratings were based on the MRC Psycholinguistic Database (Coltheart, 1981). Mean values for the frequency of the word form for the English translation equivalents of the Swahili words was 166 (to provide a reference, frequency of *bed* was 187). Both means for the concreteness and imageability were larger than 600 (out of 700). These selection criteria[2] were meant to make it possible for the participants to learn the items from scratch and proceed to automatization in one single laboratory session because frequent, concrete, and imageable words are 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).

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 1. Characteristics of the experimental 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 imageability ratings were based on the MRC psycholinguistic database (Coltheart, 1981), and such ratings were not available for four words.

Procedure

After consent, participants proceeded to the experiment conducted in a quiet lab and run on SuperLab 5, a stimulus presentation software. Overall, there was one learning block and 10 test blocks.

In the learning block, participants were presented with all Swahili words with their English translation equivalent for eight seconds, one at a time. I told the participants to learn the words as well as they could. In this learning block, the experiment proceeded automatically. This learning block was immediately followed by 10 test blocks, in each of which the participant made animacy judgments 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 ("l" for living things and "a" for nonliving things) as quickly and accurately as possible. I chose an animacy judgment task to align the present study with previous automaticity research (e.g., Lim & Godfroid, 2015; Phillips, Segalowitz, O'Brien, & Yamasaki, 2004; Segalowitz & Frenkiel-Fishman, 2005). These researchers argued that an animacy judgment task "invokes relatively strong semantic processing" (Lim & Godfroid, 2015, p. 1255).[3]

Each trial started with a fixation cross (+) presented for 330 milliseconds (ms), followed by the target Swahili word presented in the lower case (e.g., *mvulana*). The target word remained on the screen until the participant responded. Visual feedback (*Incorrect* and the Swahili-English pair) was given for 3,000 ms only when an incorrect judgment was made. No feedback was given for correct trials. Participants were able to take a short break between the test blocks. Figure 1 summarizes the procedure for each trial in the test blocks. Each word pair was tested once in each block. The order of trials was pseudorandomized in each block by Superlab. The entire procedure took about 20 minutes.

Data Analysis

To align with skill learning theory and to ensure participants' engagement in the task, I arbitrarily decided an 80% accuracy at the seventh block and beyond to be the inclusion criterion for the final analysis. Although the animacy judgment test has its own limitations (see "Limitations"), I used this test as a learning proxy to index the level of declarative knowledge, a prerequisite for automatization, at the seventh block and

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

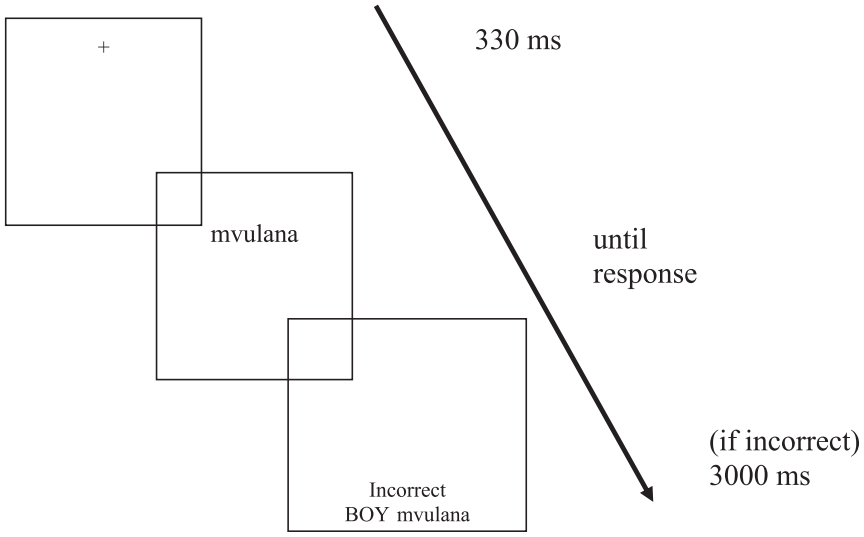


FIGURE 1. Order of presentation in a testing trial.

beyond. Descriptive statistics showed that, after exclusion, learners achieved a near-ceiling accuracy after Block 7 (at or above 95%). This level of accuracy demonstrated fulfilment of the prerequisite of declarative knowledge for automatization. As a result of this criterion, five participants were excluded. Following Hulstijn et al. (2009), Lim and Godfroid (2015), and McManus and Marsden (2019), only correct responses were analyzed. RT data trimming involved removing negative RTs (0%) and those slower than 3,000 ms (5.5%). At both participant and item levels, RTs were further trimmed by winsorizing 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 (version 1.8.4) (Revelle, 2018). The trim quantile was specified as .1, meaning that RTs faster than the 5th and slower than the 95th percentiles were replaced by values at the 5th and 95th percentiles, respectively. A CV value was computed for each participant at each test block, based on their RTs to the target words. I used the resulting dataset to build multilevel, mixed-effects regression models, using lme4 package (version 1.1-17) in R (Bates, Mächler, Bolker, & Walker, 2015). By including both fixed and random effects, mixed-effects modeling is able to account for inherent clustering between observations (Linck & Cunnings, 2015), or in technical terms, nested data structures. In the present case, each participant had 10 CV values in the dataset, one from each test block. As such, these 10 observations were nested within each individual participant (CVs clustered by participants). Because of this, these 10 observations may correlate (because they came from the same individual) in ways that were not observed between participants. Mixed-effects modeling, in this case, accounted for both within- and between-participant variation to generate more accurate estimates.

In the modeling, the *t* test used Satterthwaite's method to calculate the degrees of freedom, and hence the associated *p* values. The outcome variable was specified as CV.

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

For the predictors, I first recoded Block so that the nth Block was coded as n-1. For example, Block 1 was coded as 0, and Block 7 was coded as 6. Based on visual inspection of the observed data, orthogonal polynomials to the second order were included to capture any potential curvilinearity in the data (curve-like relationships, as opposed to linear, straight-line-like relationships). As such, I also transformed the recoded Block to an orthogonal, linear term (ot1) and to an orthogonal, quadratic term (ot2) (Mirman, 2014). This transformation centered both terms, and it made both terms independent of each other, avoiding issues associated with dependence between predictors when using natural polynomials (Mirman, 2014). As a result of the transformation, the intercept in the model reflects the average overall curve height (Mirman, 2014), not performance at the first block. These two terms (ot1 and ot2) were the primary predictors entered into the models.

In specifying the random effects structure, I engaged in model selection to identify the optimal random effects structure although there have been calls for use of the maximal random effects structure as justified by the research design (i.e., to include person-specific [random] intercepts and a person-specific [random] slope for all fixed effects) (Barr, Levy, Scheepers, & Tily, 2013; Linck & Cunnings, 2015). First, engaging in model selection can avoid issues associated with overfitting. Second, use of a maximal random effects structure also runs a risk of losing power and hence committing Type II errors (false negatives) (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Finally, from a more data-driven approach to modeling, individual differences in automatization found in previous studies (e.g., Pili-Moss et al., 2019; Suzuki, 2018) may not be observed in the data (i.e., random slopes for ot1 and ot2 may not be statistically significant).

In determining my model selection strategy, I used the forward approach (e.g., Field, Miles, & Field, 2012; Hox, Moerbeek, & Van de Schoot, 2018) to avoid early non-convergence issues. I first fitted a random intercept-only, null model with no fixed-effects factors. An intraclass correlation (ICC) was computed to indicate the proportion of variance due to the nested data structure (i.e., variance between level-2 units [participants]). Then, at each step of the model selection process, a fixed effect or its associated random component (the random slope for the fixed effect) was added to the model. The goal of each step was to test if the addition in question would result in a better model fit. Model selection was complete when an addition no longer improved model fit or when the model becomes so complex that it would not converge, resulting in a singular fit.

To compare model fit, I used the Akaike information criterion (AIC), a model fit index derived from deviance information while adjusted for model complexity (i.e., the number of parameters estimated). In general, smaller AIC values indicate a better model fit. In cases of negative AIC values, the more negative (larger absolute value) the AIC value the better the model fit. I did not use a χ^2 test of difference in deviance because an appropriate alpha (α) value was still being sorted out in the literature. Most references used the traditional alpha (α) value of .05, while Matuschek et al. (2017) advocated the use of .20 to avoid "an overly strong penalty on the model complexity" (p. 308). Because the whole model selection process involved both fixed and random effects, models were fitted using full information maximum likelihood to obtain deviance information for the computation of AIC values. For interpretation purposes, models fitted with restricted maximum likelihood (REML) were used because this method of estimation produces better estimates (Hox et al., 2018). To promote open science practices, I have made all raw data (output text files from

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

Superlab) and scripts of codes used in R for data preparation, processing, and analysis available on IRIS (https://www.iris-database.org/) (Marsden, Mackey, & Plonsky, 2015) and on Open Science Framework (https://osf.io/2ta7m/).

REANALYSES OF INCIDENTAL LEARNING EYE-TRACKING DATA

Description of the Study by Elgort et al. (2018)

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 ($k = 9$). This study was an ideal candidate for the present reanalysis because learners were under incidental vocabulary learning conditions, an important source of vocabulary learning (e.g., De Vos, Schriefers, Nivard, & Lemhöfer, 2018; Peters & Webb, 2018; Uchihara, Webb, & Yanagisawa, 2019). Also, Elgort and colleagues collected eye-movement data on both target and control words. Any potential change in eye-movement patterns for target words can be directly compared with the baseline obtained from control words. Finally, the multiple exposure to target words aligned well with the design of the intentional word learning experiment that centered the focus on the process of learning. Data analysis in the original study focused on the first eight occurrences of each target word. The most relevant finding was a decrease in time spent on the target words in measures such as first fixation and gaze duration, total reading times, and number of fixations, compared with the control words. These decreases were interpreted as evidence that learners had gained orthographic familiarity "relatively quickly and reliably" (Elgort et al., 2018, p. 342). Building upon these findings, I investigated if CV could index this gain in familiarity or the learning of orthographic forms in natural reading contexts.

Data Collection and Analysis

The authors generously made their eye-movement data available on Open Science Framework (https://osf.io/a9w7v/). 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., 2018). This decision was in accordance with the major purpose of the present study: to capture the fuller developmental trajectory. Inclusion of four more occurrences would better capture processing change further along the developmental trajectory of CV. As such, I removed four target words that occurred fewer than 12 times in the original study from the analysis. The analysis included only two early measures of eye movements: first fixation and gaze durations. First fixation durations refer to the duration the of the first fixation made in an interest area (Godfroid, 2020a). Gaze durations (or first pass time) refer to the sum of all fixation durations up until the eyes leave the interest area for the first time (Godfroid, 2020a). These early measures index such early processes in reading as word recognition or lexical access (Godfroid & Winke, 2015). In choosing these two early measures, I took into considerations how CV has been used in the literature (i.e., often computed from reaction time data obtained from

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

behavioral tasks such as lexical decision and animacy judgment tasks). These two early eye-movement measures indexing lexical processing could potentially mirror reaction time data, although there are some fundamental differences in these measures. Also, these early measures, compared with other late measures such as total reading times, have relatively less room for conscious involvement; hence better in line with the current investigations into the learning of orthographic forms. The data were not further trimmed or cleaned because Elgort et al. had already done so.

The statistical procedure included, first, computation of a CV value for each occurrence of the words in the text (e.g., the first time participants saw the words coded as 0 [intercept], the second time as 1). This CV value was based on their eye-movement measures across all 10 target and 9 control words. This CV value was specified as the outcome variable in the mixed models. Order of Occurrence in the text (e.g., first/second time the participant read the words), Word Type (i.e., target vs. control words), and their interaction term were included as the primary fixed effects of interests. Based on initial visual inspection of the observed data, I decided not to include any polynomial terms of the Order of Occurrence. As in the intentional word learning experiment, I engaged in model selection to determine the optimal random effects structure for each model to be interpreted. The model selection process followed that in the intentional word learning experiment. Again, codes used in R will be made available on IRIS (https://www.iris-database.org/) and on Open Science Framework (https://osf.io/2ta7m/).

RESULTS

INTENTIONAL WORD LEARNING EXPERIMENT

Table 2 shows the descriptive statistics for accuracy, RT, and CV across the 10 blocks of animacy tests. In general, accuracy increased, indicating that learning took place across the blocks. RT decreased, showing faster processing by the participants over time. For CV, there was an initial increasing trend followed by a plateau and then a potential decrease. From the null model (Model 1A) (see Table 3), the ICC was .19, indicating that 19% of the variation was attributed to variation between participants. The best fitting model was Model 1C whose random effect structure was a random intercept by participants and a random slope for ot1 (the linear term for Block). Interpreting results of this model, both ot1 ($p = .02$) and ot2 ($p < .001$) were significant. The positive sign of the coefficient for ot1 indicated an overall increase in CV, while the negative sign for ot2 confirmed the initial increase later leveled off or even potentially decreased from its peak. I 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. The R^2 conditional value was .40, indicating that 40% of the variance was explained by both fixed and random effects combined. In Figure 3, I plotted the distributions and the means of the observed and predicted CV values across blocks.

Although the quadratic term 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).

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 2. Descriptive statistics of the intentional word learning experiment

Block	1	2	3	4	5	6	7	8	9	10
RT (Mean, ms)	1464	1308	1233	1162	1078	1017	988	964	943	926
RT (*SD*, ms)	303	255	229	193	188	178	216	205	196	192
RT (95% CI, ms)	[1360, 1569]	[1220, 1395]	[1155, 1312]	[1095, 1228]	[1014, 1143]	[955, 1078]	[914, 1063]	[893, 1034]	[876, 1010]	[860, 992]
CV (mean)	0.208	0.272	0.301	0.336	0.318	0.337	0.320	0.311	0.322	0.296
CV (*SD*)	0.086	0.082	0.089	0.063	0.088	0.074	0.088	0.098	0.113	0.098
CV (95% CI)	[0.178, 0.238]	[0.244, 0.301]	[0.271, 0.332]	[0.315, 0.358]	[0.288, 0.348]	[0.312, 0.362]	[0.290, 0.350]	[0.277, 0.345]	[0.283, 0.360]	[0.262, 0.330]
Accuracy (Mean)	0.666	0.720	0.795	0.864	0.920	0.938	0.948	0.964	0.966	0.980
Accuracy (*SD*)	0.472	0.450	0.404	0.343	0.272	0.242	0.222	0.186	0.181	0.139
Accuracy (95% CI)	[0.627, 0.705]	[0.682, 0.757]	[0.761, 0.828]	[0.836, 0.893]	[0.897, 0.942]	[0.917, 0.958]	[0.930, 0.967]	[0.949, 0.980]	[0.951, 0.981]	[0.969, 0.992]

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 3. Model summaries for intentional word learning experiment (all blocks)

Model		Fixed Effects Coefficients			Variance Components				Intercorrelation of Random Components			R^2 (Marginal)	R^2 (Cond'l)	AIC
		Intercept (*SE*)	ot1 (*SE*)	ot2 (*SE*)	Error (*SD*)	Intercept (*SD*)	ot1 (*SD*)	ot2 (*SD*)	Int ~ ot1	Int ~ ot2	ot1 ~ ot2			
1A (null)	ML	0.31*** (0.0085)			0.0079 (0.089)	0.0017 (0.041)						0.00	0.18	–656
	REML	0.31*** (0.0086)			0.0079 (0.089)	0.0018 (0.042)						0.00	0.19	
1B (+ot1 & ot2)	ML	0.31*** (0.0085)	0.050*** (0.014)	–0.085*** (0.014)	0.0068 (0.082)	0.0018 (0.043)						0.10	0.29	–698
	REML	0.31*** (0.0085)	0.050*** (0.014)	–0.085*** (0.014)	0.0068 (0.083)	0.0019 (0.044)						0.10	0.30	
1C (+ot1 random)	**ML**	**0.31*** (0.0085)**	**0.050* (0.021)**	**–0.085*** (0.013)**	**0.0058 0.076**	**0.0019 (0.044)**	**0.0092 (0.096)**		**0.58**			**0.10**	**0.40**	**–718**
	REML	**0.31*** (0.0086)**	**0.050* (0.021)**	**–0.085*** (0.013)**	**0.0058 0.076**	**0.0020 (0.045)**	**0.0096 (0.098)**		**0.58**			**0.10**	**0.40**	
1D (+ot2 random)	ML					Singular fit								
	REML					Singular fit								

Notes: Parameters of the final, best fitting model are **bolded**; *= $p < .05$; **=$p < .01$; ***=$p < .001$

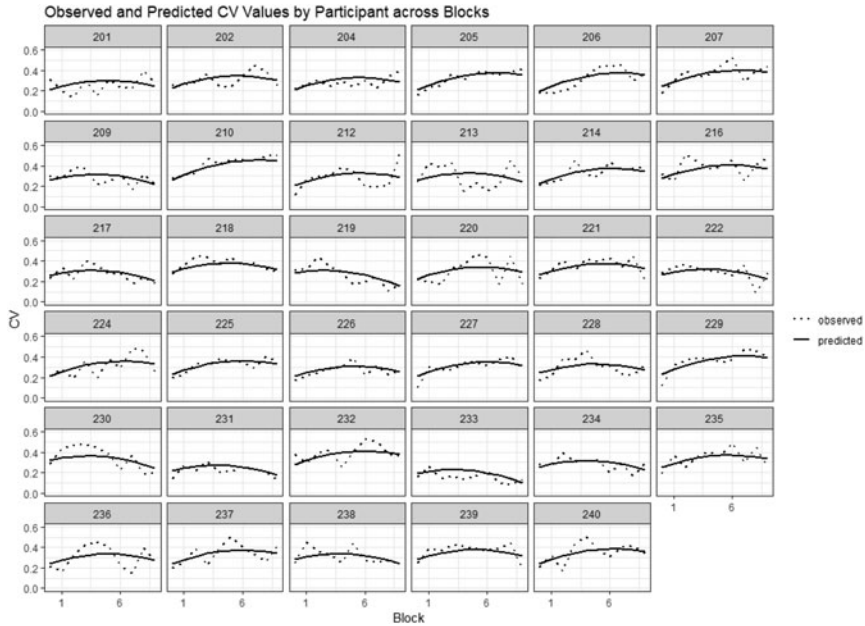


FIGURE 2. Observed and predicted CV values by participant across blocks

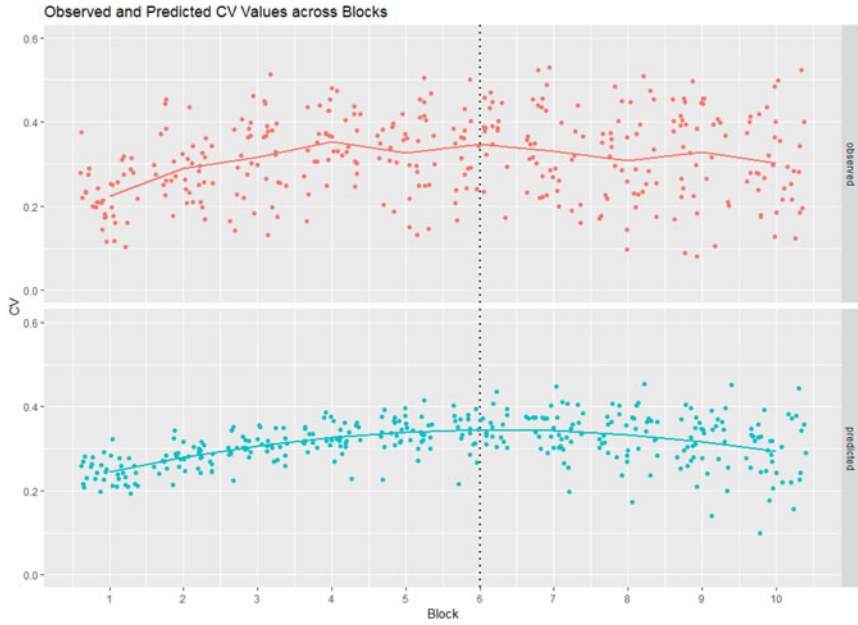


FIGURE 3. Observed (upper panel) and predicted (lower panel) CV values across blocks.

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

Block 6 was also the stage when CV peaked descriptively ($M = .337$, $SD = .074$, 95% CI [.312, .362]). I built an additional model using this subset of the data. I first recoded Block such that the nth Block in the actual experiment was coded as n-6. This way, performance as the sixth block was reflected by the intercept term. Because visual inspection of the data did not reveal any curvature in this part of the trajectory, no higher order time terms were considered necessary. As such, there were also no needs to center the recoded linear term of Block. This term was the only fixed effect attested. From the null model (Model 2A), the ICC value was .39, indicating 39% of the variance was attributed to that between participants. Following the model selection procedure described, I identified the optimal random effects structure in Model 2B, which only included a random intercept for participant. This model confirmed a significant decrease ($p = .039$) in CV from Block 6 to Block 10 (see Table 4 for a model summary). The overall R^2 conditional value was .41. Despite its significance, caution should be exercised when interpreting this decrease. First, the confidence intervals for CV were overlapping between Block 6 and Block 10, suggesting the difference may not be robust. Also, the CV value in the final block did not decrease to the level at Block 1, which I will return to in the "Limitations" section. Summarizing the results of the intentional word learning experiment, 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 afterward, forming a roughly inverted U-shaped trajectory.

REANALYSES OF INCIDENTAL LEARNING EYE-TRACKING DATA

Tables 5 and 6 present descriptive statistics for first fixation and gaze durations as well as both CV measures across occurrences of the target and control words in the text. First fixation durations were relatively consistent across occurrences while gaze durations had a downward trend. This decrease indicated gains in familiarity of orthographic forms as reported in Elgort and colleagues' original study. For CV values, first fixation CV appeared to be comparatively stable across occurrences. There was seemingly a decrease in gaze duration CV on target words across occurrences; but this potential decrease needed to be considered relative to the control words. In the mixed-effects model with CV based on first fixations specified as the outcome variable, the null model (Model 3A) suggested minimal variance due to the nested data structure (ICC = .08). The final model (Model 3B) was a random intercept-only model. Interpreting this model, only Word Type emerged as a significant predictor ($p = .03$) (see Table 7 for a model summary), meaning that the readers processed high-frequency words in a less variable (more stable) manner. Occurrence in the text ($p = .46$) and the critical interaction between the two predictors ($p = .20$) were not significant. The absence of an interaction suggested that the readers did not process the target words in a more (or less) stable manner progressively, compared with the control words. The R^2 conditional value was .09. This finding indicated that the readers' processing variability (stability) did not change as the participants encountered the unfamiliar words repeatedly.

For gaze duration CV, again, the null model (Model 4A) suggested minimal intraclass correlation (.06). The best fitting model was also a random intercept-only model (Model 4B). Only Word Type ($p < .001$) emerged as a significant predictor in this model. This significance indicated processing of familiar, high-frequency words was less variable

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 4. Model summaries for intentional word learning experiment (blocks 6–10)

		Fixed Effects Coefficients		Variance Components			Intercorrelation of Random Components			
Model		Intercept (*SE*)	Block (*SE*)	Error (*SD*)	Intercept (*SD*)	Block (*SD*)	Int ~ ot1	R^2 (Marginal)	R^2 (Cond'l)	AIC
2A (null)	ML	0.32*** (0.012)		0.0065 (0.081)	0.0041 (0.064)			0.00	0.38	–329
	REML	0.32*** (0.013)		0.0065 (0.081)	0.0042 (0.065)			0.00	0.39	
2B (+Block)	**ML**	**0.34*** (0.015)**	**–0.0089* (0.0043)**	**0.0064 (0.080)**	**0.0041 (0.064)**			**0.015**	**0.40**	**–331**
	REML	**0.34*** (0.015)**	**–0.0089* (0.0043)**	**0.0064 (0.080)**	**0.0042 (0.065)**			**0.015**	**0.41**	
2C (+Block random)	ML					Singular fit				
	REML					Singular fit				

Notes: Parameters of the final, best fitting model are **bolded**; *=$p < .05$; **=$p < .01$; ***=$p < .001$

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 5. Descriptive statistics of first fixations in the eye-tracking experiment

Occurrences	1	2	3	4	5	6	7	8	9	10	11	12
First Fixation – Target												
FF (Mean, ms)	252	256	259	257	249	261	240	248	250	245	226	247
FF (*SD*, ms)	97	94	106	93	99	112	90	91	90	88	84	83
FF (95% CI, ms)	[241, 263]	[244, 268]	[246, 272]	[246, 269]	[238, 260]	[247, 275]	[228, 251]	[237, 258]	[239, 260]	[235, 256]	[215, 237]	[237, 257]
CV (mean)	0.343	0.330	0.345	0.315	0.338	0.331	0.333	0.326	0.316	0.319	0.345	0.307
CV (*SD*)	0.124	0.103	0.132	0.094	0.100	0.117	0.112	0.107	0.099	0.106	0.126	0.087
CV (95% CI)	[0.300, 0.386]	[0.294, 0.366]	[0.299, 0.391]	[0.282, 0.348]	[0.303, 0.373]	[0.291, 0.372]	[0.294, 0.372]	[0.288, 0.363]	[0.282, 0.351]	[0.282, 0.356]	[0.301, 0.389]	[0.277, 0.338]
First Fixation) – Control												
FF (Mean, ms)	232	223	242	228	244	249	250	234	243	221	235	236
FF (*SD*, ms)	73	82	85	74	97	83	88	77	93	81	79	79
FF (95% CI, ms)	[223, 242]	[212, 234]	[231, 253]	[218, 238]	[231, 258]	[238, 260]	[238, 261]	[224, 244]	[230, 257]	[209, 233]	[224, 247]	[224, 248]
CV (mean)	0.276	0.331	0.329	0.294	0.332	0.304	0.318	0.286	0.335	0.316	0.324	0.311
CV (*SD*)	0.083	0.113	0.107	0.099	0.148	0.112	0.163	0.098	0.134	0.109	0.085	0.098
CV (95% CI)	[0.247, 0.305]	[0.291, 0.370]	[0.292, 0.366]	[0.260, 0.328]	[0.280, 0.385]	[0.265, 0.343]	[0.261, 0.375]	[0.252, 0.320]	[0.288, 0.381]	[0.278, 0.354]	[0.294, 0.354]	[0.277, 0.345]

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 6. Descriptive statistics of gaze durations in the eye-tracking experiment

Occurrences	1	2	3	4	5	6	7	8	9	10	11	12
Gaze Durations – Target												
GD (Mean, ms)	347	335	338	323	339	306	319	323	309	295	278	288
GD (*SD*, ms)	179	169	171	150	178	142	153	148	145	132	148	132
GD (95% CI, ms)	[326, 368]	[314, 356]	[318, 359]	[304, 341]	[318, 360]	[288, 324]	[300, 339]	[306, 341]	[292, 326]	[280, 311]	[259, 296]	[273, 304]
CV (mean)	0.478	0.448	0.430	0.408	0.497	0.395	0.431	0.426	0.397	0.393	0.481	0.375
CV (*SD*)	0.131	0.161	0.151	0.144	0.120	0.118	0.155	0.106	0.131	0.141	0.151	0.147
CV (95% CI)	[0.433, 0.524]	[0.392, 0.504]	[0.377, 0.483]	[0.358, 0.458]	[0.455, 0.539]	[0.354, 0.436]	[0.377, 0.485]	[0.389, 0.463]	[0.351, 0.443]	[0.344, 0.442]	[0.428, 0.533]	[0.323, 0.426]
Gaze Durations – Control												
GD (Mean, ms)	275	278	285	269	286	298	309	280	275	248	273	268
GD (*SD*, ms)	121	130	113	124	112	115	129	127	121	109	112	115
GD (95% CI, ms)	[259, 290]	[261, 296]	[271, 300]	[253, 285]	[270, 301]	[283, 313]	[292, 325]	[264, 296]	[258, 293]	[232, 263]	[257, 290]	[251, 285]
CV (mean)	0.369	0.430	0.350	0.402	0.359	0.344	0.373	0.349	0.370	0.363	0.366	0.369
CV (*SD*)	0.112	0.135	0.110	0.141	0.133	0.127	0.147	0.098	0.144	0.145	0.105	0.115
CV (95% CI)	[0.330, 0.409]	[0.383, 0.477]	[0.312, 0.389]	[0.353, 0.452]	[0.312, 0.407]	[0.300, 0.389]	[0.322, 0.424]	[0.315, 0.383]	[0.320, 0.421]	[0.312, 0.413]	[0.329, 0.403]	[0.329, 0.409]

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 7. Model summaries for eye-tracking data (first fixation)

Model		Fixed Effects Coefficients				Variance Components		R^2		
		Intercept (*SE*)	Occur (*SE*)	WType	Occur* WType	Error (*SD*)	Intercept (*SD*)	(Marginal)	R^2 (Cond'l)	AIC
3A (null)	ML	0.32*** (0.0067)				0.011 (0.11)	0.0010 (0.032)	0.00	0.084	–1281
	REML	0.32*** (0.0068)				0.011 (0.11)	0.0011 (0.033)	0.00	0.087	
3B (+ Occur* WType)	ML	**0.31*** (0.011)**	**0.0011 (0.0015)**	**0.031* (0.014)**	**–0.0028 (0.0022)**	**0.011 (0.11)**	**0.0011 (0.032)**	**0.0071**	**0.091**	**–1281**
	REML	**0.31*** (0.011)**	**0.0011 (0.0015)**	**0.031* (0.014)**	**–0.0028 (0.0022)**	**0.011 (0.11)**	**0.0011 (0.032)**	**0.0070**	**0.094**	
3C (+Occur random)	ML	Singular fit								
	REML	Singular fit								
3D (+Wtype random)	ML	Singular fit								
	REML	Singular fit								

Notes: Parameters of the final, best fitting model are **bolded**; * = $p < .05$; ** = $p < .01$; *** = $p < .001$

than that of low-frequency target words. Occurrence ($p = .22$) and the critical interaction ($p = .37$) were both not significant. The R^2 conditional value was .12. Again, the lack of an interaction revealed that the readers did not process the target words in a more (or less) stable manner progressively, compared with the control words. The model summary is presented in Table 8.

Overall, in the present reanalysis of Elgort and colleagues' data, both CV measures demonstrated a similar, stable pattern of processing variability across occurrences. At the same time, the main effects of Word Type highlighted differences in processing variability as a result of participants' prior word knowledge.

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 an index of processing variability (stability) along the fuller trajectory of novel word learning and to explore the potential application of CV with eye-tracking data. Looking at both intentional and incidental word learning paradigms, I focused on the pattern of change in processing *during* the process of learning (as opposed to processing products of learning). Results in the intentional word learning experiment involving learners establishing form-meaning mappings replicate the initial increase in CV reported in Solovyeva and DeKeyser. Crucially, the present study extends their work by capturing a quadratic (roughly inverted-U shaped) learning curve in the development of processing stability. In the reanalysis of the eye-tracking data from an incidental vocabulary learning experiment, processing variability was found to be stable despite learners' gains in familiarity with orthographic forms. The main effects of Word Type in the analysis indicated that learners processed high-frequency control words in a more stable (less variable) manner than low-frequency target words, validating the use of CV with eye-tracking measures.

A ROUGHLY INVERTED U-SHAPED TRAJECTORY OF COEFFICIENT OF VARIATON IN INTENTIONAL WORD LEARNING

First and foremost, the significance of the quadratic term (ot2) captured a roughly inverted U-shaped trajectory in processing stability in the intentional word learning experiment where CV values peaked at the sixth block. This study, to my knowledge, is one of the first to report unique evidence for a fuller trajectory of processing stability, other than a pre-/posttest design, observing a level-off or even a potential decrease after an increase in CV. This pattern of change has the potential to explain some of the inconsistency in the literature where an expected decrease in CV (i.e., an indication of automatization) was not found. This finding also empirically confirms Lim and Godfroid's (2015) claim that when the learner is still in the stage of acquiring declarative and procedural knowledge, s/he then has not proceeded to automatization, hence a decrease in CV should not be predicted.

My data suggest that, at first, the participants were still learning some of the new words even after the learning block, establishing declarative knowledge from a skill learning perspective. The accuracy rate at Block 1 was 67%, and some of the correct trials were

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

TABLE 8. Model summaries for eye-tracking data (gaze duration)

		Fixed Effects Coefficients				Variance Components			Intercorrelation of Random Components			
Model		Intercept (*SE*)	Occur (*SE*)	WType	Occur* WType	Error (*SD*)	Intercept (*SD*)	Occur (*SD*)	Int ~ Occur	R^2 (Marginal)	R^2 (Cond'l)	AIC
4A (null)	ML	0.40*** (0.0076)				0.018 (0.13)	0.0012 (0.035)			0.00	0.063	–928
	REML	0.40*** (0.0077)				0.018 (0.13)	0.0013 (0.036)			0.00	0.066	
4B (+Occur* WType)	**ML**	**0.38*** (0.014)**	**–0.0023 (0.0019)**	**0.073*** (0.017)**	**–0.0024 (0.0026)**	**0.017 (0.13)**	**0.0013 (0.035)**			**0.055**	**0.12**	**–971**
	REML	**0.38*** (0.014)**	**–0.0023 (0.0019)**	**0.073*** (0.017)**	**–0.0024 (0.0026)**	**0.017 (0.13)**	**0.0013 (0.036)**			**0.055**	**0.12**	
4C (+Occur random)	ML	0.38*** (0.013)	–0.0023 (0.0019)	0.073*** (0.017)	–0.0024 (0.0026)	0.017 (0.13)	0.0011 (0.032)	0.000015 (0.0012)	0.36	0.055	0.12	–967
	REML	0.38*** (0.013)	–0.0023 (0.0019)	0.073*** (0.017)	–0.0024 (0.0026)	0.017 (0.13)	0.0012 (0.034)	0.000032 (0.0018)	0.10	0.054	0.12	
4D (+Wtype random)	ML						Singular fit					
	REML						Singular fit					

Notes: Parameters of the final, best fitting model are **bolded**; * = $p < .05$; ** = $p < .01$; *** = $p < .001$

likely to have been guesses. Overall accuracy improved in subsequent blocks as more robust representations were being entrenched in memory. At around Blocks 4 and 5, accuracy reached a near-ceiling level of 90%, before leveling off in subsequent blocks. One interpretation, then, is that the learner had moved beyond the acquisition of declarative and procedural knowledge at Block 6. After this block, skill learning theory predicts automatization of the knowledge, hence a decrease in CV.

Given this observed accuracy trajectory, I argue that my CV data pattern well with the predicted course of development according to skill learning theory. In particular, CV increased (i.e., processing was more and more variable) during the first blocks where new representations were added to the lexicon, which echo the overall findings in Solovyeva and DeKeyser (2018). Then, the knowledge was strengthened by practice through exposure and feedback, satisfying the prerequisite of strong linguistic knowledge for automatization. This fulfillment is apparent in the light of the high accuracy rate at around Blocks 4 and 5. Beyond that point, in the latter half of the experiment, CV then peaked at Block 6, followed by a significant decrease in the remaining blocks. This trajectory is then in line with automatization as predicted by skill learning theory.

Qualifying the initial increase in CV in word learning, Solovyeva and DeKeyser discussed the heterogeneity in automatization in the lexicon. As some known words in the lexicon are automatized along with development, processing of these particular words becomes more stable; yet, learners, at the same time, add novel representations to the mental lexicon as they learn new words. This addition of new representations and processing routes introduces variability in processing. This interpretation echoes the claim that, at the level of individual words, different words are at different levels of development (Jiang, 2000, 2002). Overall processing stability, then, is a function of the proportion of words at each possible stage of development. Empirically, the overall direction of change in CV along the word-learning trajectory then depends on the strength of two forces relative to each other: automatization driving overall CV down and the addition of representations driving CV up. The data in the present study capture precisely this change in direction across the ten test blocks.

USE OF COEFFICIENT OF VARIATION TO INDEX LEXICAL PROFICIENCY

The present results and those in Solovyeva and DeKeyser (2018) are in sharp contrast with the general prediction that CV should decrease along with development (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 as automaticity develops. The evidence so far suggested that researchers should exercise caution, especially when using CV to index lexical proficiency. It is because the assumption of a general, overall decrease in CV along with development could be violated in the early stages of word learning. As pointed out by one of the reviewers, however, this violation does not necessarily make CV an unreliable measure of word knowledge. On the contrary, the unveiled relationship between accuracy and CV allows researchers to be in a better position to interpret CV data *in conjunction with* accuracy data. In particular, CV values peaked at the sixth block in the current experiment before decreasing. This peak, then, can be interpreted as the start of the automatization process. On this account, to address how robust declarative

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

knowledge needs to be for it to be automatized and at what point CV measures can be interpreted as an automaticity measure, the accuracy data at this block (M = .938, SD = .242, 95% CI [.917, .958]) can indeed shed some important light. Specifically, if this pattern is replicated in other studies, CV can be interpreted as a measure of processing automaticity only after accuracy hits more than 90%, which I rounded down from the lower bound of the confidence intervals. From this point onward, a decreasing CV can be expected, indexing development of processing automaticity.

STABILITY IN COEFFICIENT OF VARIATION IN INCIDENTAL WORD LEARNING

Another important finding of this study was that CV was relatively stable in the learning of orthographic forms in the incidental paradigm. One implication of Solovyeva and DeKeyser (2018) was that CV can potentially be extended beyond a measure of automaticity. An increase in CV, in particular, can index new knowledge representations at both levels of word forms and form-meaning mappings. The findings in the present reanalysis, however, suggest that researchers should approach this implication with caution. CV measures based on first fixation and gaze durations on target words that appeared multiple times in the reading text did not change significantly across the first 12 occurrences. This was despite the learners' gains in familiarity with the orthographic forms as revealed by a quick decrease in reading times. If the decrease in reading times can be taken as some form of learning of orthographic forms, the present reanalysis suggests that CV cannot index this learning or "the emergence of new representations [of word forms]," as Solovyeva and DeKeyser claimed (p. 237).

At the same time, it is important to note that the current reanalysis was not intended to be directly compared with the aspects of Solovyeva and DeKeyser's (2018) study concerning implicit learning of word forms. First, the items (i.e., the learning targets) in Solovyeva and DeKeyser were nonwords, while Elgort and colleagues (2018) used low-frequency words of which learners could already have some knowledge. In this regard, those low-frequency words might have already been represented in the learner's mental lexicon. Put differently, the different initial states of the targets make the two studies not directly comparable. More importantly, the two selected eye-movement measures for the present reanalysis might not be directly comparable with RT data obtained from behavioral tasks such as the lexical decision tasks used in Solovyeva and DeKeyser. Although first fixation and gaze durations are believed to index early lexical processing (Godfroid & Winke, 2015), they have also been shown to only moderately correlate with RT data (Dirix, Brysbaert, & Duyck, 2017; Kuperman, Drieghe, Keuleers, & Brysbaert, 2013). As pointed out by one of the reviewers, one potential explanation for this discrepancy is that lexical decisions involve full access to lexical representations, while early eye-movement measures, such as first fixation and gaze durations, may reflect only access to sublexical orthographic word forms. If that is the case, then, the task demand in terms of the level of lexical access is one of the keys in determining the limit of CV as an index of emerging representations. One direction for further research is to replicate the present analysis using late eye-movement measures, such as total reading times. At the same time, researchers will need to consider the conscious involvement of the readers in such behavior as rereading (regression) and, importantly, the extent to which this involvement might confound the data.

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

USE OF COEFFICIENT OF VARIATION WITH EYE-TRACKING DATA

While the use of CV to index formation of word form knowledge is to be followed up, the present reanalysis has shown that CV can be used to index processing variability with eye-tracking data. Specifically, the main effects of Word Type in the reanalysis suggest that processing of high-frequency words was significantly more stable. This finding plays a crucial role in extending CV research to word processing and learning involving eye tracking. It validates CV as a measure of variability in lexical processing as captured by eye tracking. On this account, SLA researchers, and vocabulary researchers in particular, have one more research tool at their disposal to measure lexical proficiency from a temporal, automaticity perspective. The granularity of CV also potentially helps researchers capture the incremental development of vocabulary knowledge and gain understandings of the relationship between automatic lexical access and such general language skills as reading. In broader terms, the use of CV also aligns with the reconceptualization of dimensions of lexical knowledge with a new, online aspect, revealing the extent to which lexical knowledge is available in actual language use (Godfroid, 2020b).

LIMITATIONS

I acknowledge a number of limitations in the present study. First, in the intentional word learning experiment, I sampled only highly frequent, concrete, and imageable nouns, which did not reflect entirely the learning task for most learners and can only elicit responses to a homogeneous set. I also used an animacy judgment task that was not an authentic language task and could only reflect the learning of certain aspects of form-meaning mappings. Future research can use other online methodologies, such as primed decision tasks, to examine different aspects of processing in a more fine-grained manner (e.g., processing of meaning vs form). The decrease in CV from the peak could have been captured even more clearly if there had been even more practice (e.g., additional test blocks). Ideally, CV should drop beyond the initial value at Block 1, which is not shown in the present data. Finally, the data here did not rule out any further possible increase in CV along the trajectory; in that case, an experiment with more practice potentially over multiple days can address this possibility.

CONCLUSION

Inspired by Solovyeva and DeKeyser (2018), I investigated the signature of CV in the early stages of word learning of orthographic forms and form-meaning mappings in both intentional and incidental paradigms. With a focus on measuring the process of learning, I captured a roughly inverted U-shaped trajectory in CV in the intentional paradigm. This finding confirms that processing stability suffers at the beginning of the trajectory where learners acquired declarative and procedural lexical knowledge. Only after they learn the words (as shown in the peak for accuracy), processing stability improves, and CV starts to decrease, which is then consistent with automatization.

The relative stability of CV in the eye-tracking data highlights that CV as an index of establishing new knowledge representations in the lexicon might not be universally

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

applicable, especially in the learning of orthographic word forms under incidental conditions. Specifically, although learners gained orthographic familiarity as revealed by reduced reading times, this learning had little influence on processing stability. This finding is in sharp contrast with the learning of form-meaning mappings in the intentional learning paradigm. At the same time, processing variability changed as a function of word type (i.e., high- vs, low-frequency words). This main effect validates the use of CV in eye-tracking research in the sense that the present study, echoing the literature which used mostly behavioral tasks, found more stable processing for high-frequency words of which learners are most familiar. In other words, use of CV in eye-tracking research has the ability to index lexical processing variability as a new, online dimension of word knowledge. With a better understanding of the trajectory of CV in both word learning paradigms, researchers are then in a much better position to use CV to index lexical proficiency from a processing perspective.

NOTES

[1]I would like to thank one of the reviewers for bringing this to my attention.

[2]A reviewer suggested presenting evidence of the Swahili stimuli being phonologically and phonotactically legitimate to the participants (i.e., pronounceable for L1 English speakers). While subvocal rehearsal was indeed a likely strategy that the participants had deployed, many language learners in general do not only learn foreign words that are pronounceable in their L1. For this reason, I argue that illustrating the phonotactic (dis)similarity between participants' L1 (English) and Swahili would not necessarily add to the overall arguments made in the study.

[3]A reviewer expressed concern about the task requiring responses that did not necessarily involve retrieval of meaning in that a vague degree of knowledge suffices for completing such a task. While I accept this as a limitation for the study, I argue that at least some aspects of the targets' meaning were likely to be involved in the form-meaning mapping and were tested by the tasks. In particular, animacy is closely related to the concept of a word and its association with other words, both of which are subcategories under word meaning in Nation's (2013) oft-cited framework of component word knowledge.

REFERENCES

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

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Anderson, J. R. (1983). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.

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*, 255–278.

Bartolotti, J., & Marian, V. (2014, January). Wordlikeness and novel word learning. *In Proceedings of the Cognitive Science Society*, *36*, 146–151. 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. *Journal of Statistical Software*, *67*, 1–48.

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

Brysbaert, M., & New, B. (2009). Moving beyond Kuĉera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

DeKeyser, R. (2007). Situating the concept of practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1–18). New York, NY: Cambridge University Press.

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). New York, NY: Routledge.

De Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*, 906–941.

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 887.

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

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*, 341–366.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. New York, NY: Sage.

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

Godfroid, A. (2019). Investigating instructed second language acquisition using L2 learners' eye-tracking data. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 44–57). New York, NY: Routledge.

Godfroid, A. (2020a). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide* (pp. 433–453). New York, NY: Routledge.

Godfroid, A. (2020b). Sensitive measures of vocabulary knowledge and processing: Expanding nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433–453). New York, NY: Routledge.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*, 563–584.

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*, 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* (pp. 325–348). Amsterdam, The Netherlands: John Benjamins.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). New York, NY: Routledge.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge, UK: Cambridge University Press.

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*, 555–582.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*, 617–637.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *The Quarterly Journal of Experimental Psychology*, *66*, 563–580.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York, NY: 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*, 1247–1282.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207.

Marsden, E., Mackey, A., & Plonsky, L. (2015). The IRIS repository of instruments for research into second languages. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice* (pp. 1–22), New York, NY: Routledge.

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McManus, K., & Marsden, E. (2019). Signatures of automaticity during practice: Explicit instruction about L1 processing routines can improve L2 grammatical processing. *Applied Psycholinguistics*, *40*, 205–234.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman and Hall/CRC Press.

Mohamed, A. A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*, 269–293.

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, *31*, 1–26.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

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

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577.

Phillips, N., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics*, *17*, 237–262.

Pili-Moss, D., Brill-Schuetz, K., Faretta-Stutenberg, M., & Morgan-Short, K. (2019). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition*. Advance online publication. doi: 10.1017/S1366728919000543.

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*, 155.

Revelle, W. (2018) *psych: Procedures for personality and psychological research*. Evanston, IL: Northwestern University. https://CRAN.R-project.org/package=psych Version = 1.8.3

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, *33*, 644–653.

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*, 369–385.

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*, 121–136.

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

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

Suzuki, Y. (2018). The role of procedural learning ability in automatization of L2 morphology under different learning schedules: An exploratory study. *Studies in Second Language Acquisition*, *40*, 923–937.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*, 559–599.

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*, 1256.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*, 33–52.

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

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press

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 = Imageability rating; SUBLwf = English written frequency (SUBTLEXUS); EnNLET = English word length; SwNLET = Swahili word length

https://doi.org/10.1017/S0272263119000603 Published online by Cambridge University Press