Studies in Second Language Acquisition (2026), **48**, 350–372
doi:10.1017/S0272263126101703

RESEARCH ARTICLE

Audiovisual (a)synchrony

Reading ahead as a mechanism for L2 vocabulary processing and learning benefits of reading while listening

Jonathan Malone[1] and Bronson Hui[2]

[1]Maryland English Institute, University of Maryland at College Park, College Park, MD, USA and [2]SLLC, University of Maryland at College Park, College Park, MD, USA
Corresponding author: Jonathan Malone; Email: jemalone@umd.edu

(Received 14 October 2024; Revised 11 February 2026; Accepted 14 March 2026)

Abstract

Prior research has reported vocabulary learning benefits from reading while listening (RWL), but few testable predictions have been made regarding an underlying mechanism for benefits. In this study, reading ahead of the audio was examined as a potential source of benefits during RWL. Fifty-nine participants had their eye movements tracked during a contextualized RWL task embedded with 25 pseudowords 10 times each, followed by three learning outcome measures: form recognition, meaning recognition, and meaning recall. Results indicated that (a) reading ahead facilitated initial form decoding as indexed by gaze durations, especially during early instances; (b) it only predicted the deepest level of vocabulary knowledge tested—meaning recall. These findings indicate that reading ahead may reflect the critical alignment between auditory input, orthographic decoding, and phonological encoding processes during reading, thereby facilitating faster processing when reading ahead than not reading ahead, as well as form-meaning mapping for new words.

Keywords: eye tracking; L2 processing; reading while listening; vocabulary learning

Introduction

Instructed Second Language Acquisition (ISLA) research has long been interested in how linguistic input can be manipulated to optimize L2 learning. The proliferation of digital technologies facilitates a wider variety of multimodal input, in particular, in which visual (text and/or images) is coupled with auditory input in meaning-focused activities (e.g., Hui, 2024). Studies examining the relative benefits of multimodal input on ISLA, especially in vocabulary development, have provided promising findings regarding form and form-meaning learning (e.g., Webb & Chang, 2015; Chen, 2021). However, few have provided a theoretical account for the source of these benefits, nor posited a potential mechanism for why these benefits have been found.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

In the present study, we utilized a within-participant design in seeking to isolate one potential source of bimodal learning benefits, reading ahead of the audio, in processing and learning new L2 vocabulary during a reading while listening (RWL) task. Other recent work (e.g., Wisniewska & Mora, 2018; Conklin et al., 2020) has suggested that L2 readers often read ahead when their eye movements are not synchronized with audio, with more advanced L2 readers mirroring L1 readers by doing so more often. Reading in English obliges activation of sequential orthographic decoding and phonological encoding processes (Coltheart, 2005; see Brysbaert, 2022 for a recent review), and phonological encoding of unfamiliar words in particular could be facilitated by learners encountering the audio forms *asynchronously*, during the orthographic decoding process, after orthographic forms are encountered, allowing learners to test phonological predictions for new word pronunciation. However, no prior studies have isolated the effects of reading ahead of the audio either on processing measures or learning outcomes, nor employed research designs allowing for within-participant comparisons. By utilizing eye-tracking measurement, this study examined the effects of reading ahead versus not reading ahead across encounters with new words on the initial stages of orthographic decoding of new words (i.e., gaze duration, or GD), in which initial form-oriented processing of vocabulary (Kida & Barcroft, 2018) goes through a *familiarity check* (Reichle & Sheridan, 2015).

Recent eye-tracking studies examining unfamiliar LX word processing during reading have indicated that faster GD times during initial encounters with unfamiliar items indicate a form processing benefit for new words, which frees up visual processing time during subsequent encounters to facilitate the mapping of form to meaning (Elgort et al., 2018; Pellicer-Sánchez, 2016). The present study extends this line of reasoning to test whether reading ahead of the audio would facilitate a similar form processing benefit (i.e., shorter GD when reading ahead than when reading is synchronous or behind the audio) during initial encounters with new words, facilitating form-form connections more quickly when reading ahead than not reading ahead. Additionally, we utilized the proportion of reading ahead across encounters as a predictor of three vocabulary learning posttests measuring different levels of the early stages of word learning (form/meaning recognition; meaning recall), while accounting for individual differences in proficiency and memory. In this way, we tested the extent to which reading ahead of the audio facilitated initial form-form and form-meaning learning of unfamiliar words during contextualized reading.

Background

Prominent models of continuous L1 English reading, such as EZ-Reader (e.g., Pollatsek et al., 2006; Rayner et al., 2012), emphasize the centrality of lower-level reading processes, involving orthographic decoding and obligatory phonological encoding, particularly for unfamiliar word forms (Brysbaert, 2022). Given robust evidence that reading in English can be considered the translation of print to speech (see Rayner et al., 2012), these models assume that processing during reading mimics auditory processing in being sequential in nature, and necessarily producing sublexical or prelexical spoken representations in memory (Pollatsek et al., 2006, p. 39). An initial check on word familiarity is conducted sequentially in foveal vision, with each fixated word kept in a working memory buffer. While L2 English reading involves unique challenges during accurate visual processing, it is largely assumed to operate along similar parameters, although slower, more variable, and more effortful (see Cop et al., 2015; Godfroid, 2020).

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Obligatory phonological encoding of the spoken form of words immediately following orthographic decoding during reading in English (Coltheart, 2005; Brysbaert, 2022) potentially implies psycholinguistic benefits of RWL. Simultaneous auditory and visual information could provide an unobtrusive pathway to helping L2 readers establish form familiarity with new word forms, resulting in subsequent faster processing and more efficient word recognition, an essential aspect of developing reading fluency in L2 English (Cop et al., 2015). Additionally, providing auditory and visual information simultaneously could facilitate form-form connections by providing a direct and simultaneous comparison between visual text (orthography) and auditory information (phonology) during a continuous and fluent stream of linguistic information. Orthographic decoding would be facilitated at larger grain sizes across exposures to words through accelerating bidirectional form-form mapping, which is essential to efficient word recognition in deep orthographies such as English (Coltheart, 2005; Grainger & Ziegler, 2011), while facilitating phonological and orthographic "bootstrapping" (Gor et al., 2021) in developing visual word identification. Faster and more efficient form-oriented processing from reading ahead than synchronous reading or reading behind the audio could facilitate free additional cognitive resources toward higher-level form-meaning processing (Kida & Barcroft, 2018).

This hypothesis has often been supported through vocabulary learning studies comparing learning outcomes from RWL with reading only (RO), with superior learning from RWL typically reported in both form and meaning outcomes (e.g., Brown et al., 2008; Webb & Chang, 2015; Malone, 2018; Chen, 2021), but moderated by L1 and individual difference variables (Teng, 2024, and The TwiLex Group, 2024 found no differences between RWL and RO in outcomes). Webb and Chang (2012) reported greater gains from RWL over RO among beginner-level L2 English learners, moderated by reading skill and vocabulary size. Malone (2018) found superior learning from RWL compared to RO for intermediate English learners on form and meaning outcomes, robust to proficiency and phonological short-term memory (PSTM), and Chen (2021) replicated Malone's (2018) findings regarding superior effects of form and meaning recognition from RWL compared with RO. Importantly, these general findings of benefits for bimodal input on learning come from studies that included only visual items on learning tests (Uchihara et al., 2022). However, to this point, to our knowledge, no studies have either made or tested predictions regarding the underlying *source* of possible learning benefits during RWL.

An important insight into real-time L1 and L2 reading behavior was reported in Conklin et al. (2020), who compared L1 and L2 eye movement patterns in RO and RWL groups to measure differences among university-level readers. The text consisted of two 1,500-word experimental stories, and the audio was recorded at a rate of roughly 210 words per minute. Reading was infrequently synchronized with the audio (17% for L1; 33% for L2), and for the asynchronous samples, a very high proportion (nearly 90% for the L1 group, and nearly 80% for the L2 group) read ahead of the audio (around 53% overall for L2 readers). If learners read ahead, while keeping audio in close proximity, the phonological encoding process and establishment of phonological-orthographic form-form connections could benefit substantially from immediate auditory feedback, particularly during early encounters with new words. Evidence for processing benefits would be shown through early measures of eye movements such as gaze duration, which are assumed to reflect orthographic decoding and phonological encoding processes in contextualized L2 reading and word recognition (Elgort et al., 2018; Pellicer-Sánchez, 2016; Rayner et al., 2012).

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

In other words, the temporal sequence of psycholinguistic processing during visual word recognition (visual orthographic decoding *followed by* sublexical phonological encoding) (see Jiang, 2000; Brysbaert, 2022) implies that learners see the words first, and then make hypotheses regarding their pronunciation. In this account, the audio would provide an immediate phonological exemplar during obligatory phonological encoding, thereby strengthening the fledgling representation of a newly encountered visual word. Thus, the temporal *misalignment* between the eyes and the audio could provide psychological *alignment* between phonological encoding and the exemplars provided in the audio, potentially strengthening form-form connections. Such a direct comparison would not be possible temporally during synchronous eye fixations and audio presentation, or if learners read behind the audio, since learners would not yet be at the phonological encoding stage of visual word recognition. Similarly, the theoretical benefits conferred in form-form mapping would be evidenced in behavioral, psycholinguistic-based measures of form familiarity (i.e., the familiarity check) as measured by gaze duration (GD; see Pollatsek et al., 2006). Faster GD when participants read ahead than when eye movements were synchronous or behind the audio during subsequent encounters with the target words would provide empirical processing evidence that reading ahead supports form-form mapping.

This account indirectly aligns with the findings from research into multimodality more broadly, as Wisniewska and Mora (2018) examined improvement in L2 pronunciation from different types of captioned video. They reported that L2 participants read target words ahead of the audio around 70% of the time, and significantly improved pronunciation of target items when viewing a captioned video. However, since their study included video, many targets were skipped, and they did not include asynchrony as a predictor of learning. Conversely, while the findings of Conklin and colleagues (2020) provided targeted insights into RWL behavior in L1 and L2 reading, their study made no predictions and had no research questions related to learning outcomes from these processes.

There is a growing body of research indicating vocabulary learning benefits from RWL (e.g., Webb & Chang, 2015; Malone, 2018; Teng, 2018; Chen, 2021). While many of these studies were skillfully designed and produced interesting results, few have provided a theory-based account for *why* simultaneous input modalities should benefit learners, or in what way(s), and even fewer have made testable predictions for potential mechanisms for benefits. Many have suggested possible sources—assisting the mapping of graphemes to phonemes (Goswami & Bryant, 1990), improving L2 text and phrasal segmentation (Webb & Chang, 2015), freeing up mental resources for semantic processing (Malone, 2018), and enhancing affective benefits that could impact motivation (Tragant & Vallbona, 2018). However, the empirical case for psycholinguistic benefits has been sparse, and even contested, with very few testable claims beyond possible sources of benefit in discussion sections of journal articles as avenues for future research.

To our knowledge, only one recent SLA study (Galimberti et al., 2023) examined issues surrounding the temporal relationship between visual text and auditory processing, to examine the dynamics of multimodality at a more granular level. Galimberti et al. (2023) examined the effects of different types of highlighting of text temporally synchronized to audio during video watching, in order to examine whether and which type of highlighting might benefit learners. They found that highlighting words from the onset of captions (300 ms condition) led to greater attention and rejection of mispronunciations of unfamiliar words, which they took to be evidence of superior phonolexical updating (i.e., more accurate mapping of new

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

word spelling to pronunciation). However, they did not operationalize synchrony between visual and auditory information as an independent variable, but rather at the level of condition. Additionally, the inclusion of video in both Galimberti and colleagues (2023) and Wisniewska and Mora (2018) rendered the eye-tracking data much more variable and difficult to interpret, and introduced the theoretical construct of visual-verbal dual coding (see Paivio & Desrochers, 1980).

In order to justify a research agenda focused on learning and instructional materials that include bimodal input and to examine their effects on the initial stages of memory encoding of form and meaning, the theoretical basis underlying assertions of the benefits of RWL on processing and learning must be matched by empirical tests regarding the source(s) of these benefits. The present study was designed to provide a unique connection between psycholinguistic theory, language processing, and vocabulary research in ISLA. In it, we first tested theoretical predictions at the *individual* level that reading ahead of the audio (i.e., instances where eye fixations on target items occur temporally prior to the auditory presentation) would facilitate faster familiarity checks than when participants were not reading ahead for novel pseudoword forms (measured by GD) across encounters with new words, which would be evidenced by reading ahead as a significant negative within-participant predictor of GD. If simultaneous audio facilitates the phonological encoding process during reading, we expected that it would accelerate the word recognition process itself, resulting in faster familiarity checks across encounters when participants read ahead, compared with reading synchronous to or behind the audio. Since other recent work has indicated that learners begin to process new words similarly to familiar words after 4–6 encounters within a text (Pellicer-Sánchez, 2016), we anticipated this phonological encoding processing advantage for reading ahead to reveal itself early, during the first 4–6 instances of the words in the text, although we predicted comparable GD from reading ahead and not reading ahead during the first 1–2 encounters. Secondly, we predicted that more efficient orthographic decoding and phonological encoding of unfamiliar words across 10 instances in a contextualized reading task would result in better learning both of pseudoword form and meaning, evidenced by reading ahead as a within-participant predictor of learning in mixed-effects outcome models. To examine these issues, we asked two research questions and made two predictions:

RQ1: Does reading ahead of the audio in RWL facilitate form familiarity across encounters with new words (i.e., faster GD when reading ahead than not reading ahead)?

Prediction 1: Reading ahead of the audio will be a significant *negative* predictor of GD across instances (i.e., in growth curve model), resulting in faster familiarity checks across encounters when participants read ahead of the audio than when they are not reading ahead.

RQ2: Does reading ahead of the audio in RWL contribute to learning of new word form and meaning, as evidenced by posttest outcomes?

Prediction 2: Reading ahead will be a significant predictor of form and meaning learning, evidenced by the proportion of reading ahead as a significant predictor of three explicit learning outcomes (form/meaning recognition, meaning recall).

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Method

For this study, we promoted open science by providing all Supplemental Material, data, and analytical code on Open Science Framework (https://osf.io/cpbqd/?view_only=dfeba92f11ee4301ae44154e33e18c83). This study was part of a larger, pre-registered project examining effects of bimodal input on L2 vocabulary processing and learning which can be found at https://osf.io/a95g6.

Participants

With planned generalized linear mixed-effects models, we estimated optimal sample size using Judd et al.'s (2017) recommendations and web-based application. Optimal power was set to Cohen's (1962) recommendation of .8 or above, the alpha level at .05, and the smallest effect size of interest 0.7. Given the number of targets, this resulted in an estimated sample size of 50. Since other studies utilizing similar methodology with eye-tracking data (e.g., Pellicer-Sánchez, 2016; Godfroid et al., 2018) mentioned substantial track loss, we collected initial datasets from 60 participants. One participant's session was interrupted, and those data were discarded, so analyses proceeded with full datasets from 59 participants. Institutional Research Board approval was granted at a U.S. university in the mid-Atlantic region, and participants were paid $15 for Part 1 of the study and an additional $35 for Part 2 (see Table 3). Participants were mostly undergraduate and graduate university students, from a variety of L1 backgrounds. Table 1 indicates demographic information for participants in terms of age, education, and age of fluency in speaking and reading in English, while Table 2 indicates L1 background. All participants reported normal or corrected-to-normal sight and hearing.

Experimental task and pseudoword characteristics

Participants read a modified translation of *How Much Land Does a Man Need?* by Leo Tolstoy (original version obtained from The Literature Network, 2022; see Supplemental Material A for the complete revised version). Nouns in object positions were chosen for pseudoword replacement (see Birch & Rayner, 2010; Moravcsik & Healy, 1998). Nouns were chosen to maximize parsimony in analytic models, even though this choice limited the generalizability of findings. We particularly wanted to focus on how new nouns are processed and learned, while subsequent planned research can diversify part of speech. All other words in the text above the most frequent 4,000 word families in English (see Webb & Nation, 2017) were identified utilizing the Compleat VocabProfiler (Cobb, 2022) and replaced with more frequent synonyms.

Table 1. Demographic information of participants

Variable	Mean	SD	Min	Max
Age (years)	26.07	4.11	19	41
Length of residence (months)	18.56	24.6	1	96
Formal education (years)	17.86	2.78	12	27
Age of spoken fluency	17.1	4.87	7*	29
Age of reading fluency	14.68	4.78	7*	29

*Five participants had become fluent in reading, listening, or both in English prior to the age of 12.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Table 2. Language background of participants

L1 background	Amharic: 1 Bengali: 4 Cantonese: 1 French: 1 Gujarati: 3 Hindi: 10 Indonesian: 1 Malayalam: 5 Mandarin: 19 Marathi: 5 Portuguese: 1 Sinhala: 1 Spanish: 1 Tamil: 2 Telegu: 3 Urdu: 1

This resulted in the replacement of 206 additional words in the text (2.8% of the total words), to ensure lexical coverage above 95% for the most frequent 4,000 word families in English, thereby increasing the likelihood that participants would understand the texts (Waring & Takaki, 2003). Three comprehension questions per chapter, which did not require knowledge of the target pseudowords, were piloted and implemented to ensure focus on meaning. Pilot testing indicated that participants in the same demographic and roughly the same proficiency level were sufficiently able to read the text and answer the comprehension questions accurately.

Supplemental Material B provides information regarding lexical and contextual characteristics of the original words and pseudoword replacements in the text. During initial text development, the individual words or groups of words were selected among existing concrete and imageable nouns in the text, given their centrality to communicating meaning and the likelihood that participants would try to infer meanings during reading. Given this prerogative and the topic of the text, original words were selected that ranged in meaning from family items, to clothing, to food/drink (tea, kumiss/beer, mutton/bread), to agricultural terms (i.e., hillock, meadow/pasture).[1] Words with similar or identical meanings (e.g., "child" and "son") in the original were combined into a single pseudoword item (e.g., "lidger"). In this way, original words from the English translation of the story we used that varied in frequency (e.g., "kumiss/beer") could be collapsed into a single pseudoword ("snall"), with the general meaning of "a drink" as the correct answer choice in the multiple-choice meaning recognition test. Since only the initial stages of form-meaning mapping were measured, this was considered adequate for maintaining both ecological validity and as much of the original structure of the story as possible.

[1]Anonymous reviewers also noted the possibility that the different types of original nouns selected for pseudoword replacement could have impacted their learnability. Given our focus on ecological validity for the reading task, our process for ensuring sufficient contextual information for meaning inference (see Supplemental Material B for a detailed description), and the fact that lexical characteristics of the pseudoword targets were highly similar to one another, we considered them viable to employ. We heartily agree that future studies could examine the potential role of variability in meaning of *original* words in the text, as they could potentially influence processing and form-meaning mapping of pseudoword replacements, but are beyond the scope of the present study.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Concreteness for original items was measured utilizing Brysbaert et al.'s (2014) method and was consistently high; frequency from SUBTLEX-US (Brysbaert & New, 2009) varied more than concreteness, but since original words were replaced in the text, it was considered less important than concreteness for semantic inferencing (e.g., Paivio, 2013). Pseudoword spelling consistency (following English orthographic and phonological conventions) was calculated by averaging rime scores per syllable, utilizing Chee et al.'s (2020) method, and was generally high (0.77 on a 0-1 scale), indicating that phonology and orthography for the targets were relatively connected. The average range of distribution of occurrence was between 500 and 700 words, and Elgort and Warren's (2014) procedure was followed in assessing contextual information (see Supplemental Material C for a full description).

Given the focus on developmental trajectory across instances with the text in processing, along with examples of other recent vocabulary processing studies that found developmental effects most often during the first 10 encounters (Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016), we decided to standardize the repetitions in the text at 10. Instances were distributed across the story as much as possible, typically 500–700 words apart in the text, and care was taken to ensure that knowledge of the pseudowords was not necessary to answering the comprehension questions. See Supplemental Material A for the full text, and Supplemental Material B for additional information about the pseudoword replacements and their location in the reading text.

Apparatus and audio recordings

Conklin and Alotaibi's (2023) methodological recommendations were followed for eye-tracking data collection, utilizing an EyeLink 1000 Plus desk-mounted camera that sampled monocular (right or left) eye movements at 1000 Hz (SR Research, 2010). Participants viewed a computer screen (size: 24 inches, 1920 × 1080 pixels, 60 Hz refresh rate) at a distance of 66 cm, with a chin and forehead rest for stability. Breaks followed each chapter, with re-calibration prior to the following chapter. Task presentation consisted of 190 trial screens, with text triple spaced, black font against white background, monospaced 18-point Courier New font (to ensure psycholinguistic consistency in letter width). Each screen consisted of 2–7 lines of text, left-aligned, and 1-inch margins to minimize spatial recording error. No more than four targets were presented on a single screen, none fewer than four words apart, or on the first or last line, as the first or last word in a line, or the first or last word in a sentence (Rayner & Pollatsek, 2016; Godfroid, 2020). One instance of one word was erroneously included as the final word in a sentence, and these data (59 total points) were discarded.

Each screen was visible for the duration of the audio recording, with a two-second delay after the audio concluded. Participants could not return to a prior screen (as explained during the initial instructions to each participant) and were provided with a nine-point calibration prior to each session and after each break, in addition to drift corrections in the top-left at the onset of each screen. Based on visual inspection of the data in real-time, additional calibrations were performed as needed. Participants were reminded to read along with the audio if they were reading noticeably ahead or behind, in order to ensure that they were attending to the reading task without ignoring either the visual text or the audio. Very few reminders were necessary for the participants, indicating that they were both reading and listening to the text. Length of audio recordings varied from 20 to 39 seconds per trial, with audio manipulated to standard

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

rate and volume with Audacity software. Pilot testing indicated that participants at this level of English ability could comfortably read while listening at an audio rate of 160–180 words per minute, so it was manipulated to that speed (following Pawlas et al., 1996). Audio versions of the story were recorded by a female native speaker of English and standardized to this rate. To ensure that participants were comfortable during the task and to minimize fatigue, breaks were embedded at the end of each chapter, where participants could relax and move around freely before continuing the task.

Vocabulary outcomes

Three unannounced vocabulary posttests measured initial learning—form/meaning recognition and meaning recall. These were selected to follow a theoretically based pattern of development in receptive vocabulary knowledge at the initial stages of learning from form-form to form-meaning connections, measured through receptive form recognition → receptive meaning recognition → receptive meaning recall (L2-L1 translation) measures (Chaudron, 1985; Webb, 2007; Webb & Nation, 2017). These three measures were deployed utilizing the Psychopy3 Builder interface (Peirce et al., 2019) and its online Pavlovia platform. All items were scored dichotomously and presented in randomized order and modality (visual or auditory) within each test. First, the form recognition task presented 50 items (25 targets, 25 distractors) one at a time with the question "was this word in the story"? (Supplemental Material D). Secondly, participants were instructed to write an L1 translation of each target in a meaning recall task (Supplemental Material E). Finally, a multiple-choice meaning recognition measure employed one target and two distractors (Supplemental Material F). The meaning recognition task was created with each answer (correct and two distractors) including either the correct semantic category for more general terms (i.e., the correct answer for "Berrow" (clothing terms in the original text) was "something you wear," rather than "something you eat" or "something you look at") or the specific meaning of targets that were more narrow in meaning (e.g., the correct answer for "Taive" was "a tool," rather than "a vehicle" or "a person"). Each of these distractors was a viable choice from the text compared with the original and was piloted with a group of 10 participants from a similar demographic (see Supplemental Material B for the full list). For each of the three learning outcomes, auditory items could be heard up to three times.

Individual difference measures

Given previous findings regarding the facilitative roles of L2 proficiency (e.g., Tekmen & Daloğlu, 2006; Chen, 2021) and phonological short-term memory (PSTM) (e.g., Martin & Ellis, 2012; Malone, 2018) in contextualized vocabulary learning under incidental conditions, it was considered important to account for both. Because the treatment included both visual and auditory information, we also sought to include proficiency and PSTM measures in both modalities, for a more complete picture of each construct. Three measures of proficiency were administered: (a) an untimed cloze measure of English ability (Brown, 1980) (Supplemental Material G), (b) an auditory version of the LexTALE task (Lemhöfer & Broersma, 2012; Supplemental Material H), and (c) first-pass reading times for a two-screen story preview (with no audio) as a measure of reading speed (Supplemental Material I).

Two PSTM measures were used from the High-Level Language Aptitude Battery (Hi-LAB; Linck et al., 2013; Doughty, 2019): (a) Auditory running memory span (RMS)

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Table 3. Study Tasks and Procedure in Chronological Sequence

Part	Task(s)	Measure(s)	Duration
1	Informed consent	N/A	<2 min
1	Proficiency tasks	Cloze measure; auditory LexTALE	15–20 min
1	Phonological short-term memory/ general processing	RMS; NWS	20–25 min
1	Language background questionnaire/participant payment	N/A	5–10 min
End of Part 1; Part 2 scheduled with each participant on a subsequent day			
2	Informed consent	N/A	<2 min
2	Reading task & comprehension questions	Reading speed (summary pages); eye-tracking data; comprehension questions	65–80 min
2	Vocabulary posttests	Form recognition; Meaning recognition; Form-meaning recall (auditory and visual items)	15–30 min
2	Debriefing/participant payment	N/A	5–10 min

(Pollack et al., 1959; Bunting et al., 2006) presented auditory lists of 12–20 Roman letters at a rate of three per second, with participants instructed to recall the most recent six in order; and (b) a visual nonword span task (NWS) (Gathercole et al., 2001; Linck et al., 2013) presented lists of seven nonwords visually one by one, followed by presentation of a nonword with the question of whether it was in the previous list.

Procedure

Table 3 describes the procedure for the study, including a language background questionnaire (Supplemental Material J) based on LEAP-Q (Marian et al., 2007). Part 1 consisted of language background, proficiency, and memory measures, while Part 2 included the RWL task, comprehension questions (Supplemental Material K), vocabulary learning outcomes, and debriefing (Supplemental Material L). Table 3 also includes the time duration for each task in each of the two parts of the study.

Data analysis

Eye-tracking data cleaning and readahead coding

The four-stage fixation cleaning procedure from the Eyelink DataViewer program was utilized (SR Research Ltd.; https://www.sr-research.com/), along with their default cognitive configuration for parsing. Vertical drifts were manually corrected, and no track loss was detected. The initial eye-tracking dataset included 14,691 data points (59 participants × 249 interest areas). Skips of target interest areas were removed to minimize skew in the dataset, which resulted in the removal of 351 data points (2.4%). The cleaned eye-tracking dataset also included extracted start (onset) and finish (offset) timings for interest areas including target words (following Conklin & Alotaibi, 2023). The readahead variable was then coded dichotomously (1) for instances where the start and end of the first pass reading time were prior to the onset of the audio, and (0) for instances where the start and end of the first pass time were after the onset of the audio.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Model building

Following data cleaning, to answer RQ1, we conducted a growth curve analysis (Godfroid, 2020; Hui, 2020; Mirman, 2014) with GD as the outcome. Model fitting proceeded utilizing the *lmer* function in the *lme4* package (Bates et al., 2015) in R version 4.1.2 (R Core Team, 2024). During model fitting, restricted maximum likelihood (REML) was used as the estimation method, and GD was specified as the outcome. Distributions of GD data are positively skewed, so they were log transformed with base e. Predictor variables in the initial model were reading ahead (dummy-coded 0 (no) and 1 (yes)), and instance (i.e., the repeated occurrences of the target words), along with the interaction term. To minimize potential confounds, we included five covariate predictors in the initial model: the three individual difference measures in proficiency (cloze, LexTALE, and reading speed) and the two measures of PSTM (RMS and NWS).

For RQ2, three binomial logistic mixed-effects models (Jaeger, 2008) were built utilizing the *glmer* function in the same *lme4* package in R (Bates et al., 2015) version 4.1.2 (R Core Team, 2024). Because scores on form recognition distractor items were high (70–85% accuracy), this was considered strong evidence of accurate rejection of false positives, so distractor items for form recognition were removed from the analysis for parsimony (following Godfroid et al., 2018). Models were built for each learning outcome, with the continuous predictor of the proportion of reading ahead of the audio included as a fixed effect of interest. In the initial models, all covariate predictors of proficiency and PSTM variables were included. Additionally, given the role of total reading time (TRT) as a measure of individual attention and a predictor of vocabulary learning gains in other recent eye-tracking studies (e.g., Godfroid et al., 2018), we utilized summed total reading time across all instances for each item as a covariate predictor in the initial models for the learning outcomes.

Comparability for all continuous measures was attained through z-transformations to center and scale. Non-significant covariate fixed effects were removed in a stepwise fashion to arrive at the optimal predictor set, and random slopes for each model (for both RQ1 and RQ2) were included when they did not result in model non-convergence or singularity. Effect sizes were reported as odds ratios for best-fitting logistic models. Supplemental Material M includes all variables and syntax for both initial and best-fitting models for GD and learning outcomes, with syntax for the initial models including all predictors, fixed effects, random effects, and random slopes.

Results

Individual differences, comprehension scores, and measure reliability

Figure 1 indicates correlations between z-transformed individual difference variables, and Table 4 indicates descriptive statistics for these variables. Correlations were mostly weak to moderate, indicating distinct aspects of each construct, so each measure was included in model building. Comprehension question scores were near ceiling (98%), indicating that participants were focused on meaning during the RWL task. Supplemental Material N provides reliability statistics for each vocabulary learning outcome and individual difference variable of interest, with reliability information (Rasch IRT person reliability and Cronbach's α) ranging from .64 to 90. Supplemental Material O summarizes model-fit comparisons for initial and best-fitting models for growth curve analyses for GD data (RQ1) and vocabulary learning measures as outcomes (RQ2).

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

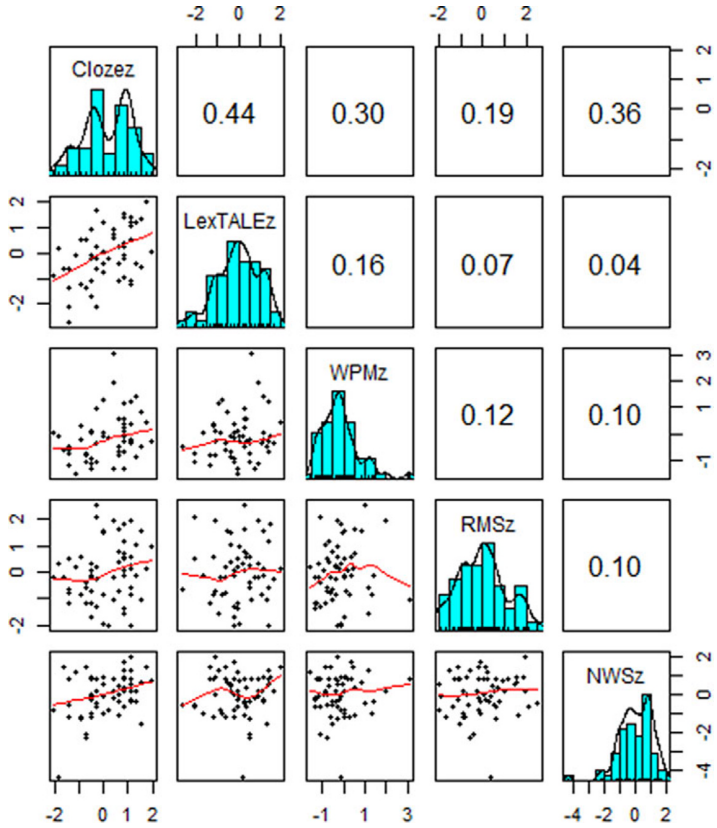


Figure 1. Pearson correlation coefficients and graphs of individual difference variables.
Note: WPM = reading speed in words per minute; RMS = running memory span; NWS = nonword span.

Table 4. Descriptive Statistics for Individual Difference Variables

	Mean	SD	Min	Max	95% CI
Cloze (%)	79.5	4.37	60	96	70.76, 88.2
Auditory LexTALE (%)	66.9	11.2	37.5	88.75	44.5, 89.3
Reading speed (WPM)	225	62.8	125	458	99.4, 350.6
RMS (PSTM)	3.38	0.7	2.05	5.05	1.98, 4.78
NWS (PSTM)	174.5	17.5	106	206	139.5, 209.5

Reading ahead predicting GD (RQ1)

Overall, participants read ahead of the audio nearly 50% of the time, following Conklin et al.'s (2020) findings for RWL (roughly 53% reading ahead for advanced L2 readers). Table 5 and Figure 2 provide and visualize descriptive statistics for gaze duration, separated by instances of reading ahead versus not reading ahead across encounters in the text. The best-fitting linear growth model with GD as the outcome (summarized in Table 6 and visualized in Figure 3) indicated significant effects of reading ahead of the audio ($b = -.16$, $SE = .01$, $p < .001$), but not instance ($b = -.04$, $SE = .02$, $p = .136$), as well

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Table 5. Descriptive Statistics for Gaze Duration for Target Interest Areas by Instance When Reading Ahead of the Audio and Not Reading Ahead

	Reading ahead				Not reading ahead			
Instance	*n*	*M*	*SD*	*95% CI*	*n*	*M*	*SD*	*95% CI*
1	665	339.48	202.97	[324, 355]	781	421.77	228.42	[405, 438]
2	800	353.87	206.74	[340, 368]	654	437.26	213.53	[421, 453]
3	578	337.89	203.89	[321, 355]	875	420.94	215.41	[407, 435]
4	715	341.73	204.44	[327, 357]	727	420.54	218.89	[405, 436]
5	733	345.47	196.58	[331, 360]	695	415.96	211.23	[400, 432]
6	601	345.43	191.20	[330, 361]	838	385.08	202.82	[371, 399]
7	858	393.01	215.96	[379, 407]	528	388.06	217.13	[370, 407]
8	946	352.92	202.82	[340, 366]	487	388.35	192.69	[371, 405]
9	788	347.09	186.94	[334, 360]	643	432.54	218.19	[416, 460]
10	544	349.76	183.84	[334, 365]	884	388.31	216.13	[374, 403]

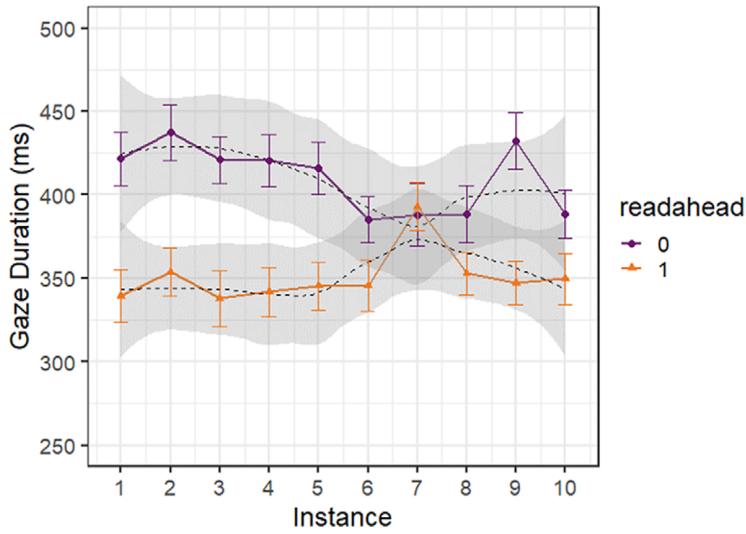


Figure 2. Data visualization for gaze duration raw means across instances by reading ahead (1) versus not reading ahead (0).

as a significant reading ahead by instance interaction ($b = .10$, $SE = .03$, $p = .001$). Both auditory LexTALE ($b = -.06$, $SE = .02$, $p = .003$) and reading speed ($b = -.05$, $SE = .02$, $p = .023$) were significant negative covariate predictors of GD.

In other words, model results indicate that in instances where participants read ahead of the audio, GD was significantly shorter across all 10 instances of pseudowords in the text than when they read synchronously or behind the audio (see Figure 3). Additionally, the reading ahead by instance interaction indicated that any reduction in GD across instances overall was due to a reduction in GD across instances in cases where participants either were synchronized or behind the audio when they initially looked at the targets, but there was no indication that GD times decreased across instances when participants were reading ahead of the audio.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Table 6. Best-fitting Model Summary for Gaze Duration Outcome

Fixed effects	b	SE	95% CI	*t*	*p*	Exp(b)
Intercept	5.87	.02	5.82 to 5.91	260.260	<.001***	354.249
Readahead	−.23	.02	−.26 to −.19	−11.427	<.001***	.79
Instance	−.04	.03	−.09 to .01	−1.490	0.136	.96
Readahead:Instance	.10	.03	.04 to .15	3.308	.001**	1.11
Covariates						
LexTALE	−.06	.02	−.09 to −.02	−3.004	.003	.94
WPM (reading speed)	−.05	.02	−.09 to −.01	−2.266	.023	.95
Random Effects	**Variance**	**SD**				
1|Participant	.02	.09				
1|Item	.01	.14				
Residual	.25	.50				
$N_{participant}$	59					
N_{item}	25					
Observations	14340					
Marginal R^2/Conditional R^2	.048/.146					

Notes. ***p < .001, **p <.01, *p <.05

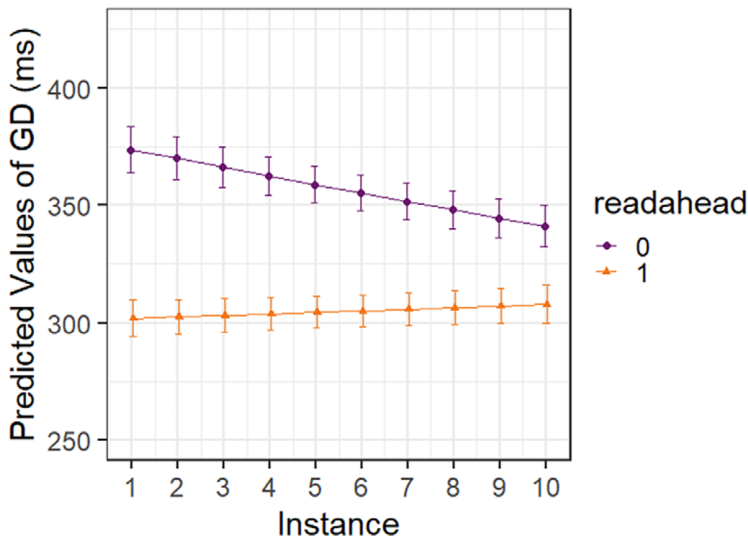


Figure 3. Best-fitting model estimates of readahead by instance interaction for gaze duration.

Reading ahead predicting vocabulary learning (RQ2)

Table 7 indicates descriptive statistics for the three vocabulary learning outcomes. The best-fitting logistic mixed-effects model for form recognition (Table 8) indicated that

Table 7. Descriptive Statistics for Learning Outcome Raw Scores

	k	M (SD)	95% CI
Form Recognition	25	18.71 (4.16)	[17.65, 19.77]
Meaning Recognition	25	14.22 (3.98)	[13.20, 15.24]
Meaning Recall	25	4.12 (4.38)	[3.00, 5.24]

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Table 8. Best-fitting Model Estimates for Learning Outcomes

		Form recognition						Meaning recognition						Meaning recall					
Fixed effects		*b*	*SE*	*95% CI*	*z*	*p*	*OR*	*b*	*SE*	*95% CI*	*z*	*p*	*OR*	*b*	*SE*	*95% CI*	*z*	*p*	*OR*
Intercept		1.31	.17	2.66–5.12	7.847	<.001***	3.69	.40	.20	1.00–2.20	1.984	.047*	1.49	−2.25	.25	.06 to .17	−8.938	<.001***	.11
Readahead		.13	.08	.96–1.34	1.516	.129	1.13	.10	.08	.95–1.29	1.301	.193	1.11	.35	.16	1.05–1.93	2.265	.024*	1.42
Covariate effects																			
LexTALE		N/A	N/A	N/A	N/A	N/A		.31	.10	1.11–1.65	3.012	.003**	1.35	N/A	N/A	N/A	N/A	N/A	N/A
Reading speed (WPM)		N/A	N/A	N/A	N/A	N/A		.26	.11	1.05–1.61	2.385	.017*	1.30	.84	.23	1.47–3.63	3.626	<.001***	2.31
Random effects		**Variance**			***SD***			**Variance**			***SD***			**Variance**			***SD***		
1|Participant	(intercept)	.77			.88			.26			.51			1.71			1.31		
1|Trial	(intercept)	.24			.49			.44			.66			.18			.42		
LexTALE|Item		N/A			N/A			.05			.23			N/A			N/A		
Readahead|Item		N/A			N/A			N/A			N/A			0.19			0.43		

Notes. ***p < .001, **p <.01, *p <.05

Table 9. Summary of Predictions and Alignment of Results

RQ(s)	Prediction (s)	Findings	Alignment between RQ and prediction
1	1	Reading ahead resulted in faster GD than not reading ahead; GD × Instance interaction (decreasing line when not reading ahead; stable line when reading ahead)	✓
2	2	Reading ahead was not a significant predictor of form recognition	✗
2	2	Reading ahead was not a significant predictor of meaning recognition	✗
2	2	Reading ahead was a significant predictor of meaning recall	✓

the proportion of reading ahead for each target was not a significant predictor of form recognition learning (b = .13, SE = .08, p = .129). None of the covariate measures were significant predictors of form recognition in the model.

The best-fitting logistic mixed-effects model for the meaning recognition outcome (Table 8) indicated that reading ahead was not a significant predictor of meaning recognition learning (b = .10, SE = .08, p = .193). Auditory LexTALE (b = .30, SE = .10, p = .003) and reading speed (b = .26, SE = .11, p = .017) were significant covariate predictors of meaning recognition in the model.

The best-fitting logistic mixed-effects model for meaning recall (Table 8) indicated that reading ahead was a significant predictor of meaning recall learning (b = .35, SE = .16, p = .024). Reading speed (b = .84, SE = .23, p <.001) was a significant covariate predictor of meaning recall.

Table 9 summarizes RQs, predictions, findings, and alignment between predictions and findings related to audiovisual synchrony, processing, and learning outcomes. Reading ahead of the audio was not a significant predictor of form or meaning recognition learning (contra predictions), but it was a significant predictor of meaning recall (as predicted), the most difficult of the three learning tasks.

Discussion

Reading ahead facilitates efficient form processing

To our knowledge, the present study was the first to utilize reading ahead as a real-time behavioral predictor of form familiarity during early instances of contextualized vocabulary learning, while accounting for multiple individual differences in L2 proficiency and memory. We defined reading ahead as instances where eye fixations on target items occur temporally prior to the auditory presentation, and utilized reading ahead as an item-level predictor both of processing speed (RQ1) and learning gains (RQ2). In answering RQ1, we found that reading ahead was a significant, negative predictor of gaze duration (GD), indicating that when participants read ahead of the audio, they were significantly faster in form processing (Reichle & Sheridan, 2015) compared with instances in which they did not read ahead, and this finding was durable across 10 instances of the words in the text. The significant effect of instance and the reading-ahead-by-instance interaction for GD revealed that reading ahead provided a processing advantage for the initial form familiarity check, compared with reading

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

synchronously or after the audio, across encounters with unfamiliar words. We predicted this effect, although our expectation was that during initial encounters, GD times would be roughly equivalent when reading ahead or not reading ahead of the audio. As such, the stability of GD from reading ahead as indicated in Figure 3 was surprising. However, our findings (see Figure 3) clearly indicate a processing advantage from the first instance of unfamiliar words in the text.

These findings align with both L1 (Pollatsek et al., 2006) and L2 (Cop et al., 2015) versions of the EZ Reader model, in which initial word recognition during reading involves a temporal sequence of lower-level processes (orthographic decoding → phonological encoding), with faster word recognition indicated by faster GD times (Rayner et al., 2012). Building on previous work examining the development of word recognition across instances in the text (Godfroid et al., 2018; Pellicer-Sánchez, 2016), these findings also provide evidence that reading ahead of the audio may facilitate form processing for unfamiliar L2 words during RWL. Even when individual reading speed, reading ability, and vocabulary knowledge were accounted for in statistical models, participants read the unfamiliar targets significantly faster when reading ahead than when eye movements and audio were synchronized, or when participants were reading behind the simultaneous audio. This *misalignment* between visual and auditory verbal information in the input in this case (i.e., eyes ahead of audio) may indicate psychological *alignment* between presentation of the audio and phonological encoding, an essential stage of orthographic decoding across prominent models of word recognition in English (see Coltheart, 2005; Grainger & Ziegler, 2011; Rayner et al., 2012). In other words, auditory presentation *following* the initial visual decoding process, which aligns the audio and phonological encoding, may facilitate less dependence on initial time spent looking at a word to encode it phonologically, thereby triggering faster eye movement to subsequent words (i.e., word recognition). In contrast, we interpret these findings to indicate that when participants were reading precisely with or behind the audio, the lack of synchrony between the visual decoding/phonological encoding process in the input and in the mind resulted in a slower and more effortful word recognition process, resulting in longer GD times.

Interestingly, this finding was revealed even during the first encounters with unfamiliar items and remained notably stable across instances. Our within-participant analysis showed that participants were consistently faster in the initial decoding process when reading ahead than when not. In cases where participants did not read ahead, the decreasing pattern of GD across instances followed behavioral findings of other reading-only L2 vocabulary studies utilizing eye-tracking methods in developing form familiarity for new words (e.g., Pellicer-Sánchez, 2016), but the stable pattern of GD in cases of reading ahead provides some evidence that reading ahead of the audio may reduce the form-form decoding burden during RWL, even during initial encounters with new words. Additionally, this finding was robust to the within-participant GD analysis, even while accounting for participant differences in reading speed, multifaceted L2 reading and vocabulary proficiency, and PSTM. While other recent studies found evidence that L2 readers often read ahead of the audio (Conklin et al., 2020) and that RWL facilitates the learning of explicit form recognition compared with reading only (Chen, 2021; Malone, 2018; Tuzcu, 2023), this study provides unique initial evidence for an input processing advantage for initial form familiarity as reading ahead *within* RWL. We cautiously interpret these findings to indicate that reading ahead during RWL may facilitate form decoding for unfamiliar words, even during initial encounters. However, since other studies examining the processing of unfamiliar L2

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

words have not operationalized reading ahead as a predictor of processing, additional work in this area is needed.

Reading ahead facilitates form-meaning mapping for new words

RQ2 examined reading ahead as a predictor of three vocabulary learning outcomes, and results indicated that it only predicted meaning recall, the most difficult of the three learning outcome measures. Given high scores on both form and meaning recognition outcomes, we interpret this finding alongside the GD finding to indicate that reading ahead of the audio during initial encounters could facilitate form-form knowledge (evidenced by the GD data), freeing up cognitive resources for semantic processing during subsequent encounters, resulting in stronger connections between form and meaning. This account is supported by the finding that reading ahead predicted the most difficult of the two form-meaning mapping outcomes (meaning recall). Indirectly, this finding also supports the claim that reading ahead facilitates form-form mapping, given tight theoretical links between form and meaning in the establishment and development of L2 lexical items (see Bordag et al., 2022).

After 10 instances in the text, differentiating effects of reading ahead on form and meaning recognition measures were not found (contra our predictions). However, given other findings regarding the central role that repetition plays in contextual vocabulary learning (see Uchihara et al., 2022), it is plausible that the frequency of exposure to new words mitigated effects of reading ahead on form and meaning recognition after 10 instances, and these effects were only uncovered in the most difficult outcome (meaning recall). In this study, the proficiency level of participants was high, so it was unsurprising that substantial learning of pseudoword form and meaning was revealed across 10 instances of pseudowords in the text, indicated by recognition scores (see Table 7), and even under incidental conditions. However, for the productive meaning recall task, which required more precise form-meaning mapping between L2 form and L1 concepts, the effects of reading ahead emerged, above and beyond individual differences in proficiency, reading speed, and memory.

Since no other studies have examined these relationships between reading behavior (i.e., reading ahead of the audio) during RWL and learning gains for contextualized vocabulary, comparisons to other studies are challenging. These findings follow other recent studies examining only learning outcomes from reading only and RWL groups, in reporting significant learning of both form and meaning from RWL (e.g., Webb & Chang, 2015; Teng, 2018). Conklin and colleagues (2020) reported differences in reading behavior during RWL in comparing high-ability L1 readers with high-proficiency L2 readers, in that L2 readers read ahead of the audio less than L1 readers, had fewer skips, and read more slowly than L1 readers. We controlled between-participant variability by measuring and accounting for a battery of cognitive individual differences, including reading speed, to examine whether there would be effects of reading ahead on learning gains. Our findings both of reading ahead as a predictor of GD and meaning recall indicate that for L2 learners, audiovisual asynchrony (i.e., reading ahead of audio) during bimodal reading while listening may play a crucial facilitative role during reading in form familiarity and subsequent processing across exposures, resulting in more efficient word recognition and form-meaning mapping, resulting in better outcomes. In other words, even accounting for individual variability in proficiency, reading speed, and memory, participants learned more about the meaning of the words when they read ahead than when they did not. From an input

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

processing perspective (Barcroft, 2015), these findings could indicate that reading ahead of the audio during RWL could reduce structural demands on processing through aligning phonological encoding with auditory information, thereby facilitating form-meaning mapping. However, since our study is the first to uncover these effects, they should be tested rigorously in future designs, while comparing gains from reading ahead with instances where participants were reading synchronously or behind the audio.

Limitations and future directions

This study had a number of limitations. It included only immediate and receptive learning posttests of form and form-meaning mapping, which limits any claims we can make regarding durability of learning, and did not include a test of word production. Additionally, the type of mapping selected (new L2 form to existing L1 concept) was specific within the broad possibilities for form-meaning mappings in L2 vocabulary learning (see Kida & Barcroft, 2018). As such, generalizable claims regarding learning are limited to L2 form-form and form-meaning mapping. One anonymous reviewer noted that we included both general and specific distractor types on the meaning recognition outcome test, tailored to the meaning associated with that pseudoword item, which could have impacted item responses; 76% of meaning recognition items (19/25) included only specific distractors, 12% (3/25) included only general distractors, and 12% (3/25) had one of each. Descriptive scores indicated notable differences in accuracy between items with specific (58% accurate), generic (67% accurate), or mixed (38% accurate) distractors, so we re-ran the main outcome analysis for the meaning recognition measure with distractor type (specific, generic, mixed) as a categorical predictor. In this analysis, distractor type was not a significant predictor of meaning recognition when other variables were included ($b = -.41$, $SE = .27$, $p = .127$), and did not change other results of the model, so the model without distractor type was retained as the best-fitting model. However, future studies utilizing similar outcome presentation modalities and examining learning of *only* specific or general items can better control for this potential source of variability in learning.

The study was highly controlled, with pseudoword targets rather than real words, while the audio rate was standardized rather than modified to follow reading speed more closely at the individual level. Reading ahead was defined dichotomously yes/no in the analytic models, so it was not possible to examine the relative effects of lagging behind the audio when reading. Since our RQs focused only on the potential benefits of reading ahead, we did not explore differential impacts of reading precisely synchronously or lagging behind the audio; future studies are warranted which do so, thereby increasing precision and allowing for more specific theoretical predictions for instances when the eye and ear are synchronous, and when the eye lags behind the ear during RWL. The L1s among participants were diverse (see Table 2), which did not allow for an examination of the effects by L1, while recent findings from L2 English reading report evidence that L1-L2 differences may result in variable reading behavior, demonstrated by eye movements (Kuperman et al., 2025). This would be a fruitful area for future research.

As the present study provides initial evidence that reading ahead of the audio during RWL facilitates faster word form recognition across instances and mapping form to meaning during contextualized learning of new L2 vocabulary, compared with reading synchronously or behind the audio, it has both research and pedagogical implications. For research, future studies should measure the specific (mis)alignment between visual and auditory information, at group and individual levels, to examine the optimal

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

temporal distance between visual and auditory processing, and further explore the theoretical psychological alignment of phonological encoding and auditory information during reading and audio presentation. At the pedagogical level, instructors should be sensitive to student proficiency levels, and the role of L2 ability and reading speed may be important to how RWL facilitates new word learning. Given the proliferation of multimodal digital materials in language learning contexts that utilize authentic materials (e.g., multimedia newspaper articles, podcasts with automatically generated transcripts), language instructors should be sensitive to differences in ability among their students, and how they could impact processing and learning. Additionally, instructors can utilize technological tools, particularly in adaptive artificial intelligence environments, to design and assign multimodal materials at the relative synchrony of individual students, and embed reminders to follow along with the auditory information.

The present study provides unique insights into a potential mechanism for vocabulary benefits of RWL, indicating that reading ahead of the audio predicted faster familiarity checks (i.e., word recognition) and superior form-meaning mapping of new words compared to reading synchronously or behind the audio, during a contextualized reading task under incidental conditions. The stable rate of predicted GD across instances indicates that RWL may change the nature of early encounters with unfamiliar L2 vocabulary in context, by facilitating form-form (phonological-orthographic) mapping, phonological encoding, and subsequent learning. As such, the findings of this study provide crucial insights into a potential underlying mechanism (audiovisual asynchrony) for previously reported benefits of RWL both on processing and outcomes during contextualized L2 vocabulary learning.

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The data files and R code used in this study are available on OSF: https://osf.io/a95g6.

Acknowledgments. The authors would like to thank Duolingo, The International Research Foundation for English Language Education (TIRF), and the *Language Learning* journal for their grant support for the larger project from which this study came. The authors would also like to thank Mona Alsalmi, Sanshiroh Ogawa, Mireia Toda Cosi, Tetiana Tytko, and Jill Woodward for their assistance with the project.

References

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition*, *38*(6), 740–752. https://doi.org/10.3758/MC.38.6.740

Bordag, D., Gor, K., & Opitz, A. (2022). Ontogenesis model of the L2 lexical representation. *Bilingualism: Language and Cognition*, *25*(2), 185–201. https://doi.org/10.1017/S1366728921000250

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*(10), 1691–1700. https://doi.org/10.1080/17470210600848402

Brown, J. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, *64*(3), 311–317. https://doi.org/10.2307/324497

Brown, R., Waring, B., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language 20*(2), 136–163. https://doi.org/10.64152/10125/66816

Brysbaert, M. (2022). Word recognition II: Phonological coding in reading. In M. Snowling, C. Hulme, & K. Nation (Eds.), *The science of reading: A handbook* (pp.79–101). Wiley-Blackwell.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Brysbaert, M., & New, B. (2009). Moving beyond Kučera 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*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, *7*(1), 1–14. https://doi.org/10.1017/S027226310000512X

Chee, Q., Chow, K., Yap, M., & Goh, W. (2020). Consistency norms for 37,677 English words. *Behavior Research Methods*, *52*(6), 2535–2555. https://doi.org/10.3758/s13428-020-01391-7

Chen, Y. (2021). Comparing incidental vocabulary learning from reading-only and reading-while-listening. *System*, *97*, 102442. https://doi.org/10.1016/j.system.2020.102442

Cobb, T. (2022). Compleat Web VP (Version 2.6) [Computer software]. https://www.lextutor.ca/vp/comp/

Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *The Journal of Abnormal and Social Psychology*, *65*(3), 145. https://psycnet.apa.org/doi/10.1037/h0045186

Coltheart, M. (2005). Modeling reading: the dual-route approach. In Snowling & Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Malden, MA: Blackwell.

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257–276. https://doi.org/10.1177/0267658320921496

Conklin, K., & Alotaibi, S. (2023). Eye-tracking reading-while-listening: Challenges and methodological considerations in vocabulary research. *Research Methods in Applied Linguistics*, *2*(3), 100086. https://doi.org/10.1016/j.rmal.2023.100086

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS one*, *10*(8), e0134008. https://doi.org/10.1371/journal.pone.0134008

Doughty, C. (2019). Cognitive language aptitude. *Language Learning*, *69*(S1), 101–126. https://doi.org/10.1111/lang.12322

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*(2), 365–414. https://doi.org/10.1111/lang.12052

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*(2), 341–366. https://doi.org/10.1017/S0272263117000109

Galimberti, V., Mora, J., & Gilabert, R. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System*, *116*, 103078. https://doi.org/10.1016/j.system.2023.103078

Gathercole, S., Pickering, S., Hall, M., & Peaker, S. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 1–30. https://doi.org/10.1080/02724980042000002

Godfroid, A. (2020). *Eye tracking in Second Language Acquisition and Bilingualism: A research synthesis and methodological guide*. New York: Routledge.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S, Sarkar, A., & Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy lexical representations in adult second language speakers. *Frontiers in Psychology*, *12*, 732030. https://doi.org/10.3389/fpsyg.2021.732030

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hillsdale, NJ: Erlbaum.

Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*(54), 1–13. https://doi.org/10.3389/fpsyg.2011.00054

Hui, B. (2020). Processing variability in intentional and incidental word learning: An extension of Solovyeva and DeKeyser (2018). *Studies in Second Language Acquisition*, *42*(2), 327–357. https://doi.org/10.1017/S0272263119000603

Hui, B. (2024). Scaffolding comprehension with reading-while-listening and the role of reading speed and text complexity. *The Modern Language Journal*, *108*(1), 183–200. https://doi.org/10.1111/modl.12905

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Judd, C., Westfall, J., & Kenny, D. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Kida, S., & Barcroft, J. (2018). Semantic and structural tasks for the mapping component of L2 vocabulary learning: Testing the TOPRA model from a new angle. *Studies in Second Language Acquisition*, *40*(3), 477–502.

Kuperman, V., Schroeder, S., Acartürk, C., Agrawal, N., Alexandre, D. M., Bolliger, L. S., … Siegelman, N. (2025). New data on text reading in English as a second language: The Wave 2 expansion of the Multilingual Eye-Movement Corpus (MECO). *Studies in Second Language Acquisition*, *47*(2) 1–19. https://doi.org/10.1017/S0272263125000105

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid test for advanced learners of English. *Behavioral Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Linck, J., Hughes, M., Campbell, S., Silbert, N., Tare, M., Jackson, S., Smith, B., Bunting, M., & Doughty, C. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*(3), 530–566. https://doi.org/10.1111/lang.12011

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, *40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Marian, V., Blumenfield, H., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martin, K. & Ellis, N. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*(3), 379–413. https://doi.org/10.1017/S0272263112000125

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Chapman and Hall.

Mohamed, A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(2), 269–293. https://doi.org/10.1017/S0272263117000092

Moravcsik, J., & Healy, A. (1998). Effect of syntactic role and syntactic prominence on letter detection. *Psychonomic Bulletin & Review*, *5*(1), 96–100. https://doi.org/10.3758/BF03209462

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General*, *142*(1), 282–287. https://doi.org/10.1037/a0027004

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *34*(4), 388–399. https://doi.org/10.1037/h0081101

Pawlas, A.A., Ramig, L.O., & Countryman, S. (1996). Perceptual Voice and Speech Characteristics in Patients with Idiopathic Parkinson's Disease. *NCVS Status Report #10*, 79–87. Available at http://www.ncvs.org/ProgressReports/NCVS%20Status%20&%20Progress%20Report%20Vol.%2010,%20Nov%201996%20copy.pdf

Peirce, J., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pollack, I., Johnson, I., & Knaff, P. (1959). Running memory span. *Journal of Experimental Psychology*, *57*, 137–146. https://psycnet.apa.org/doi/10.1037/h0046137

Pollatsek, A., Reichle, E., and Rayner, K. (2006). Tests of the EZ Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*(1), 1–56. https://doi.org/10.1016/j.cogpsych.2005.06.001

R Core Team. (2024). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading*. Psychology Press.

Rayner, K., & Pollatsek, A. (2016). Eye movements in reading: A tutorial review. *Attention and Performance XII*, 327–362.

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press

Reichle, E. D., & Sheridan, H. (2015). EZ reader: An overview of the model and two recent applications. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 277–290). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199324576.013.17

SR Research (2010). EyeLink 1000 user manual (Version 1.5.0). http://srresearch.jp/support/EyeLink%201000%20User%20Manual%201.5.0.pdf

Tekmen, E., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, *39*(2), 220–243. https://doi.org/10.1111/j.1944-9720.2006.tb02263.x

Teng, M.F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, *12*(3), 274–288. https://doi.org/10.1080/17501229.2016.1203328

Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, *124*, 103381. https://doi.org/10.1016/j.system.2024.103381

Tragant, E., & Vallbona, A. (2018). Reading while listening to learn: young EFL learners' perceptions. *ELT Journal*, *72*(4), 395–404. https://doi.org/10.1093/elt/ccy009

Tuzcu, A. (2023). Unimodal and bimodal input in incidental vocabulary learning: cognitive processes and the development of different knowledge types (30528835) [Doctoral Dissertation, Michigan State University]. ProQuest Dissertations & Theses Global.

The Literature Network (2022). *How much land does a man need?* Leo Tolstoy. https://www.online-literature.com/tolstoy/2738/

The TwiLex Group (2024). First language effects on incidental vocabulary learning through bimodal input: A multisite, preregistered, and close replication of Malone (2018). *Studies in Second Language Acquisition.* Advanced online publication. https://doi.org/10.1017/S0272263124000275

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words. *Language Learning*, *73*(1), 84–125. https://doi.org/10.1111/lang.12517

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163. http://hdl.handle.net/10125/66776

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68*(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *L*anguage Teaching Research, *19*(6), 667–686. https://doi.org/10.1177/1362168814559800

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Wisniewska, N., & Mora, J. (2018). Pronunciation learning through captioned videos. In J. Levis (Ed.), *Proceedings of the 9th annual pronunciation in second language learning and teaching conference* (pp. 204–215). Iowa State University.

Cite this article: Malone, J., & Hui, B. (2026). Audiovisual (a)synchrony: Reading ahead as a mechanism for L2 vocabulary processing and learning benefits of reading while listening. *Studies in Second Language Acquisition,* 48: 350–372. https://doi.org/10.1017/S0272263126101703

https://doi.org/10.1017/S0272263126101703 Published online by Cambridge University Press