



Linguistically-driven text formatting improves reading comprehension for ELLs and EL1s

Jack Dempsey^{1,5} · Kiel Christianson^{1,2} · Julie A. Van Dyke^{3,4,5}

Accepted: 15 April 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Typical print formatting provides no information regarding the linguistic features of a text, although texts vary considerably with respect to grammatical complexity and readability. Complex texts may be particularly challenging for individuals with weak language knowledge, such as English language learners. This paper investigates the usefulness of a text format referred to as Linguistically-Driven Text Formatting (LDTF), which provides visual cues to grammatical structure for in-the-moment language support during reading. We assessed reading comprehension in adult English Language Learners after a two-session exposure to the new format (also called Cascade Format). Participants' primary languages were Mandarin and Korean, which have substantially different syntactic structures from English. Ninety participants (30 L1 English, 30 L1 Mandarin, 30 L1 Korean) were randomly assigned to either the traditional or the LDTF format and read six English passages across two sessions within the same week. Comprehension was assessed via questions that probe sentence comprehension and global text properties. Participants also completed a TOEFL assessment, presented in either LDTF or traditional format. Bayesian analyses showed that the Cascade Format improved sentence comprehension relative to control participants for all language groups and experience levels. Effects on the TOEFL assessment, which taps inferencing and meta-linguistic skills, were not observed. Syntactic knowledge plays a fundamental role in reading comprehension, and LDTF appears to support comprehension by providing visual cues to this knowledge that can be used at the very moment of meaning construction.

This manuscript reports original, unpublished research conducted under the oversight of the University of Illinois Institutional Review Board. The study was judged to be Exempt and was not required to be registered as a clinical trial. Funding was provided by Cascade Reading, Inc. and Jack Dempsey and Julie Van Dyke both received compensation for their participation in the study. Kiel Christianson did not receive compensation for his participation in the study. Data from the study is available at <https://osf.io/7s5tb/>.

Extended author information available on the last page of the article

Keywords Text formatting · Reading comprehension · English language learners · Syntactic knowledge · Linguistically-driven text formatting

Introduction

Mastering reading comprehension remains a serious challenge for all learners, especially in the wake of a multi-year pandemic (e.g., Kuhfeld et al., 2023). For example, the National Center for Education Statistics released the National Assessment of Educational Progress (NAEP) Report Card (also known as the Nation's Report Card) for reading comprehension skills, showing that average reading scores for US twelfth grade students (and likewise for fourth and eighth grade students) were significantly lower from when the first reading assessment was administered back in 1992 (US Department of Education, 2019). As of the most recent assessment, taken in 2019 before the effects of Covid, only 37% of twelfth grade students performed at or above NAEP Proficient levels, quantifying a real crisis for graduating students. Low literacy presents a major obstacle to attaining educational and occupational goals, and has been associated with increased mental and physical health risks (Kutner, 2007; National Institute for Literacy, 2008). Even more troubling is that English Language Learners (ELLs) have consistently scored between thirty and fifty points *behind* their Native English-speaking (EL1) peers (for reference, the difference between NAEP Basic and NAEP Proficient is 38 points). Thus, quality of life challenges for ELLs living in a predominantly English-speaking society may be especially acute.

It has long been documented that ELLs fall behind their EL1 peers in reading comprehension in elementary school and that this gap tends to increase with each subsequent grade (e.g., Farnia & Geva, 2013). Studies have revealed a variety of causes for this gap, such as differences in vocabulary knowledge (Droop & Verhoeven, 2003), listening comprehension (Li et al., 2021), language-specific knowledge, and metalinguistic skills, including morphological and syntactic knowledge, syntactic awareness, and the ability to build a mental model of the text (Farnia & Geva, 2013; Guo et al., 2011; Li et al., 2021; Raudszus et al., 2021). Although there is no definitive research regarding which of these factors has the greatest impact in determining comprehension outcomes (Choi & Zhang, 2021), some studies suggest that syntactic knowledge, e.g., recognizing constituent boundaries and relations between them, may be particularly important in elementary grades (Farnia & Geva, 2013; Shiotsu & Weir, 2007) and into adulthood (Jeon & Yamashita, 2014, 2022; Zarei & Neya, 2014), just as it is for EL1 readers (Andrews & Veldre, 2021; Balthazar & Scott, 2023; Breen et al., 2016; Brimo et al., 2017; Groen et al., 2019; Mackay et al., 2021; Van Dyke, 2021; Van Dyke et al., 2014).

This is consistent with a long tradition in psycholinguistic and neurolinguistic research which points to syntactic processing as a core component of language *composition*—the process of joining individual words into meaningful phrases and sentences (e.g., Brennan & Pykkänen, 2012; Brothers & Traxler, 2016; Chen et al., 2021; Clifton et al., 2003; Flick & Pykkänen, 2020; Frazier & Rayner, 1982; Friederici, 2002, 2011; Mollica et al., 2020; Shain et al., 2022, *inter alia*). This evidence base suggests that increased syntactic knowledge, or more immediate accessibility

to a sentence's structure, is the foundation for constructing a mental model during reading. In the ELL context, this is a crucial point, as learners often have poor syntactic knowledge and processing ability in their non-native language. Multiple studies of adult ELLs and bilingual children have shown that syntactic knowledge predicts comprehension outcomes (e.g., Gottardo et al., 2018; Guo, 2008; Siu & Ho, 2020; Sohail et al., 2022; Zhang, 2012), and evidence also shows that syntactic knowledge in one language can improve reading comprehension and syntactic awareness¹ and knowledge in another language (Siu & Ho, 2020; Sohail et al., 2022).

Since syntactic knowledge is an important determinant of comprehension, it is critical to consider that there are many reported differences in how EL1s and ELLs process syntactic information during reading. One influential line of work in second language processing posits that syntactic reasoning is more difficult in L2 processing than in L1 processing, resulting in a heavier reliance on non-syntactic factors like semantics and pragmatics. This Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, b, c, 2018) is important to consider in conjunction with ELL reading comprehension outcomes since it claims that syntactic knowledge is often underutilized in favor of other linguistic cues. Alternatively, Good-Enough theories of language processing (e.g., Ferreira et al., 2002; Christianson, 2016) argue that these differences between L1 and L2 processing are quantitative in nature rather than qualitative, with L2 readers differing in their sensitivity to syntactic cues rather than in their ability to use them (Lim & Christianson, 2013a, b, 2015), given adequate proficiency. Regardless of whether this difference in processing is quantitative or qualitative, researchers agree that syntax is not as reliably accessible for L2 readers compared to L1 readers (Christianson, 2016; Lim & Christianson, 2013a, b, 2015; Clahsen & Felser, 2006a, b, c, 2018; Ferreira et al., 2002). Hence, methods of increasing and supporting syntactic processing ability may be especially effective for improving reading comprehension for ELLs, whose access and use of syntactic information may be less automatic. In sum, both EL1s and ELLs are expected to benefit from syntax-based reading interventions, with a potential for even larger benefits for ELLs, since such interventions may provide added support for less developed or less efficient syntactic processing.

Indeed, a number of educators have recently emphasized the importance of explicitly teaching grammatical structures to improve ELL reading (Alqahtani, 2019; Cai & Yao, 2022; Zipoli, 2017). This approach is synergistic with the long debate about whether second language instruction is better taught via methods that draw explicit attention to linguistic forms or rules versus methods that rely on implicit learning of structures presented in natural contexts (see Goo et al., 2015; Kang et al., 2019; Norris & Ortega, 2000; Spada & Tomita, 2010 for metaanalyses.) Across all these studies, explicit instruction emerged with consistently large effect sizes, usually Cohen's *d*

¹ Whereas syntactic knowledge can be defined as the ability to recognize and interpret syntactic grammatical structures during reading, syntactic awareness represents a more metalinguistic skill that allows readers to reflect upon and manipulate these structures (Tunmer et al. (1988)). Although both are predictors of comprehension, recent work by Brimo and colleagues (2017) show that syntactic awareness is only a significant predictor when mediated by syntactic knowledge, thus syntactic knowledge is likely the better predictor for more immediate comprehension abilities.

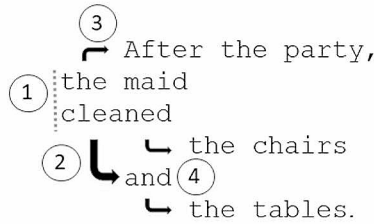
Hedges' $g > 1$,² with larger effect sizes compared with implicit instruction in all but one study (Kang et al., 2019). These authors attributed the discrepancy to the inclusion of 39 new studies that investigated implicit instruction, in response to heightened interest in this topic in more recent years. The Kang et al., meta-analysis is also notable in that it incorporated modern random-effects statistical methods to control for sample-size bias and investigate moderator variables on individual effect sizes (Borenstein et al., 2011; Shintani et al., 2013), therein producing a more nuanced analysis of summative results. For example, Kang et al. found that for short-term learning (i.e., immediate assessments) there was no difference between explicit and implicit instruction, while implicit instruction produced significantly larger effect-sizes for long term learning. Finally, when investigating the impact of specific types of linguistic instruction, their meta-analysis found that the largest effect size was found for instructing syntax ($g=0.94$), compared to instructing morphology ($g=0.85$) or pragmatics ($g=0.32$).

These results bode well for the relevance of linguistically-driven text formatting (LDTF; Van Dyke et al., 2021) for ELL learners, as the format presents visual cues to emphasize syntactic information while learners are reading naturally. Although we are not claiming to improve explicit syntactic knowledge via LDTF, we do believe that LDTF—also known as Cascade Reading—offers a novel approach to presenting form-focused (Nassaji, 2015) exposure to syntax during authentic reading. Specifically, the Cascade format utilizes line breaks to chunk syntactic constituents onto separate lines and indentations to cue dependency relations between them. A more detailed description of the algorithm can be found in Van Dyke and Dempsey (submitted), however the basic principles are briefly elaborated here. First, syntactic subjects and their heads are aligned with one another. Second, other dependent constituents are indented one level past their head, whether they precede or follow linearly. Lastly, members of conjunctions are indented one level past the conjoining term, which assumes a head constituent role. These rules allow participants to see a visual map of the syntactic structure in each sentence *as they are reading*—providing an unobtrusive scaffold from which readers can access (or develop) their syntactic knowledge while building interpretations. An example can be found in Fig. 1 below.

Although no previous research has investigated whether this type of alternative text formatting can improve syntactic knowledge in ELLs or ELIs, previous studies have shown that chunking based on syntactic constituents improves reading comprehension of ELIs (Graf & Torry, 1966; Levasseur et al., 2006; Tate et al., 2019; Walker et al., 2007). For example, Van Dyke and Dempsey (submitted) asked 4th and 5th grade ELIs to read passages in either a Cascade group or a control group before switching halfway through the semester. They observed larger growth in reading comprehension ability for students reading in the Cascade Format compared to those reading in the traditional text format. Moreover, students who benefitted most from this format were significantly more likely to self-report higher instances of implicit prosody (i.e., hearing how a sentence would sound were it to be read out loud.) This finding is consistent with other research demonstrating that prosodic processing and

² The lowest effect size of 0.73 was found by Spada and Tomita (2010) for explicit instruction of simple linguistic features.

Cascade Reading LDTF Principles



Principle 1: Subjects and verbs are aligned.

Principle 2: Dependents are indented under their head.

Principle 3: Introductory information is indented.

Principle 4: Elements of a conjunction are aligned and preceded by the conjoining term.

Fig. 1 Example of the Cascade Reading LDTF and its core principles

syntactic knowledge are tightly linked, particularly in students with lower reading comprehension ability (Breen, 2014; Breen et al., 2016, [in press](#)).

The current study seeks to determine whether the LDTF format can improve reading comprehension for adult EL1 and ELLs. While we believe that the mechanism for this improvement is due to increased access to syntactic knowledge due to form-focused cuing, at this time we do not assess syntactic knowledge. We discuss this further below under Limitations, although given the discrepancy in ELL reading scores summarized above, we believe that identifying interventions that can improve reading is a worthy goal in itself. Hence, the current research seeks to address the following two questions: (1) Can Cascade Reading's LDTF improve reading comprehension for adult EL1s? (2) And can it also improve, perhaps to a greater extent, reading comprehension for adult ELLs during reading?

Methods

Participants

We recruited adult EL1, L1 Mandarin ELL, and L1 Korean ELL readers at an American university. Korean and Mandarin ELLs were selected as participant groups for several strategic reasons. First, they are among the 10 most commonly reported home languages of ELLs in the United States (NCES, 2022). Second, they represent

typologically and orthographically different languages. Mandarin uses a logographic system where characters denote semantic and phonological information, whereas Korean uses an alphabet expressing phonological information and syllabic information via arrangement of the letters into “syllable blocks.” Perhaps more importantly, in terms of each language’s syntax, Korean is agglutinative, where meanings are often expressed in productive affixes (Koopman, 2005), whereas Mandarin is isolating, where meanings are often expressed in single, isolated morphemes (Huang & Liu, 2014). English, on the other hand, uses an alphabet, similar but distinct to Korean, and is an analytic language where meaning is expressed partially through productive affixes but largely through word order (Jespersen, 1984). Thus, Korean and Mandarin represent ELL populations who would likely have different sources of issues with English syntactic processing due to interference from their respective first language, which typically represents over one third of errors ELLs make (Pichette & Leśniewska, 2018). Finally, these population groups are readily available on the university campus where the research was conducted.

Participants were all at least 18 years of age and earned a total of \$80 for participation across all three sessions of the experiment. Participants in the EL1 group were required to have grown up speaking English as (one of) their first language(s). To be eligible to participate in one of the ELL groups, participants had to have grown up speaking either Mandarin or Korean and had to have not grown up speaking English in the home. Although education level was not a criterion for eligibility, all participants indicated at least a bachelor’s degree if not higher post-secondary degrees. Language experience and history were assessed by using an adapted LEAP-Q (Kaushanskaya et al., 2020; Marian et al., 2007). Thirty participants were initially recruited from each language group and were randomly assigned to either the Cascade format group or the control group, who read in a traditional text format. Some of these participants were not included in the final analysis (see [Results](#) section), resulting in 27 L1 English, 28 L1 Korean, and 30 L1 Mandarin participants in the final analysis. A post-hoc power analysis conducted in GPower (Erdfelder et al., 1996) shows that our design achieved 70.0% power for finding the main effect of Format (Odds Ratio=1.65, Probability of H0=0.5, Alpha=0.05, Total Sample Size=85).

Language experience differed in terms of exposure to English reading and age of acquisition between groups. The former finding revealed that English participants have a higher tendency to read in English, which is not very surprising. The age of acquisition difference shows that Mandarin participants reported an overall higher age of acquisition compared to Korean participants; however, this significance did not hold across format groups. To ensure that the age of acquisition imbalance was not driving differences in the data, we ran an additional model only including ELL participants, revealing that the Mandarin ELLs scored higher than Korean ELLs even when controlling for age of acquisition. That model can be found in the supplemental code, and this language-based difference is discussed in the general discussion. Differences also existed between groups for the individual differences observed in Session 1, outlined below; however, these differences were mostly expected given language requirements for these groups, and these variables were all added as fixed effects in the main models to control for their influence on the effects of interest.

Language experience and individual difference variables across the three groups are reported below in Table 1.

We administered tasks before the first reading session to collect information about various participant abilities, including English reading comprehension as assessed by the Gates-MacGinitie Reading Test (GMRT; MacGinitie et al., 2000) 4th Ed (Form T, Level AR), word-decoding ability as assessed by a spelling recognition task (Andrews et al., 2020) and oral reading fluency as assessed by a read-aloud task (Breen et al., *in press*) and scored according to the NAEP oral reading fluency scoring rubric (White et al., 2021). This latter measure was scored by two independent raters with high interrater reliability ($ICC [A,2]=0.996, p < .001$). Means and standard deviations on these measures by group are reported in Table 1. Due to the theoretical importance of these variables on reading comprehension success, GMRT scores, spelling recognition scores, and ORF scores were included as control predictors in all inferential models (see [Results](#) section).

Materials

For the two reading sessions, passages were selected and used with permission from ReadWorks.org. Passages were all over 1000 words in length and were evaluated as being suitable for the 12th grade reading level by ReadWorks. Comprehension assessments were administered as a combination of multiple choice and binary choice probes immediately following each passage. These included four multiple-choice questions taken from ReadWorks that probed explicitly mentioned information, the main topic of the passage, text structure and evidence, and inferences about what was read. We created an additional ten binary choice (yes or no) questions which were written to explicitly probe information conveyed by the syntactic relationships expressed within the texts (e.g., who did what to whom) on the sentence level. An example of these binary choice questions can be found in the supplemental materials.

In each of these sessions, participants also completed a TOEFL (Enright et al., 2000) practice test passage presented alongside ten comprehension questions, appearing in either traditional or Cascaded text depending on experimental group. These passages were taken from the Educational Testing Service's website and are freely available. Since these tests are publicly available, participants were also asked if they had ever read the passage. No participant indicated that they had seen the TOEFL passage prior to their session. These passages were included for a few reasons. First, although many of the questions in the TOEFL target higher-level, metalinguistic skills unrelated to sentence-level comprehension, it is possible that a scaffold to sentence comprehension could lead to better inferencing and metalinguistic understanding of passage content, which could then lead to increases in TOEFL scores. Second, we wanted to evaluate the effect of Cascade on an assessment that is of considerable interest for the ELL community.

Procedure

The study was administered across three separate sessions. Each session took place in a laboratory setting where the participant was seated at a computer within sight

Table 1 Demographic and reading ability means and standard deviations by language and format group. English exposure is a self-reported proportion of written material read in English versus another language and L1 dominance is a binary measure of whether participants view their L1 as their dominant tongue. GMRT scores are based on number of responses scored correct out of 48. Spelling scores are based on number of words correctly marked as correct or incorrect out of 88. ORF ratings are based on a derivation of the NAEP scoring rubric converted to a score from 1 to 5 with 5 indicating high proficiency. *Asterisks denote groups that significantly differ from other groups

Language/Format	Eng./Cascade	Eng./Traditional	Kor./Cascade	Kor./Traditional	Man./Cascade	Man./Traditional
<i>Number of Participants</i>	13	14	15	13	15	15
<i>Age in years</i>	21.58* (3.34)	21.1* (2.89)	29.07 (6.40)	26.31 (4.11)	25.06 (2.59)	26.53 (3.78)
<i>Age of acquisition</i>	-	-	9.79 (3.45)	7.85 (3.08)	7.41* (2.37)	6.47* (2.61)
<i>English Exposure</i>	0.90* (0.21)	0.90* (0.16)	0.49 (0.28)	0.31 (0.29)	0.30 (0.15)	0.43 (0.27)
<i>L1 dominance binary</i>	0.92 (0.29)	1.0 (0.0)	0.86 (0.36)	0.92 (0.28)	1.0 (0.0)	0.87 (0.35)
<i>Gates reading</i>	40.11 (4.86)	39.80 (5.81)	35.88* (4.20)	38.91* (6.17)	36.68* (4.68)	35.76* (5.80)
<i>Spelling recognition</i>	75.41 (6.66)	75.28 (7.01)	74.19* (5.30)	70.35* (6.72)	64.63* (6.66)	66.38* (8.26)
<i>TOEFL SCORE</i>	-	-	26.33 (2.79)	26.99 (3.13)	27.31* (2.51)	28.06* (1.73)
<i>Oral reading proficiency</i>	4.30 (0.55)	4.57 (0.32)	3.64* (0.71)	3.57* (0.61)	3.16* (0.68)	3.24* (0.63)

of the researcher. The researcher followed a written protocol to ensure participants were given the same instructions in the same order. The first session collected data on all the individual difference and language background measures described above, including the GMRT, a spelling recognition task, self-reported TOEFL questionnaire, LEAP-Q, and a read-aloud task. The second and third sessions of the study occurred within seven days of one another to ensure that the Cascade format was not forgotten between sessions, and these sessions never occurred on the same day. For each of the reading sessions, participants read three passages either in Cascade or traditional format, each followed by a series of fourteen comprehension questions. Passages and questions were presented using a private Ibex Farm (Drummond et al., 2016) server hosted online. Participants were allowed to take as much time as they needed to finish reading each passage, but they could not go back and read the passage once they viewed the comprehension questions. This process iterated through three passages until participants reached the TOEFL practice passage. Participants were limited to twenty minutes to read this TOEFL passage, presented in either the Cascade Format or the traditional format, and answer the ten questions about the reading. Unlike the first three passages, the TOEFL passage was presented alongside the comprehension questions, so participants were allowed to freely move back and forth between answering questions and reading the text.

Results

Data cleaning & analytic approach

Participants' data were removed prior to analysis if they did not complete one of the reading sessions, leading to a loss of three L1 English participants, two L1 Korean participants, and no L1 Mandarin participants. Inspection of passage reading times and question reading times revealed no discernable lower boundary cutoff that needed to be implemented since neither continuous measure showed a clear lower-end outlier. No data trimming was required due to longer reading times or responses since participants were given as much time as they needed to complete the passages and questions.

We use Bayesian inference in all our statistical models for this study. The Bayesian approach is an alternative, and now highly popular, approach for statistical inference to frequentism. Whereas frequentist analyses quantify the probability that the null hypothesis is false given the data, Bayesian analyses allow researchers to conduct direct model comparisons to quantify how well the data support a particular hypothesis (e.g., the null or the hypothesized model; Masson, 2011). The method has increased sensitivity because it incorporates a priori knowledge of an effect's size and shape. Even when specific effect sizes are not known, so-called mildly informative priors can be used to approximate general information like the shape of distributions (Gelman et al., 2008). These prior distributions are computed along with the data to derive a posterior distribution, or a distribution of posterior estimates, that is the result of a Markov chain Monte Carlo sampling method used to approximate a true effect size. Thus, rather than producing a *t* or *z* statistic, the output of a Bayesian model is

an entire distribution. From this distribution, rejection of the null hypothesis can be inferred by looking at so-called credible intervals (CrI) around the null estimate (i.e., 0). For example, an 89% CrI that does not contain zero indicates that 94.5% of the posterior estimates in the posterior distribution showed an effect in a given direction. Although not a decision criterion like a specific p -value in the frequentist framework, this is often likened to statistical significance.

There are several reasons we decided to use Bayesian inference in the current study. Since Bayesian inference does not use a significance cutoff such as a p -value, probability for a null hypothesis, if borne out, can be computed via Bayes Factors (for a review, see e.g., Schad et al., 2021; Wagenmakers et al., 2017). Importantly, p -values cannot be used as a measure of probabilistic magnitude (i.e., using p -values to compare significance of models), whereas such comparisons between models are possible in Bayesian statistics (Wagenmakers, 2007). Therefore, Bayesian inference offers a higher degree of analytic flexibility. Also, the same structure of models, like mixed effects or hierarchical models, can be used via the brms package (Bürkner et al., 2017) in R, just as they can be used within a frequentist framework. Moreover, models with both random and fixed effects are more optimized with Bayesian inference because the sampling procedure causes fewer convergence issues, meaning that researchers can fit maximal random effects models without issues caused by quasi-separation (Eager & Roy, 2017; Kimball et al., 2019). Finally, lower-powered samples, as this one, are less problematic when using Bayesian inference versus frequentist inference thanks to the former's magnitudinal probabilistic approach, although higher-powered samples still give more accurate estimates on true effect sizes.

We fit comprehension question accuracy separately for ReadWorks and TOEFL passages to Bayesian hierarchical models using the brms package in R Version 4.0.3 (R Core Team, 2020). These included fixed effects of Format (sum coded: Control = -0.5, Cascade = 0.5), Language (treatment coded with English baseline), and Session (sum coded: 1 = -0.5, 2 = 0.5), along with their interactions. Additional fixed effects of Gates MacGinitie Scores, Spelling, and ORF, all numeric, were added as controls. Random effects were maximal and included intercepts by Participant and by Question with slopes of Passage by Participant and Group, Language, and their interaction by Question. All models were fit to a Bernoulli distribution and included mildly informative priors specifying a wide range of a priori possible values (intercept = normal(0,1), beta = normal(0,1), sd = truncated normal (0,1)). Models were run on four chains for 7500 iterations, 2500 of which were warm-up. No models failed to converge, and all population-level effects had Rhat values of 1.00, indicating good fit. Following convention, model outputs were interpreted by using 89% credible intervals (CrIs), denoting that at least 94.5% of estimates from the posterior distribution indicate an effect in a given direction when 0 is not included in the interval.

Comprehension models

Means and standard deviations for ReadWorks passage accuracy and for TOEFL passage accuracy are reported below in Table 2. These are also visualized in Figs. 2 and 3, respectively. The model outputs for ReadWorks passage and TOEFL passage accuracy are reported in Tables 3 and 4, respectively, and only include the fixed effects that are part of interactions. The full model outputs are available in the supplemental code. For the TOEFL passage accuracy model, the only effect found was a simple effect showing that Mandarin ELLs answered these questions more accurately than EL1 participants (estimate=0.82, 89% CrI = [0.30, 1.34], as shown in Table 4. This effect was not mediated by the format in which participants read, suggesting this effect is indicative of an overall group difference in TOEFL experience. Overall, the Cascade Format and traditional format led to similar TOEFL accuracy results regardless of language group.

The analysis of ReadWorks passage accuracy revealed a main effect of Format such that the Cascade group performed better across languages and across sessions (estimate=0.50, 89% CrI = [0.06, 0.93]). There were also simple effects of Language showing that both Korean (estimate=0.35, 89% CrI = [0.02, 0.68]) and Mandarin (estimate=0.74, 89% CrI = [0.35, 1.14]) ELL groups scored higher than EL1 participants on average. This was qualified by an interaction with Session such that the Mandarin group did better overall, particularly in the third session, compared to the English group (estimate=0.64, 89% CrI = [0.24, 1.05]); however, this did not interact with Format, suggesting again that this group may just have more experience in this kind of testing environment.

These results suggest that Cascade Reading's LDTF led to higher accuracy for questions following ReadWorks passages, but this benefit did not significantly vary across language or session. This could be due to insufficient power for finding smaller interaction effects, so more work is needed to ascertain the exact degree to which different ELL populations enjoy a larger benefit. There also seems to be an advantage when answering questions for Mandarin speakers on both TOEFL and ReadWorks questions and for Korean speakers on ReadWorks questions. This may be due to an increased familiarity with English comprehension exams compared to L1 English speakers. Despite the lack of significance, it appears that effects were greater in the second reading session, so we quantified this benefit with separate models for the Session 2 data for each language group. This revealed a 2.6% format advantage for EL1s, 5.3% for Korean ELLs, and 11.1% for Mandarin ELLs. On the other hand, there was no evidence that Cascade Reading's LDTF led to higher accuracy for questions probing comprehension of the TOEFL passages. One potential reason for the benefit difference in readings is the type of questions that were asked: the questions following ReadWorks passages probed sentence-level syntactic relations, whereas the questions asked alongside the TOEFL passages probed many other aspects of comprehension, including inference, author intent, vocabulary, and paragraph structure. This difference is addressed further in the Discussion.

Table 2 Means and standard deviations for accuracy to all comprehension questions across language and format groups and reading sessions. S1 = Session 1; S2 = Session 2

Language/Format	Eng./Cascade	Eng./Traditional	Kor./Cascade	Kor./Traditional	Man./Cascade	Man./Traditional
<i>ReadWorks passage accuracy</i>	S1: 0.81 (0.39)	S1: 0.75 (0.44)	S1: 0.76 (0.43)	S1: 0.76 (0.43)	S1: 0.73 (0.45)	S1: 0.78 (0.42)
	S2: 0.85 (0.36)	S2: 0.74 (0.44)	S2: 0.86 (0.35)	S2: 0.77 (0.42)	S2: 0.88 (0.32)	S2: 0.79 (0.41)
<i>TOEFL passage accuracy</i>	S1: 0.56 (0.50)	S1: 0.52 (0.50)	S1: 0.52 (0.50)	S1: 0.56 (0.50)	S1: 0.52 (0.50)	S1: 0.50 (0.50)
	S2: 0.47 (0.50)	S2: 0.45 (0.50)	S2: 0.47 (0.50)	S2: 0.46 (0.50)	S2: 0.49 (0.50)	S2: 0.51 (0.50)

Discussion

In the current study, we examined reading comprehension outcomes for three groups differing in language background following the reading of texts presented in either traditional formatting or in Cascade's LDTF. Our central finding was that, regardless of language background, Cascade's LDTF improved comprehension on ReadWorks passages, as evidenced by accuracy to comprehension questions probing sentence-level interpretation of the texts. Moreover, ELLs appeared to receive an even greater advantage compared to EL1 readers, although additional research comparing language groups is required.

We argue that the benefit conferred by reading in the LDTF is due to the scaffolding of syntactic knowledge during meaning construction. Whereas syntactic knowledge is a latent ability that a reader possesses, the visual syntactic cues in the LDTF makes accessing syntactic structures easier compared to reading in traditional text format, which provides no guidance regarding relationships within the text. Syntactic knowledge guides a readers' understanding of "who did what to whom," allowing them to more efficiently and accurately build mental models while reading. We believe that the LDTF works in a similar fashion by explicitly laying bare the relationships between entities in a sentence, as communicated by its syntax. We acknowledge that our decision not to directly assess syntactic knowledge makes it impossible to verify our argument, however even if LDTF is not improving access to a readers' general syntactic knowledge, the positive result found here demonstrates that explicit demarcation of syntactic relationships leads to improved comprehension. This is consistent with a wealth of evidence that suggests that syntactic processing provides the foundation for sentence comprehension in EL1s (e.g., Frazier & Rayner, 1982; Christianson et al., 2001; McElree & Griffith, 1995; Van Dyke & McElree, 2011). Moreover, those with poor comprehension also display deficits in syntactic knowledge (Adlof & Catts, 2015; Bowey, 1986; Campanelli et al., 2023) and have a reduced ability to signal linguistic constituents during oral reading (Breen et al., [in press](#)). Future work should focus on directly testing effects of LDTF on syntactic knowledge.

The finding that ELL reading comprehension was improved by LDTF suggests that any impoverished use of syntactic cues—as suggested by both the Shallow Structure and Good-Enough Processing hypotheses—does not obviate the benefits of LDTF during reading; in fact, the opposite is more likely to be true. This result is promising even when we consider that the ELLs in this study were highly proficient, most of them being students who use English on a daily basis. That is, even though their syntactic knowledge was presumably quite advanced, they nevertheless derived additional benefit from the visual cuing provided by the novel format. Given our findings that the LDTF improves comprehension for less experienced elementary-aged EL1 readers (Van Dyke and Dempsey, [submitted](#)), we believe this format also holds promise for teaching English as a foreign or second language to younger learners.

The current results provide a hint that the format may be especially helpful for those whose L1 has substantially different syntactic properties compared to English; however, further research is necessary to specify any interactions stemming from known languages and varying levels of proficiency. For example, Mandarin is an isolating language whereas Korean is agglutinative, putting both languages at polar

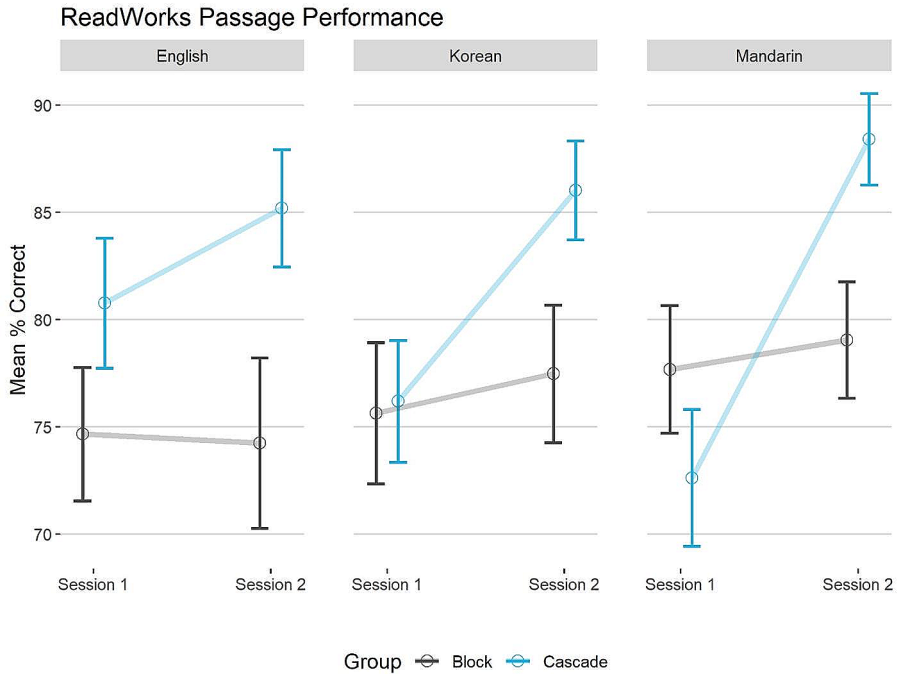


Fig. 2 Accuracy to questions follow ReadWorks passages across reading sessions and across language and format groups

extremes in terms of average number of morphemes expressed in a single word. It could be the case that Mandarin ELLs could make better or more immediate use of the Cascade LDTF since they are more accustomed to multi-word expressions and their relations with one another compared to Korean ELLs. The Mandarin ELLs also learned English at a younger age compared to Korean ELLs; however, Mandarin ELLs still outperformed Korean ELLs when controlling for age of acquisition. More research is needed to determine whether a possible interaction between first language typology and age of acquisition exists in terms of predicting LDTF's success for supporting reading in these populations.

The lack of effects for TOEFL passages was disappointing, but not wholly unexpected given the fact that many TOEFL questions required higher-level rhetorical reasoning that Cascade's LDTF is not designed to scaffold. In contrast, the binary-choice questions associated with the ReadWorks passages were designed to probe sentence-level interpretations of the text—meaning that every question had a corresponding sentence in the passage that contained the information needed to answer the question correctly. This means that the ReadWorks assessments were more closely calibrated to the syntactic processing that the LDTF is designed to support. To be clear, it is not that the TOEFL materials are more complex in terms of text readability; rather, the comprehension questions themselves may require higher-level metalinguistic reasoning abilities than sentence-level syntactic cues alone can address.

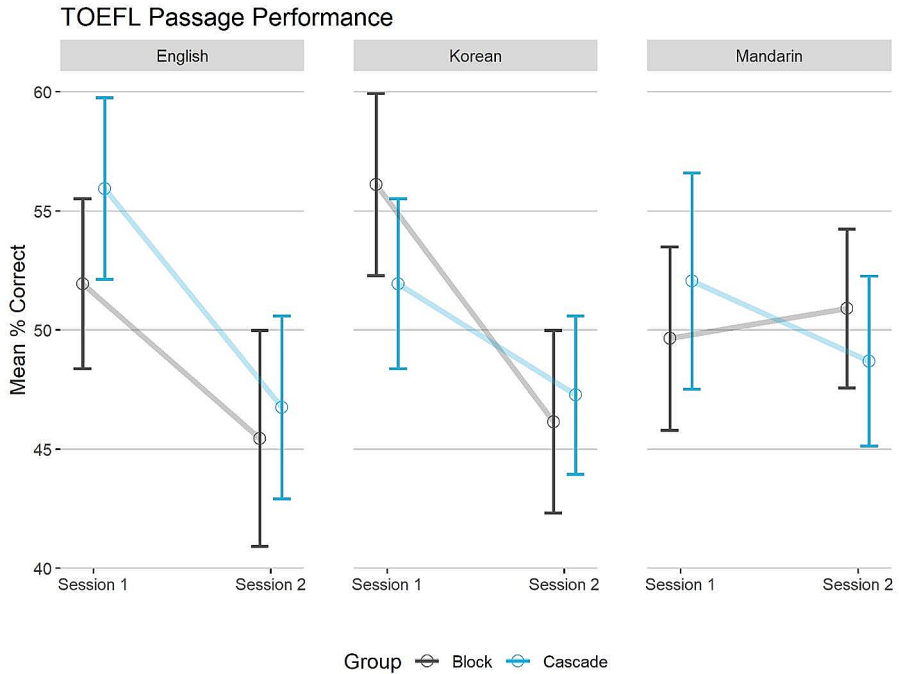


Fig. 3 Accuracy to questions presented alongside TOEFL passages across reading sessions and across language and format groups

Table 3 Model outputs for ReadWorks passage question accuracy. Asterisks denote credible intervals that do not cross 0

Effect	Estimate	SE	89% CrI
Format	0.50	0.27	[0.06, 0.93]*
Korean vs. English	0.35	0.21	[0.02, 0.68]*
Mandarin vs. English	0.74	0.25	[0.35, 1.14]*
Session	0.22	0.35	[-0.34, 0.77]
Format*Korean	-0.15	0.37	[-0.74, 0.44]
Format*Mandarin	-0.22	0.34	[-0.77, 0.32]
Format*Session	0.48	0.36	[-0.10, 1.05]
Korean*Session	0.27	0.22	[-0.07, 0.62]
Mandarin*Session	0.64	0.25	[0.24, 1.05]*
Format*Korean*Session	0.19	0.40	[-0.46, 0.83]
Format*Mandarin*Session	0.49	0.41	[-0.17, 1.13]

Another important difference between the two assessments is the fact that the TOEFL assessment allowed participants to go back and forth between passage and questions, while this was not possible for ReadWorks questions. Hence, the TOEFL assessment enabled a more general problem-solving approach where students could consult the text to identify answers regardless of the comprehension they attained during initial reading. This makes this assessment less related to reading comprehension per se, and therefore the benefit of LDTF may be more difficult to measure. In

Table 4 Model outputs for TOEFL passage question accuracy. Asterisks denote credible intervals that do not cross 0

Effect	Estimate	SE	89% CrI
Format	0.06	0.31	[-0.43, 0.54]
Korean vs. English	0.21	0.29	[-0.25, 0.66]
Mandarin vs. English	0.82	0.32	[0.30, 1.34]*
Session	-0.21	0.66	[-1.26, 0.86]
Format*Korean	-0.10	0.47	[-0.85, 0.64]
Format*Mandarin	-0.03	0.43	[-0.72, 0.66]
Format*Session	-0.12	0.45	[-0.83, 0.58]
Korean*Session	0.07	0.42	[-0.60, 0.74]
Mandarin*Session	0.17	0.41	[-0.50, 0.82]
Format*Korean*Session	0.36	0.64	[-0.67, 1.40]
Format*Mandarin*Session	-0.13	0.62	[-1.12, 0.87]

contrast, the ReadWorks assessment required students to read the text for comprehension and retrieve information from the mental model they created during reading—they could not look back into the text to find the answers after initial reading. Hence, the significant effect of the LDTF suggests not only that overall comprehension is improved, but also that it may facilitate memory for the specific text that was read.

In sum, Cascade Reading provides a promising new format for supporting reading comprehension for both EL1 and ELL adult students. It is entirely unique in offering visual cues that provide in-the-moment guidance about how to construct a mental model, including relationships between characters and concepts in a text, and appears to extend a particular benefit when readers are tested for their memory of what was read. Taken with previous research showing improved comprehension in elementary learners, Cascade Reading is poised to become an important pedagogical tool to improve English comprehension for all learners.

Limitations

The most obvious limitation in the current study is that we did not evaluate whether reading in Cascade has a direct effect on syntactic knowledge. As discussed above, implicit methods of instructing linguistic forms have proven successful, especially for syntactic knowledge (Kang et al., 2019). Hence, it would be useful to determine whether reading Cascaded text can improve linguistic competence as well as reading comprehension. In our estimation, the brief duration of the current study makes it unlikely that syntactic knowledge could have been generally improved, especially since we did not focus on any specific linguistic forms in our intervention, but instead used naturally occurring texts that contained a mixture of forms. A study seeking to produce a syntactic benefit would likely require greater exposure to Cascade format than provided here, and was therefore outside of the scope of the current project. In addition, the participants in the current study already had relatively high language proficiency (i.e., sufficient for admission to an American university), so the specific effect on syntactic competency could be difficult to assess. It should be emphasized, however, that improving reading comprehension is itself a valuable outcome, even if the exact mechanism of action is unknown. The fact that LDTF could have an effect—even for advanced ELLs—is especially promising, and future research will determine the role it may have in ESL pedagogy.

A further limitation is the focus on only two ELL groups with modest sample sizes. Future research should explore other native languages to better understand whether LDTF differentially impacts learners with particular linguistic backgrounds. Although issues with power are somewhat mitigated with our Bayesian approach, higher-powered samples and replications or corroborations would help solidify the current findings and grow our certainty for LDTF-driven improvements in reading comprehension. Longer-duration studies, as well as studies that keep track of a greater number of individual differences, would also be beneficial, as consistent exposure to this format could help higher-order comprehension skills improve over time, eventually supporting the metalinguistic reasoning needed to do well on high-stakes assessments like the TOEFL.

Conclusions

The patented LDTF, developed by Cascade Reading, uses line breaks and indentations to chunk syntactic constituents and show their relationships to one another. This system is hypothesized to work by aiding in the accessibility of syntactic structures *during* reading. The current study showed this to be true for EL1, Korean ELLs, and Mandarin ELLs while reading expository texts, as evidenced by higher accuracy rates to comprehension questions following the readings. This improvement in reading comprehension was greater for both ELL groups and greatest for Mandarin ELLs. Future work is needed to explore the generalizability of these reading benefits across different populations and reading contexts.

Data availability All data and analyses are publicly available on OSF: <https://osf.io/7s5tb/>

Declarations

Conflict of interest The funding for this study was provided by Cascade Reading, Inc., and two of the three authors involved in this research received compensation for their part in the study.

References

- Adlof, S. M., & Catts, H. W. (2015). Morphosyntax in poor comprehenders. *Reading and Writing*, 28, 1051–1070. <https://doi.org/10.1007/s11145-015-9562-3>
- Alqahtani, F. (2019). Investigating the Effect of structure complexity on students' recognition of the subject slot. *Online Submission*, 2(6), 77–82. <https://files.eric.ed.gov/fulltext/ED602312.pdf>
- Andrews, S., & Veldre, A. (2021). Wrapping up sentence comprehension: The role of task demands and individual differences. *Scientific Studies of Reading*, 25(2), 123–140. <https://doi.org/10.1080/1088438.2020.1817028>
- Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring lexical quality: The role of spelling ability. *Behavior Research Methods*, 52, 2257–2282. <https://doi.org/10.3758/s13428-020-01387-3>
- Balthazar, C. H., & Scott, C. M. (2023). Sentences are key: Helping school-age children and adolescents build sentence skills needed for real language. *American Journal of Speech-Language Pathology*, 1–16. https://doi.org/10.1044/2023_AJSLP-23-00038

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to metaanalysis*. Wiley.
- Bowey, J. A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology*, 41(2), 282–299. [https://doi.org/10.1016/0022-0965\(86\)90041-X](https://doi.org/10.1016/0022-0965(86)90041-X)
- Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2), 37–50. <https://doi.org/10.1111/lnc3.12061>
- Breen, M., Kaswer, L., Van Dyke, J. A., Krivokapić, J., & Landi, N. (2016). Imitated prosodic fluency predicts reading comprehension ability in good and poor high school readers. *Frontiers in Psychology*, 7, 1026. <https://doi.org/10.3389/fpsyg.2016.01026>
- Breen, M., Van Dyke, J. A., Krivokapić, J., & Landi, N. (in press). Prosodic features in production reflect reading comprehension skill in high school students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Brennan, J., & Pyllkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage*, 60(2), 1139–1148. <https://doi.org/10.1016/j.neuroimage.2012.01.030>
- Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, 40(1), 57–74. <https://doi.org/10.1111/1467-9817.12050>
- Brothers, T., & Traxler, M. J. (2016). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, 42(12), 1894. <https://doi.org/10.1037/xlm0000257>
- Bürkner, P. (2017). Brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28.
- Cai, Y., & Yao, X. (2022). Teaching complex sentences in ESL Reading: Structural analysis. *International Journal of English Linguistics*, 12(5). <https://doi.org/10.5539/ijel.v12n5p59>
- Campanelli, L., Van Dyke, J. A., & Landi, N. (2023). Semantic and syntactic processing in poor comprehenders: Evidence from eye-tracking and computational modeling. Paper presented at the Convention of the American Speech-Language-Hearing Association, Boston, MA, November 16–18.
- Chen, L., Goucha, T., Männel, C., Friederici, A. D., & Zaccarella, E. (2021). Hierarchical syntactic processing is beyond mere associating: Functional magnetic resonance imaging evidence from a novel artificial grammar. *Human Brain Mapping*, 42(10), 3253–3268. <https://doi.org/10.1002/hbm.25432>
- Choi, Y., & Zhang, D. (2021). The relative role of vocabulary and grammatical knowledge in L2 reading comprehension: A systematic review of literature. *International Review of Applied Linguistics in Language Teaching*, 59(1), 1–30. <https://doi.org/10.1515/iral-2017-0033>
- Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 817–828. <https://doi.org/10.1080/17470218.2015.1134603>
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407.
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3–42. <https://doi.org/10.1017/S0142716406060024>
- Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(1), 107–126. <https://doi.org/10.1017/S0142716406060206>
- Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10(12), 564–570. <https://doi.org/10.1016/j.tics.2006.10.002>
- Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. <https://doi.org/10.1017/S0272263117000250>
- Clifton Jr, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49(3), 317–334. [https://doi.org/10.1016/S0749-596X\(03\)00070-6](https://doi.org/10.1016/S0749-596X(03)00070-6)
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2019 Reading Assessment <https://www.nation-reporthcard.gov/reading/?grade=12>
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly*, 38(1), 78–103. <https://doi.org/10.1598/RRQ.38.1.4>
- Drummond, A., von der Malsburg, T., Erlewine, M. Y., & Vafaie, M. (2016). Ibex Farm. <https://github.com/addrummond/ibex>

- Eager, C. D., & Roy, J. (2017). Mixed models are sometimes terrible. *Linguistic Society of America, Austin, TX*<https://publish.illinois.edu/quantitativelinguistics/files/2017/01/LSA2017.Mixed-Models-are-Sometimes-Terrible-final.pdf>
- Enright, M., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework*. Educational Testing Service.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods Instruments & Computers*, 28, 1–11. <https://doi.org/10.3758/BF03203630>
- Farnia, F., & Geva, E. (2013). Growth and predictors of change in English language learners' reading comprehension. *Journal of Research in Reading*, 36(4), 389–421. <https://doi.org/10.1111/jrir.12003>
- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>
- Flick, G., & Pykkänen, L. (2020). Isolating syntax in natural language: MEG evidence for an early contribution of left posterior temporal cortex. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 127, 42–57. <https://doi.org/10.1016/j.cortex.2020.01.025>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84. [https://doi.org/10.1016/S1364-6613\(00\)01839-8](https://doi.org/10.1016/S1364-6613(00)01839-8)
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. (2008). A default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, 2(4), 1360–1383. <https://www.jstor.org/stable/30245139>
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–483). John Benjamins.
- Gottardo, A., Mirza, A., Koh, P. W., Ferreira, A., & Javier, C. (2018). Unpacking listening comprehension: The role of vocabulary, morphological awareness, and syntactic knowledge in reading comprehension. *Reading and Writing*, 31, 1741–1764. <https://doi.org/10.1007/s11145-017-9736-2>. <https://link.springer.com/article/>
- Graf, R., & Torrey, J. W. (1966). Perception of phrase structure in written language. In *American Psychological Association Convention Proceedings* (Vol. 83, p. 84).
- Groen, M. A., Veenendaal, N. J., & Verhoeven, L. (2019). The role of prosody in reading comprehension: Evidence from poor comprehenders. *Journal of Research in Reading*, 42(1), 37–57. <https://doi.org/10.1111/1467-9817.12133>
- Guo, Y. (2008). *The role of vocabulary knowledge, syntactic awareness and metacognitive awareness in reading comprehension of adult English language learners*. The Florida State University. [https://www.proquest.com/docview/250788062?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations & Theses](https://www.proquest.com/docview/250788062?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20and%20Theses)
- Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: Is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, 43(2), 159–183. <https://doi.org/10.1177/1086296X11403086>
- Huang, J. C. T., & Liu, N. (2014). A new passive form in Mandarin: Its syntax and implications. *International Journal of Chinese Linguistics*, 1(1), 1–34.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160–212. <https://doi.org/10.1111/lang.12034>
- Jeon, E. H., & Yamashita, J. (2022). L2 reading comprehension and its correlates. In *Bilingual processing and acquisition (BPA). Understanding l2 proficiency: Theoretical and meta-analytic investigations* (pp. 29–86). John Benjamins Publishing Company. <https://www.jbe-platform.com/docserver/fulltext/9789027257697.pdf#page=42>
- Jespersen, O. (1984). *Analytic syntax*. University of Chicago Press.
- Kang, E. Y., Sok, S., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*, 23(4), 428–453. <https://doi.org/10.1177/1362168818776671>
- Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The language experience and proficiency questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, 23(5), 945–950. <https://doi.org/10.1017/S1366728919000038>

- Kimball, A. E., Shantz, K., Eager, C., & Roy, J. (2019). Confronting quasi-separation in logistic mixed effects for linguistic data: A bayesian approach. *Journal of Quantitative Linguistics*, 26(3), 231–255. <https://doi.org/10.1080/09296174.2018.1499457>
- Koopman, H. (2005). Korean (and Japanese) morphology from a syntactic perspective. *Linguistic Inquiry*, 36(4), 601–633.
- Kuhfeld, M., Lewis, K., & Peltier, T. (2023). Reading achievement declines during the COVID-19 pandemic: Evidence from 5 million U.S. students in grades 3–8. *Reading and Writing*, 36, 245–261. <https://doi.org/10.1007/s11145-022-10345-8>
- Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y-C., & Dunleavy, E. (2007). *Literacy in everyday life: Results from the 2003 national assessment of adult literacy*. Institute of Educational Sciences, U.S. Department of Education. <https://files.eric.ed.gov/fulltext/ED495996.pdf>
- Levasseur, V. M., Macaruso, P., Palumbo, L. C., & Shankweiler, D. (2006). Syntactically cued text facilitates oral reading fluency in developing readers. *Applied Psycholinguistics*, 27(3), 423–445. <https://doi.org/10.1017/S0142716406060346>
- Li, L., Zhu, D., & Wu, X. (2021). The effects of vocabulary breadth and depth on reading comprehension in middle childhood: The mediator role of listening comprehension. *Reading & Writing Quarterly*, 37(4), 336–347. <https://doi.org/10.1080/10573569.2020.1809585>
- Lim, J. H., & Christianson, K. (2013a). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, 16(3), 518–537. <https://doi.org/10.1017/S1366728912000351>
- Lim, J. H., & Christianson, K. (2013b). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Research*, 29(3), 233–256. <https://doi.org/10.1177/0267658312462019>
- Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(6), 1283–1315. <https://doi.org/10.1017/S0142716414000290>
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading tests—Fourth Edition*. Riverside Publishing.
- Mackay, E., Lynch, E., Duncan, S. T., & Deacon, S. H. (2021). Informing the science of reading: Students' awareness of sentence-level information is important for reading comprehension. *Reading Research Quarterly*, 56, S221–S230. <https://doi.org/10.1002/rrq.397>
- Marian, V., Blumenfeld, H. K., & 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\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Masson, M. E. (2011). A tutorial on a practical bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43, 679–690. <https://doi.org/10.3758/s13428-010-0049-5>
- McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 134.
- McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning Memory and Cognition*, 24(2), 432. <https://doi.org/10.1037/0278-7393.24.2.432>
- Mollica, F., Siegelman, M., Diachek, E., Piantadosi, S. T., Mineroff, Z., Futrell, R., & Fedorenko, E. (2020). Composition is the core driver of the language-selective network. *Neurobiology of Language*, 1(1), 104–134. https://doi.org/10.1162/nol_a_00005
- National Center for Education Statistics (2022). State nonfiscal survey of public elementary and secondary education, 2008–2021, EDFacts file 141, Data Group 678, Office of Special Education Programs, IDEA database.
- National Institute for Literacy. (2008). *Developing early literacy: Report of the national early literacy panel*. National Institute for Literacy. <https://files.eric.ed.gov/fulltext/ED508381.pdf>
- Nassaji, H. (2015). The scope of language teaching studies. *Language Teaching Research*, 19(4), 394–396.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, 50(3), 417–528.
- Pichette, F., & Leśniewska, J. (2018). Percentage of L1-based errors in ESL: An update on Ellis (1985). *International Journal of Language Studies*, 12(2), 1–16. <http://www.ijls.net/pages/volume/vol12no2.html>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

- Raudszus, H., Segers, E., & Verhoeven, L. (2021). Patterns and predictors of reading comprehension growth in first and second language readers. *Journal of Research in Reading*, 44(2), 400–417. <https://doi.org/10.1111/1467-9817.12347>
- Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled bayesian workflow in cognitive science. *Psychological Methods*, 26(1), 103. <https://doi.org/10.1037/met0000275>
- Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., & Schuler, W. (2022). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *Journal of Neuroscience*, 42(39), 7412–7430. <https://doi.org/10.1523/JNEUROSCI.1894-21.2022>
- Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, 63, 296–329. <https://doi.org/10.1111/lang.12001>
- Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24(1), 99–128. <https://doi.org/10.1177/0265532207071513>
- Siu, T. S. C., & Ho, S. H. C. (2020). A longitudinal investigation of syntactic awareness and reading comprehension in chinese-english bilingual children. *Learning and Instruction*, 67, 101327. <https://doi.org/10.1016/j.learninstruc.2020.101327>
- Sohail, J., Sorenson Duncan, T., Koh, P. W., et al. (2022). How syntactic awareness might influence reading comprehension in English–French bilingual children. *Reading and Writing*, 35, 1289–1313. <https://doi.org/10.1007/s11145-021-10245-3>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60, 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Tate, T. P., Collins, P., Xu, Y., Yau, J. C., Krishnan, J., Prado, Y., & Warschauer, M. (2019). Visual-syntactic text format: Improving adolescent literacy. *Scientific Studies of Reading*, 23(4), 287–304. <https://doi.org/10.1080/10888438.2018.1561700>
- Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading research quarterly*, 134–158.
- Van Dyke, J. A. (2021). Introduction to the special issue: Mechanisms of variation in reading comprehension: Processes and products. *Scientific Studies of Reading*, 25(2), 93–103. <https://doi.org/10.1080/10888438.2021.1873347>
- Van Dyke, J. A., & Dempsey, J. (submitted). Linguistically-driven text formatting improves reading comprehension: Evidence from 4th and 5th graders.
- Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65(3), 247–263. <https://doi.org/10.1016/j.jml.2011.05.002>
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131(3), 373–403. <https://doi.org/10.1016/j.cognition.2014.01.007>
- Van Dyke, J. A., Gorman, M., & Lacey, M. (2021). Linguistically-driven automated text formatting (U.S. Patent No. 11,170,154 B1). U.S. Patent and Trademark Office. <https://imagepubs.uspto.gov/dirsearch-public/print/downloadPdf/11170154>
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>
- Wagenmakers, E. J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 123–138. <https://doi.org/10.1002/9781119095910.ch8>
- Walker, R. C., Gordon, A. S., Schloss, P., Fletcher, C. R., Voge, C. A., & Walker, S. (2007). Visual-Syntactic Text Formatting: Theoretical Basis and Empirical Evidence for Impact on Human Reading, *IEEE International Professional Communication Conference*, Seattle, WA, pp. 1–14, <https://doi.org/10.1109/IPCC.2007.4464068>
- White, S., Sabatini, J., Park, B. J., Chen, J., Bernstein, J., & Li, M. (2021). The 2018 NAEP oral reading fluency study. *NCES*. <https://files.eric.ed.gov/fulltext/ED612205.pdf>
- Zarei, A. A., & Neyfa, S. S. (2014). The effect of vocabulary, syntax, and discourse-oriented activities on short and long-term L2 reading comprehension. *International Journal of Language & Linguistics*, 1(1), 29–39. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8747d9490ba5115eac6afa90a1a60250aa952373>

Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. *The Modern Language Journal*, 96(4), 558–575. <https://doi.org/10.1111/j.1540-4781.2012.01398.x>

Zipoli Jr, R. P. (2017). Unraveling difficult sentences: Strategies to support reading comprehension. *Intervention in School and Clinic*, 52(4), 218–227. <https://doi.org/10.1177/1053451216659465>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Jack Dempsey^{1,5}  · Kiel Christianson^{1,2} · Julie A. Van Dyke^{3,4,5}

✉ Jack Dempsey
jkdemps2@illinois.edu

¹ University of Illinois at Urbana-Champaign, Champaign, USA

² Beckman Institute for Advanced Science and Technology, Urbana, USA

³ Child Study Center, Yale University School of Medicine, New Haven, USA

⁴ Institute for Brain and Cognitive Sciences, University of Connecticut, Storrs, USA

⁵ Cascade Reading, Inc, Minneapolis, USA