



Ángel Garralda Ortega*, Abel Hon Man Cheung and
Michelle Yuen Shan Fong

Developing e-reading pedagogies informed by research

<https://doi.org/10.1515/jwl-2021-0024>

Received October 25, 2021; accepted June 15, 2022; published online August 18, 2022

Abstract: This paper explores whether the reading comprehension of complex texts can be facilitated through an online reading platform designed for novice readers with English proficiencies below the Cognitive Academic Language Proficiency (CALP) threshold. We hypothesize that computer-mediated text glosses could speed up lower-level processing in the readers' working memory and thus enhance the overall comprehension of complex texts for study purposes. We tested 46 participants with estimated International English Language Testing System (IELTS) reading scores between 5 and 5.5, sampled from a pool of 1,406 students who took a diagnostic reading test and 631 students who completed a survey on their reading practices. Our participants were randomly assigned to read one General Reading IELTS text and one Academic Reading IELTS text, either on an on-screen Word file or on the e-reading platform with the glossing tool. The tests were video-recorded and the participants completed post-test interviews for further qualitative analysis. While the Mixed-model ANOVA did not suggest an interaction effect between the two language proficiency categories and the mode in which the tests were administered, it revealed a main effect on the use of online reading ($p < 0.01$) across the 5–5.5 IELTS spectrum, indicating that the electronic glosses enhanced reading comprehension. Implications for further research and pedagogy are discussed.

Keywords: design-based research; electronic glossing; e-reading instruction

*Corresponding author: Ángel Garralda Ortega, Centre for English and Additional Languages, Lingnan University, Hong Kong, China, E-mail: agarralda@ln.edu.hk

Abel Hon Man Cheung, School of General Education and Languages, Technological and Higher Education Institute of Hongkong, Hong Kong, China, E-mail: abelcheung13@thei.edu.hk

Michelle Yuen Shan Fong, Language Centre, Hong Kong Baptist University, Hong Kong, China, E-mail: michellef@hkbu.edu.hk

1 Introduction

A major challenge universities face in Hong Kong comes from the fact that students often lack the proficiency to benefit from English-medium education fully. Back in 2003, the *Final Report of the Language Education Review* by the Standing Committee on Language Education and Research (SCOLAR) already concluded that if Hong Kong aligned the local minimum requirement for university admission with international standards, not enough students might qualify for the publicly funded places at Hong Kong universities each year (SCOLAR 2003). English-medium instruction at degree level requires, among other things, the ability to handle advanced texts, summarize and synthesize complex information and use that information in constructing complex abstract arguments following specific disciplinary conventions. It has been proven that Cognitive Academic Language Proficiency (CALP), or the linguistic competence to employ language in highly abstract and decontextualized contexts, requires an upper threshold of linguistic competence (Cummins 1976, 1983). This language threshold could be estimated somewhere between levels 4 and 5 of the Hong Kong's Diploma in Secondary Education (overall IELTS equivalent band scores between 6.31 and 6.51 for level 4 and 6.81–6.99 for level 5), judging from the official grade descriptors and samples of students' work (Garralda Ortega 2018). It has also been suggested that CALP may not necessarily result from being exposed to language in everyday non-academic contexts (Ellis 1994). If these assumptions prove correct, the kind of pedagogical intervention required for academic success at university should combine an extensive use of English in informal everyday contexts with adequately scaffolded exposure to language use in academic settings.

If language standards are already problematic across publicly funded programs in Hong Kong, the challenge to teach English for academic purposes to students across self-funded tertiary institutions, which typically attract students with lower language proficiency, is even more pressing. Reading can be a logical starting point in this endeavor, as high second language (L2) reading proficiency and effective skills constitute a basic pillar not only for academic success (Grabe 2014; Hermida 2009) but also for language learning, a process which naturally starts with comprehensible input (Krashen 2004). Effective reading skills are paramount in academic settings. They propitiate the effective processing of information needed for knowledge construction in tasks such as academic essays and final year project work. Reading also helps learners acquire a critical mass of vocabulary in a process which nevertheless may require explicit instruction (Roessingh et al. 2005).

Arguably, this explicit instruction can be best implemented in technology-enhanced self-access learning contexts followed by small group face-to-face discussion rather than during whole-group class-time. This would enable learners with differing levels of proficiency and reading skills to engage in reading at a self-paced

and more controlled fashion before subsequent discussion with a teacher takes place. Class-time could then be best employed for meaningful group interaction between teachers and students aimed at joint knowledge construction, where reading tasks can be integrated with speaking or writing activities (Hirvela 2016).

The present study focuses on the use of technology in reading instruction, in the context of an academic reading program for undergraduates at a self-funded institution in Hong Kong. This pedagogical intervention was centered around an online reading platform, purposely designed to cater for novice readers with English proficiency below the CALP threshold. The platform was piloted during the course of the program to explore the extent to which the comprehension of academic texts could be enhanced by providing appropriate cognitive and linguistic scaffolding during the reading process.

Pedagogical interventions of this kind can be best studied by employing design-based research, a relatively novel approach in educational research whose primary aim is to “increase the impact, transfer, and translation of education research into improved practice” while stressing “the need for theory building and development of design principles that guide, inform and improve both practice and research in educational contexts” (Anderson and Shattuck 2012: 16). Design-based research suits technology-enhanced learning innovations like the one discussed in this paper because it offers the potential for “advancing design, research and practice concurrently” (Wang and Hannafin 2005: 5), thus helping breach a pervasive gap between research and practice often found in education (Reeves 2005; Vanderlinde and van Braak 2010).

However, this blending of context-specific empirical educational research with the theory-driven design of innovative learning environments, characteristic of design-based research, may often lead to contradictions in research agendas and to methodological shortcomings (Anderson and Shattuck 2012; Collins et al. 2004; Dede 2005; Kelly 2004). The former are the result of the dual role of designer and researcher being adopted. The latter are often associated with the need to tackle complex real-world situations, which may undermine the empirical validity of design-based research or limit the applicability of its findings to broader educational contexts.

All these contradictions, rooted in the need to achieve an adequate balance between theory and practice in our pedagogical intervention, require adopting a hybrid methodology, flexible and problem-oriented, where the distinction between designers, teachers, and researchers can often be blurred. This methodology will be discussed in Section 4 below, after the learning context which motivated our pedagogical intervention is explained and our research objectives are formulated.

2 The online reading platform

The online reading platform used in our study includes the following features: First and foremost, there is a tooltip glossing device for difficult lexis (see Lee and Lee 2013 for a full description). This can be activated by pointing at a hyperlink on the computer screen with the cursor. Learners can choose to look up words during reading or simply ignore the glosses if they do not encounter any comprehension problems. Each gloss typically includes a Chinese translation of the English term as well as the kind of information found in the *Collins Cobuild Learner's Dictionary*: A definition of the term, the part of speech the word belongs to, example(s) of the word being used in context and synonyms (optional).

The platform also provides the option of activating various types of interactive questions aligned with the text on the right-hand side of the screen, which pop up after clicking on a link. These include questions to train inferential reading comprehension as well as note-taking questions, intended to help students summarize the contents of the text being read. Users can choose to activate these two types of questions simultaneously or just focus on practicing either inferential reading or note-taking separately. An answer key is provided at the end of each question and the student responses are automatically recorded on the platform. These responses can be exported onto a PDF file at the end of each reading session.

The design features of the reading platform discussed above are aimed at training academic reading skills in ways that capture the complexity of the reading process effectively. Fluent reading, according to Grabe and Stoller (2013), can be understood as a highly complex interactive, strategic, evaluating, and purposeful operation where readers employ their processing memory to comprehend the information presented to them in the text. In doing so, they juggle between lower-level components (such as lexical access, syntactic parsing, semantic processing formation) and higher-level components (text model of comprehension, situation model of reader interpretation, use of background knowledge, and inferencing). Readers need to be able to process information in their working memory rapidly for comprehension to take place. Otherwise, the information will fade quickly from memory and would have to be reactivated, thus hindering the effectiveness of the reading process.

Electronic glosses provide easy access to difficult lexis during reading, so as to maximize lower-level syntactic parsing and semantic processing in the working memory of the readers. The inferential and note-taking questions focus on the higher-level components of the reading process identified by Grabe and Stoller (2013). They are intended to train complex inferencing as well as reading for gist

and reading for detail. Other features which could be incorporated onto our prototype platform are pre-reading questions to help activate schemata and exercises to pre-teach key lexis.

Admittedly, the effectiveness of these pedagogical tools in our reading platform could be partly offset by some negative effects on reading comprehension attributed to digital reading versus reading in print. While research recognizes the widespread use of digital media in producing and delivering content and the advantages digital media offer in terms of speed, breadth and cost of delivery, it is generally acknowledged that people who read online and/or on-screen tend to employ fewer high-order strategies such as highlighting and note-taking and that they may struggle harder with abstract questions and inferential reasoning (Kaufman and Flanagan 2016; Schugar et al. 2011). This can be attributed to the fact that digital reading encourages multitasking as well as skimming, with less time spent on in-depth concentrated reading, resulting in shallower comprehension and lower critical reflection (Baron 2017). Similarly, the use of hyperlinks during reading may increase cognitive load and produce disorientation (Eveland and Dunwoody 2001). Some of these negative effects can be minimized by design features already present in our platform. These include having interactive questions aligned with the section of the text they refer to, having the possibility of de-activating note-taking or inferential questions during a reading session and presenting the information glossed in layers, starting with a Chinese translation of the term, so as to facilitate selective reading.

3 Computer-assisted language learning (CALL)-based glossing in reading pedagogy

Research on glossing and reading pedagogy has traditionally focused on two main areas. The one examined in this study is whether electronic glosses can facilitate the comprehension of complex texts when read autonomously. A related area of research focuses on the potential glossing offered for vocabulary acquisition (Boers 2022). In both cases, the nature and effectiveness of glossing and gloss use behavior can vary significantly depending on the format (i.e. paper vs. CALL-based glossing; text vs. multimedia glosses, highlighted vs. non-highlighted glosses); the location of glosses in the text (i.e. a popup window next to the term, a note in the margin, etc.); the types of glossing (i.e. L1 vs. L2 glosses, definitions vs. grammatical explanations) and the readers' attitudes toward glossing, to name a few (De Ridder 2002; Marefat et al. 2016; Mohsen and Balakumar 2011; O'Donnell 2013; Taylor 2009). In any case, there is very little evidence in the literature that more

intensive taught-word comprehension interventions are more effective in improving generalized comprehension than less labor-intensive approaches such as CALL-based glossing, which does not require direct teacher intervention (Wright and Cervetti 2017).

Reading aided by text glosses potentially enhances comprehension (Zou 2016), as it can help raise the L2 reader's lexical threshold, facilitating the transfer of L1 skills to L2 reading and promoting greater access to authentic texts (Gettys et al. 2001). Immediate access to glossed L2 lexis can ease comprehension, as it saves the reader considerable time and effort during reading (Taylor 2009). This may help readers focus on higher-level strategies during reading, according to Ko (2005). In a study comparing the reading strategies used by L2 readers under gloss and no-gloss conditions, she found that although readers employed more reading strategies when reading non-glossed texts, these tended to be low-level strategies which did not result in better reading comprehension, whereas readers using glossed texts were better at higher-level strategies such as inferencing.

However, the frequent use of glosses could make the reading process more cumbersome, resulting in overall shallower processing and short-term vocabulary retention (Chan 2011). Such shallow processing can be attributed to the effects of high cognitive load on the reader's working memory after consulting glosses frequently. Short-term vocabulary retention can be explained by the theory of cognitive depth, which predicts that the higher and more active mental effort exerted when reading without glosses may be conducive to better vocabulary retention (Hulstijn 1992).

It has also been argued that electronic glossing may facilitate reading comprehension better than traditional paper-based glossing. Taylor's quantitative meta-analysis of 32 studies comparing the effects on reading comprehension of CALL-based versus paper-based glossing revealed significant differences in reading comprehension in support of this hypothesis (Taylor 2009). Other studies, however, have not shown significant differences between electronic and paper-based glosses in enhancing reading comprehension but have concluded that electronic glossing resulted in greater short-term gains in vocabulary and higher instructional efficiency in terms of the perceived cognitive load required during the reading task (Lee et al. 2016). Other glossing studies comparing tooltip electronic glossing with marginal electronic glossing have shown that the former approach results in lower levels of cognitive load due to lesser split-attention (Marefat et al. 2016).

The most important factors when studying the effect of glossing in reading comprehension are the complexity of the target text (Lyman-Hager and Davis 1996; Taylor 2002), the nature of the reading task (i.e. leisure reading vs. reading for study purposes, reading for gist vs. reading for detail, etc.) (De Ridder 2002) and,

above all, the reader's language proficiency level. Learner preferences toward glossing and gloss types tend to vary depending on the readers' proficiency. Lower-proficiency learners are said to favor glossing to other types of reading aid (Jacobs 1994) and they generally prefer L1 glosses to L2 ones (Bland et al. 1990; Ko 2005). Researchers have attributed this preference to the "naive lexical hypothesis", which argues that the higher the readers' language proficiency is, the less reliant readers are on one-to-one lexical matching behavior during reading (Bland et al. 1990). This hypothesis has been confirmed in a subsequent study by Kroll and Sunderman (2003) upon finding "a formative shift from a word association model to a conceptual model as L2 readers' proficiency increases" (O'Donnell 2013: 545).

The positive effects of glossing on reading comprehension are typically more significant among low-intermediate and intermediate learners. That is why these learner categories are the ones usually chosen in glossing research. If the text(s) being read are too simple or complex (conceptually, rhetorically, and/or linguistically), the reading tasks are more or less demanding, and the language proficiency of the participants is too low, too high and/or too dissimilar, the potential effect of glossing on reading comprehension could vary significantly, to the point of being obscured, if not altogether annulled. Hence, the effects of glossing on reading comprehension need to be analyzed in relative terms because, at some point, both complex and simple texts could become equally accessible to seasoned readers sufficiently proficient in English and at such a point, glossing would no longer be needed. Alternatively, in the case of low-proficiency learners confronted with authentic texts, the aid provided by glossing can prove insufficient all along. These methodological difficulties, paired with an enormous variability in research focus in terms of gloss design, text types, learner's background, and selection complicate the study of glossing significantly. To date, research on the effects of glossing on reading comprehension and on vocabulary acquisition remains largely inconclusive (O'Donnell 2013).

4 Research design

The present study aimed to determine the extent to which our online reading platform could help low-intermediate L2 learners read academic texts for study purposes. Our findings could help us design appropriate pedagogical tools for secondary and tertiary education using blended learning in the future. Our focus was mainly on a single feature in the platform, the online glossary, with two objectives in mind: (1) Determine the extent to which electronic glosses could facilitate the comprehension of complex academic texts by our learners. (2) Better understand how individual learners approach the reading tasks and describe the problems they encounter, especially those cases where electronic glosses prove to

be ineffective in facilitating comprehension. Focusing on the online glossary also allowed us to simplify the statistical analysis of the reading test scores by reducing the explanatory variables in our analysis to a minimum. The questions used in the reading tests could then be turned as interactive exercises for the platform by providing the answer key on-screen.

In order to measure the effectiveness of electronic glosses on reading comprehension more accurately and independently from other factors, we needed to consider three main variables before conducting the reading tests: (1) the target readers (i.e. their English proficiency, estimated reading ability, and level of background knowledge); (2) the linguistic complexity of the texts, by considering their lexico-grammar, syntax, semantics, and rhetorical structure; and (3) the reading demands implicit in the reading task, which in our case required testing deep inferential reading for study purposes as opposed to shallow reading for leisure. These three variables were then studied independently and sequentially using a combination of quantitative and qualitative approaches to help us select the most appropriate texts for the study as well as learners of similar language proficiency and reading ability.

4.1 Defining the target readers

Our first step was to outline a learner profile representative of our student population, which could be used as a basis for selecting the participants in the study. This was done by (1) assessing the reading comprehension ability of a representative sample of students using two abridged IELTS academic reading tests for diagnosis and by (2) surveying this same group of students, so as to understand their reading practices and perceived difficulties when dealing with academic texts better.

A total of 1,406 students completed the abridged IELTS academic reading diagnostic test. Their average estimated IELTS score was between 5 and 5.5 (see Figure 1). The DSE (Diploma of Secondary Education) grade equivalent to this average IELTS score would be “between 2 and 3”, as estimated by the Hong Kong Examination Authority in a large-scale benchmarking study:¹ This estimate coincides with the actual DSE grades reported by the students who responded to a survey conducted after undergoing the diagnostic test, to be detailed below.

A total of 631 students, out of the 1,406 who took the diagnostic reading test, completed the survey on reading practices. The demographics of the population surveyed were as follows:

¹ <http://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/>.

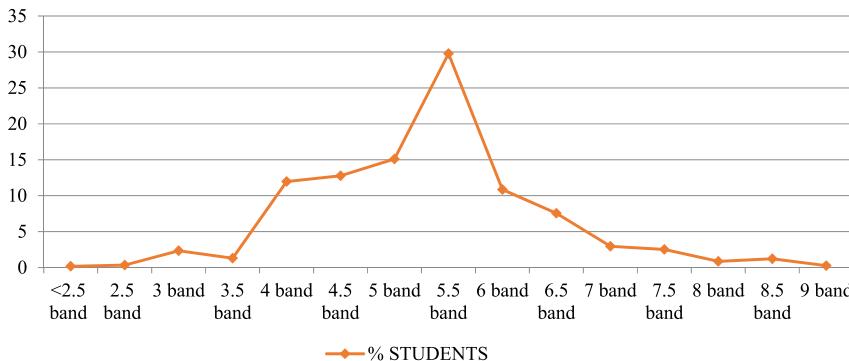


Figure 1: Diagnostic test: IELTS scores of students.

- mean age: 19.56 ($SD = 2.85$);
- gender: 52.5% male; 47.5% female;
- study mode: 99% full-time;
- year of study: 45.6% (Year 1); 24.9% (Year 2); 28.8% (Year 3); 0.6% (Year 4);
- year of entry: 71.9% (Year 1 entry); 28.1% (Year 3 entry);
- students came from 21 different degree programs; and
- parents with university degrees: father (15.21%); mother (7.92%).

The profile that resulted from the IELTS diagnostic test of 1,406 students and the survey of nearly 45% of those students pointed at a typical learner

able to understand simple texts if the topic is familiar and able to follow the development or parts of the development of an explicit argument and identify explicit opinions when they are clearly signaled, as well as make (at times) straightforward inferences and work out the meaning of unfamiliar words when a simple and familiar context is given, and either respond in part to simple written instructions requiring relevant information from the texts to complete a task (HKDSE 3) or follow simple instructions to locate and transfer some information relevant to a given task (HKDSE 2).²

This reading ability can be considered below the Cognitive Academic Language Proficiency Threshold needed for English-medium education at degree level (Cummins 1981, 1983, 1985).

An examination of the leisure reading practices of the learners surveyed in terms of frequency and main text types, revealed interesting differences between the use of Chinese and English as well as the extent to which students engaged with complex academically-oriented texts during leisure time (see Figures 2 and 3).

2 http://www.hkeaa.edu.hk/en/hkdse/hkdse_subj.html?A1&I2_4 (accessed July 10, 2022).

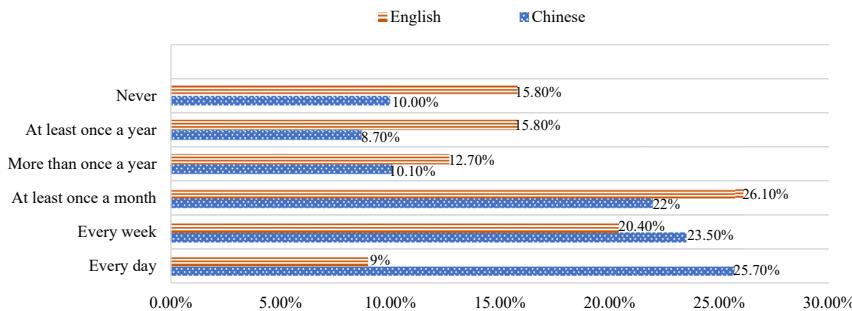


Figure 2: Frequency of leisure reading.

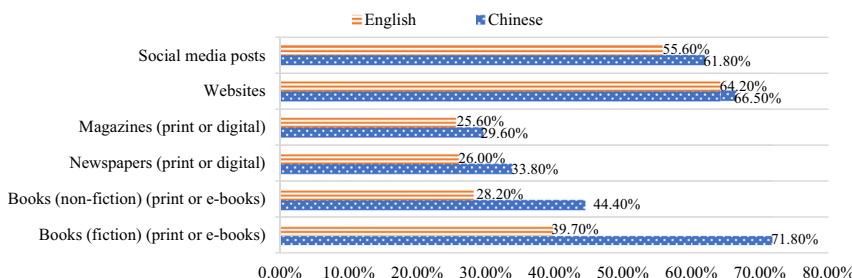


Figure 3: Leisure reading habits (main text types).

Our target learners also tended to read more often in Chinese. Out of the 631 students surveyed, only 9 and 20.40% reported reading daily or weekly in English. A large percentage of students could not be considered frequent readers in neither Chinese nor English. Over 50% of the respondents declared to read for pleasure in Chinese “at least once a month” or less than that. The percentage of infrequent readers in English reached 70%. The kinds of texts our respondents read most frequently were also indicative of a lower-intermediate learner profile, who may struggle to deal with academic readings and literary texts. In view of this, it could be assumed that the reading ability of the subjects in this study could be partly affected by first language (L1) negative transfer, as Chinese is not an alphabetic language (Wang and Koda 2007) and, especially, by their low English proficiency: Poor word recognition and syntactic parsing can inhibit semantic proposition formation at the lower level (i.e. sentences and paragraphs) and hence, higher-level processing leading to overall text comprehension can be seriously impaired,

if it ever occurs. This happens because a reader's working memory can be overburdened by slow lower-level processing, thus inhibiting the automaticity needed for higher-level interpretation (Presley 2006).

4.2 Determining text complexity

Next, we aimed at explaining text complexity in reading comprehension, so as to define the criteria for selecting texts with varying degrees of difficulty and measure the interaction effect between variables such as the language proficiency of students and/or the use of the online glossary that can facilitate text comprehension across a variety of texts. This was accomplished by comparing three different sets of texts (eight General Reading IELTS, eight Academic Reading IELTS and eight Journal Article Introductions) quantitatively, in terms of (1) general readability, (2) lexical complexity, (3) syntactic complexity, and (4) semantic complexity, and qualitatively, in terms of (5) intended audience, (6) communicative purpose, and (7) rhetorical organization. Descriptive and inferential statistics using mixed model ANOVA were employed in the quantitative analysis of text complexity.

A corpus of 24 texts belonging to three pre-determined categories (Category A. General Reading IELTS, Category B. Academic Reading IELTS, and Category C. Journal Article Introduction) was built for comparison purposes. Such comparison was needed to ascertain the extent to which we were actually dealing with three distinctive text categories in terms of reading complexity as well as to identify any linguistic features typically found in those texts for possible pedagogical exploitation in the future. Defining text complexity could also help us select three representative samples, one from each text category, for the final reading tests, in order to measure the effectiveness of the online glossary on reading comprehension. Category C texts (Journal Article Introduction) were shortened by eliminating in-text citations when these were not author-led, so as to reduce the number of words while preserving the information in the texts. Using texts employed in IELTS reading tests in categories A and B considerably facilitated our comparative analysis, as the design of the IELTS test includes a rigorous standardization process both in choosing reading texts and in formulating test questions. The details of the corpus can be found in Table 1.

A mixture of quantitative and qualitative analysis was used to compare text complexity across these three categories. The quantitative analysis examined (1) general readability, (2) lexical complexity, and (3) syntactic complexity, whereas the qualitative analysis looked into issues such as intended audience, communicative purpose, and rhetorical organization.

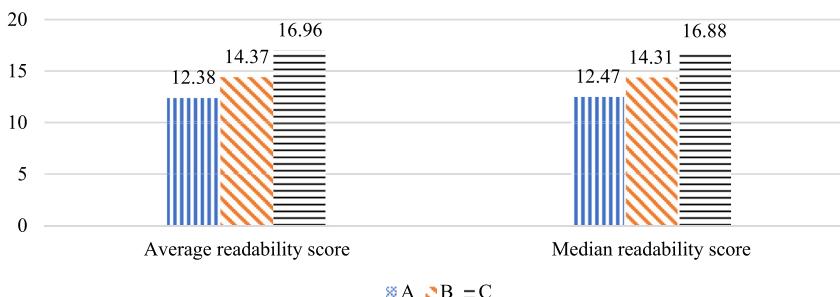
Table 1: Corpus description.

	Tokens	Types	Token/Type ratio
Category A (IELTS general reading)	6,822	2,135	31%
Category B (IELTS academic reading)	7,124	2,021	28%
Category C (Journal article introduction)	8,473	2,167	25%
Overall	22,419	4,564	20%

4.2.1 Quantitative analysis: general readability

General readability was measured using a combination of five indexes based on information such as word and sentence length and syllable count: (1) Gunning Fog, (2) Flesch Kinkaid, (3) SMOG, (4) Coleman-Liau, and (5) Automated Reading Index. As each of these indexes employs a slightly different formula, we opted for averaging the readability scores in pursuing higher reliability in our analysis. Figure 4 shows the average and median readability scores across the three categories.

The one-way ANOVA results on Table 2 indicate that the differences in general readability between the three text categories are highly significant and that there is a clear progression in terms of difficulty between the three text-type categories (General Reading IELTS being the easiest and Academic Journal Introductions being much more difficult in terms of readability). Further post-hoc analysis is presented in Table 3.

**Figure 4:** Average and median readability scores across the three text categories.**Table 2:** One-way ANOVA (general readability).

	<i>n</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p-value</i>
Average readability score	24	85.09	42.55	25.99	25.99	>0.001
Median readability score	24	78.59	39.30	26.08	26.08	>0.001

Table 3: Post-hoc analysis (general readability).

	Comparisons	M_{diff}	SE	p-value	[95%CI]	
					Lower bound	Upper bound
Average readability score	A B	-2.01	0.64	0.01	-3.62	-0.40
	C B	-4.60	0.64	>0.001	-6.21	-2.99
	C A	2.01	0.64	0.01	0.40	3.62
Median readability score	B C	-2.59	0.64	>0.001	-4.20	-0.98
	C A	4.60	0.64	>0.001	2.99	6.21
	C B	2.59	0.64	>0.001	0.98	4.20
B A	B C	-1.84	0.61	0.02	-3.39	-0.29
	A C	-4.41	0.61	>0.001	-5.96	-2.87
	B A	1.84	0.61	0.02	0.29	3.39
C A	C B	-2.57	0.61	>0.001	-4.12	-1.02
	A C	-2.57	0.61	>0.001	-4.12	-1.02
	A B	4.41	0.61	>0.001	2.87	5.96
C B	B C	2.57	0.61	>0.001	1.02	4.12
	B A	2.57	0.61	>0.001	1.02	4.12

4.2.2 Quantitative analysis: lexical complexity

General readability can be more practical in estimating reading complexity among L1 users than in L2 learners, who may have additional trouble with word recognition. It has been proposed that readers need to be able to understand between 96 and 98% of the lexis in a text (Hu and Nation 2000; Schmitt et al. 2011) for effective comprehension. With L2 learners, it can also be assumed that less frequent words as well as academic lexis can pose a significant obstacle to reading comprehension (Brown 2012; Hyland and Tse 2009; Milton 2009). Based on this assumption, one can compare the differences across the three text-type categories in terms of percentages of less frequent and academic lexis. This comparison was carried out using vocabulary profiles from a variety of well-established word lists: (1) the Common European Framework of Reference for Languages (CEFR), (2) the General Service List (GSL) + Academic Word List (AWL) or “Classic”, (3) the Billuroglu-Neufeld List (BNL), and the British National Corpus (BNC) + Corpus of American Contemporary English (COCA 25) off list.³

Figure 5 shows the differences across the three text-type categories in the CEFR Lists 1, 2 and off-list (being CEFR 1 and 2 the most common 2,000 words and the off-list the rest of the lexis) expressed in types and tokens. Similarly, Figure 6 shows the differences across the three text types by comparing the percentages of the most common words (K1 and K2) as well as those percentages of the words in the Academic Word List and the rest of the words (off-list).

Figure 7 compares the percentage of most frequent words across the three text-type categories by grouping the lexis in two sets: least frequent words (BNL 4K

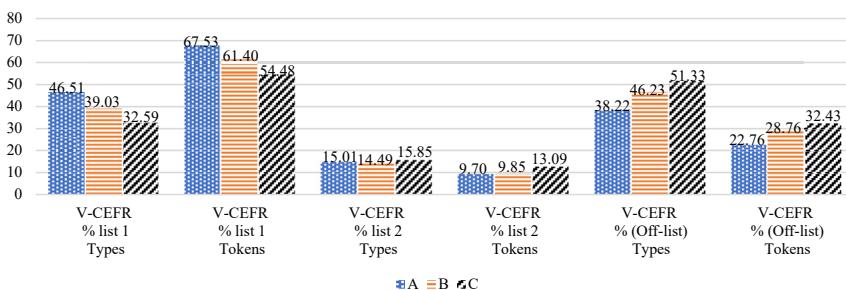


Figure 5: Comparative lexical analysis (CEFR wordlists).

3 <https://www.lextutor.ca/vp/comp/>.

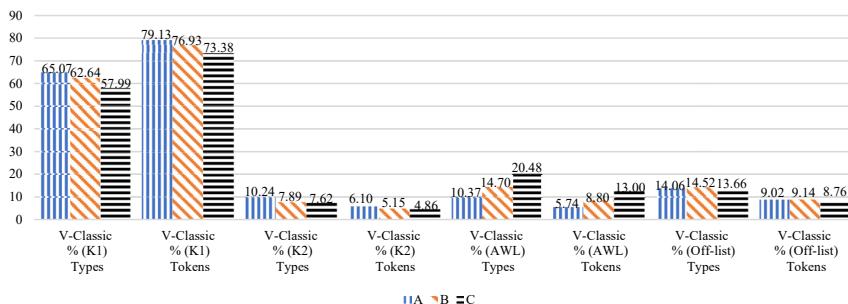


Figure 6: Comparative lexical analysis (V-Classic wordlist).

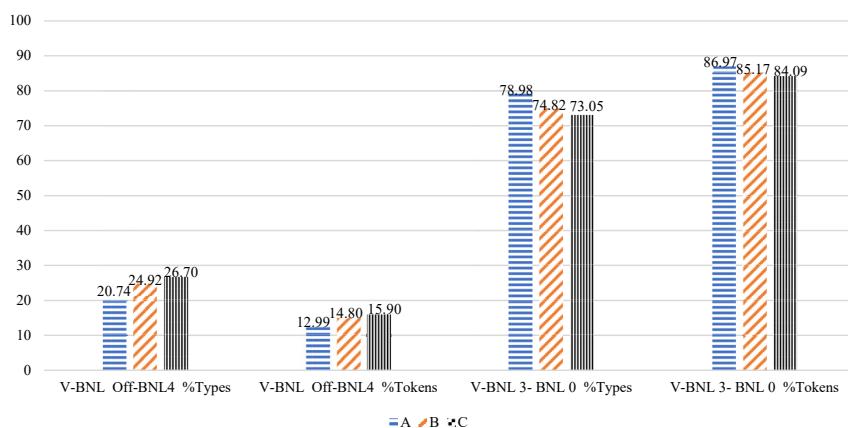


Figure 7: Comparative lexical analysis (Billuroglu-Neufeld wordlist).

words and BNL off-list words) versus the most frequent ones (BNL 3K to BNL 0). Finally, Figure 8 illustrates the lexical distribution of the off-list lexis expressed in types and tokens of BNC/COCA 25 across the three text-type categories.

A general trend can be observed in the distribution of lexis across the three text types. No matter which word list is chosen for comparison purposes, Category A (General Reading IELTS) contains a larger percentage of the most frequent lexis, followed by category B (Academic Reading IELTS) and Category C respectively. The opposite trend is observed in the case of least frequent lexis (i.e. off-list words). The only exception to this trend is the percentage of off-list lexis in the V-Classic word list, resulting from a significantly higher concentration of lexis from the Academic Word List in Category C texts. This higher presence of AWL lexis, which also happens to be infrequent, actually confirms that Category C texts are more

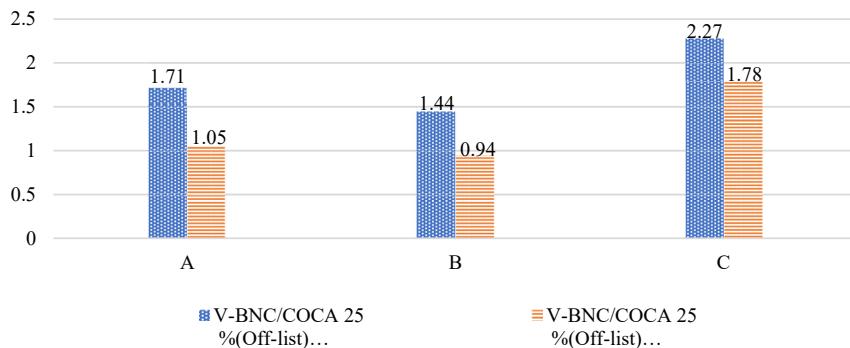


Figure 8: Comparative lexical analysis (V-BNC/COCA25 wordlists).

academically oriented than those in categories A and B respectively. Table 4 highlights the word list categories where the distribution of lexis across text types can be considered statistically significant according to one-way ANOVA.

4.2.3 Quantitative analysis: syntactic complexity

It can also be assumed that texts with longer sentences and complex syntactic structures would add additional load to a reader's processing memory, thus hindering comprehension. Syntactic complexity was measured by comparing the three text type categories in terms of lexical density as well as factors associated with lexical density, such as "nominalization" and "grammatical metaphor".

Table 4: One-way ANOVA (lexical complexity).

	<i>n</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p-value</i>
V-CEFR % (Off-list) types	24	698.74	2	349.37	16.39	<0.001
V-CEFR % (Off-list) tokens	24	381.45	2	190.73	9.41	<0.001
V-CEFR % (List 1) types	24	777.08	2	388.54	18.08	<0.001
V-CEFR % (List 1) tokens	24	682.06	2	341.03	14.36	<0.001
V-CEFR % (List 2) tokens	24	58.61	2	29.31	8.65	<0.001
V-classic % (K1) types	24	206.94	2	103.47	4.10	0.03
V-classic % (K1) tokens	24	134.67	2	67.34	3.88	0.04
V-classic % (AWL) types	24	412.14	2	206.07	18.86	<0.001
V-classic % (AWL) tokens	24	212.70	2	106.35	22.42	<0.001
V-BNL off BNL 4% types	24	149.67	2	74.84	5.30	0.01
V-BNL 3-BNL 0% types	24	148.46	2	74.23	5.27	0.01
V-BNC/COCA 25% (Off-list) types	24	2.91	2	1.46	4.76	0.02

(To 2018). Lexical density refers to “the density of information in any passage of text, according to how tightly the lexical items (“content words”) have been packed into the grammatical structure” (Halliday 1993: 76).

The relationship between lexical density and nominalization is illustrated in Figure 9 below. One can see how information can be packed into complex sentences employing hypotaxis and/or large nominal groups. The latter are typically characterized by the presence of nominalization (highlighted in bold), consisting of encoding “processes” and “qualities” as nouns (i.e. comparisons, quality) instead of opting for the more congruent options of verbs (i.e. compare) and adjectives (i.e. good/bad). When the congruent form can be recovered by rephrasing the text, we are in the presence of grammatical metaphor, which hence can be considered a subset of nominalization (see 1a–1b).

- (1) a. Such **prevalence** of nature deficit affects many millions urban inhabitants [...].
- b. When nature deficit **prevails**, that affects many millions urban inhabitants [...].

Typically, a higher presence of hypotaxis combined with nominalization leads to higher levels of lexical density, which is said to affect readability (To 2018). A further challenge to effective reading comprehension may come from unpacking heavily nominalized texts which, in addition to being lexically denser, tend to be

Hypotaxis without nominalization (A6)

||| While it is **possible** that much of our **food** will be **grown** in **skyscrapers** in **future**,|| (α) most experts currently **believe** it is far more **likely** that we will simply **use** the **space available** on **urban rooftops**|||(β) [16: 5/11] [20: 8/12] **LD: 13:2=6.5**

A heavily nominalized single-clause sentence (B7)

||| “Such **prevalence** of **nature deficit** affects many millions **urban inhabitants**, with **widespread implications** on the **quality** of the **living environment** and **quality** of **life**.” ||| [24: 13/11] **LD=13**

A complex sentence combining hypotaxis and nominalization (B1)

||| “Cross-societal **comparisons** where **culture** is considered a significant **determinant** have become increasingly important for studying **quality** of **life** and factors accounting for its **variance**.” ||| [24: 14/10] **LD=14**

Key:

	Clause complex boundary
	Ranking clause boundary
α, β	Hypotactic structure
Lexical items	
Nominalization	bold
Grammatical metaphor	<i>italics underlined</i>

Figure 9: Lexical density and nominalization.

highly abstract, technical as well as impersonal (Fang 2004). It has been argued that nominalization can affect reading comprehension in the case of low proficiency readers like the ones who took part in this study (Duffelmeier 1979).

Figures 10 and 11 compare the differences in the average lexical densities and nominalization/grammatical metaphor among the three text-type categories. A clear progression can be observed between General Reading IELTS, where lexical density and nominalization/grammatical metaphor are lowest, and Research Article Introductions, characterized by significantly higher lexical density and nominalization/grammatical metaphor.

The ANOVA results shown in Table 5 indicate that the statistical differences among the three text-type categories are significant in all cases (as well as in the average percentage of nouns across text types but not in the average percentage of sentences containing subordinate clauses).

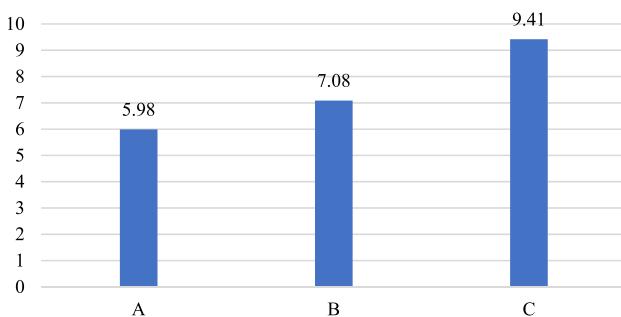


Figure 10: Average lexical densities across the three text categories.

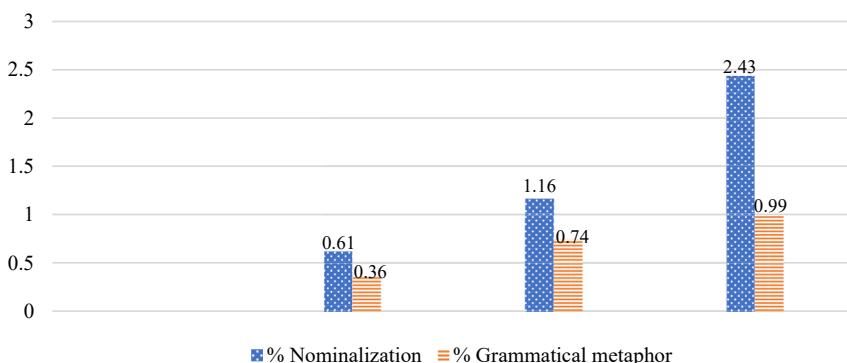


Figure 11: Average percentage of nominalization and grammatical metaphor across the three text categories.

Table 5: One-way ANOVA (syntactic complexity).

	<i>n</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p-value</i>
% Subordination	24	111.37	2	55.69	0.30	0.75
% Nouns	24	93.00	2	46.50	6.82	0.01
% Lexical density	24	49.01	2	24.50	14.62	<0.001
% Nominalization	24	13.85	2	6.93	30.28	<0.001
% Grammatical metaphor	24	1.59	2	0.79	7.63	<0.001

4.2.4 Qualitative analysis of text complexity

Table 6 illustrates the differences between the different text-type categories in terms of the intended audience(s), communicative purpose(s), and main rhetorical patterns. These discourse features can have a significant impact on the reading comprehension of novice L2 readers, particularly in how they make use of higher-level components, such as the text and situation model of comprehension, at the processing stage in the working memory.

Comparatively speaking, the expository and simple argumentative texts in categories A (General Reading IELTS) and B (Academic Reading IELTS) do not usually require significant prior background knowledge for successful reading comprehension and their overall textual organization tends to be quite linear and explicit. For instance, in Category A texts, information tends to be organized chronologically or descriptively. Although these two patterns may still feature in Category B texts, the latter are mainly characterized by simple argumentative structures employing topic sentences. On the contrary, Category C texts require significant specialist background knowledge from readers, as many of the concepts introduced are often taken for granted (and therefore left unexplained). Also,

Table 6: Comparative qualitative analysis of the three text types.

	A (General reading IELTS)	B (Academic reading IELTS)	C (Journal article introductions)
Intended audience(s)	General: Young adults, L2 learners	Academic: Young adults, L2 learners	Specialist: Researchers (discipline specific)
Communicative purpose(s)	Inform, describe	Inform, explain	Explain, justify
Rhetorical pattern	Expository	Mixed expository-argumentative	Argumentative: Create a research gap (Swales 1990)

novice L2 readers would encounter considerable difficulty in following the arguments in those texts because of the shift between moves in research article introductions (“establish a research territory”, “establish a research niche” and “occupy the research niche”), embedded in a literature review, are not always explicitly signposted in journal article introductions. As a result of that, non-expert readers may end up being confused by what they may perceive as a topic shift instead of a shift between moves, where existing knowledge is being problematized by the writers in order to establish a research niche.

All in all, if one adds to this qualitative analysis the statistically significant differences across the three text categories in terms of general readability as well as linguistic complexity at the lexical and syntactic levels discussed before, it seems safe to assume we are dealing with three distinctive text categories, which can pose markedly different challenges to our target learners (defined as novice readers with a language proficiency below the CALP threshold). Based on this insight, it seems possible to use samples of these three text categories to estimate the extent to which the hyperlink glossary function in the reading platform can make category A, B, and C texts more accessible to our target readers.

4.3 Conducting the reading tests

After selecting three sample texts, one from each of the above-mentioned categories, three separate reading tests were conducted with students identified during the diagnostic test and the survey as representative of our student population (refer to Figure 4). These students, whose L1 was Cantonese, were divided into two groups based on the score they had obtained in the diagnostic test: The control group students, having fallen within the IELTS band 5 (10–12 correct responses), and the testing group students, having reached a 5.5 IELTS band (13–16 correct responses). Each student was randomly assigned to read each of the three texts either on a Word File on a computer screen or in our online platform with hyper-linked glosses, intended to facilitate lexical comprehension by speeding up working memory processing.

Chosen from Categories A and B in our corpus (IELTS general reading and academic reading respectively), tests 1 and 2 lasted a maximum of 25 min each and they were administered consecutively during a single session. Each test included a total of 15 questions resembling those used in the IELTS diagnostic test for validation purposes: (1) Paragraph gist questions, where the readers chose the titles for different paragraphs from a multiple-choice set; (2) Inference questions using the True/False/Not Given options, and (3) information recall questions, where students needed to fill in the blanks of a paraphrased version of the reading text with

words used in the original version. Test 3, based on the introduction of a journal article, comprised a total of 25 questions of the same type, administered during a separate session lasting a maximum of 45 min.

All the tests were conducted individually and observed live by a researcher, who took notes during the process. The tests were also video-recorded using an on-screen recording software for record-keeping and double-checking. During the tests, the participants were instructed to highlight on the computer screen those parts in the texts which they found difficult to understand and post-test interviews were conducted immediately after each test to further discuss the kinds of problems encountered by the readers during the tests. Other features of the online reading platform, such as the inference and note-taking questions, were also discussed during post-test interviews to seek the views of the students for future research.

A total of 46 students completed Tests 1 and 2. Unfortunately, only 19 of those also completed Test 3, as the rest of the students opted for discontinuing the project or did not reply to our messages. This rendered further statistical analyses impossible for the third test.

5 Results and discussion

Table 7 shows the mean scores obtained by the 46 students in Tests 1 and 2 (General Reading IELTS and Academic Reading IELTS) either done using the online platform or in a plain Word file shown on a computer screen. Several trends can be observed: (1) The overall mean score of the online reading tests is higher than the one of the plain Word tests; (2) The mean score among both Control group and Testing group students is also higher in the online reading version of the tests; (3) T students tend to perform better than C students in both online reading and plain Word tests. Given the insufficient sample size of students who completed Test 3, no useful conclusions can be drawn about their performance in such test, which was supposedly more difficult than Tests 1 and 2.

A mixed-model/split-plot ANOVA was used to examine the interaction effect of English proficiency (between-subject factor) and the use of online glosses (within-subject factor) on the reading comprehension of the participants. However, the interaction effect observed was not significant, $F(1,44) = 0.14, p = 0.72$. In other words, English proficiency did not interact with the use of online glosses on their overall reading comprehension. This lack of interaction may be attributed to the small sample size of learners in the control versus the testing group. Nonetheless, there was a between-subject main effect, that is, their language proficiency, on their reading comprehension, $F(1,44) = 4.91, p < 0.05$. In general, T students performed better on both tests ($M = 7.75$) than C students ($M = 6.39$), as predicted by

Table 7: Results of reading Tests 1 and 2 (General Reading and Academic Reading IELTS).

	English proficiency	Mean	Std. deviation	<i>n</i>
Online_Score	Control group	7.09	2.202	22
	Testing group	8.29	2.726	24
	Total	7.72	2.536	46
Plain text_Score	Control group	5.68	2.750	22
	Testing group	7.21	2.553	24
	Total	6.48	2.730	46

the diagnostic test used to place students in either the C or the T group. There was also a within-subject main effect, that is, the use of online glossary, on their overall reading comprehension, $F(1,44) = 7.90, p < 0.05$. In other words, students in both C and T groups performed significantly better statistically-speaking with the online version ($M = 7.72, SD = 2.57$) than the plain text version ($M = 6.48, SD = 2.73$).

The qualitative analysis carried out during in-test observation and post-test interviews can throw further light on the effect the online glosses had on the reading comprehension of General Reading and Academic Reading IELTS texts among the participants in our study. On the one hand, it was confirmed that low frequency and/or academic lexis was often perceived as particularly problematic by our learners. On the other hand, although a majority of participants found the online glosses quite effective in enhancing their reading comprehension by making the target texts more transparent, a number of shortcomings were revealed. The online glosses proved to be more effective when the meaning of an unknown keyword, which was needed for grasping the meaning in the text, could be retrieved quickly. However, when the readers encountered several unknown words concentrated on a small stretch of text, the glosses became less effective, probably due to a significant slow-down in the processing speed of the working memory. Some readers reported that, despite having understood the individual words on the text, they proved unable to process the information at whole-text level by encoding propositions based on the text effectively. This may have been due to a general lack of reading fluency on the part of the learners in question, and/or higher-level comprehension may have been hindered by excessive syntactic or semantic complexity in the text. This finding seems to confirm the cognitive load hypothesis proposed by Chan (2011) among others.

With regard to the perceived hindrance to understanding the text, the participants overwhelmingly pointed to difficult words as the leading source, followed by long or complicated sentences, unfamiliarity with the style or genre, difficult topics, and taxing of working memory, respectively. When interpreting the above comments, the concern is the extent to which the participants are conscious of the causes of the

difficulty and able to articulate them accurately. For instance, the video recording reveals that many participants frequently referred to the paragraphs they had already read, which may indicate that the task placed a high working memory demand on them. One might speculate that some of the participants were unaware of the role of working memory in the task while an unknown word is unlikely to have escaped their notice. As for the perceived difficulty of the test questions, the most frequently cited item was paragraph gist, followed closely by inferential reading. Interestingly, the treatment group reported significantly more difficult items than the control group, most particularly the inferential questions. One might hypothesize that the treatment group was more sensitive to the difficulty of different question types, which could have been the result of a higher language proficiency.

All in all, the present study confirms a moderately positive effect of online glosses in the comprehension of complex texts among low-intermediate learners. Such positive effect can be hindered by an excessive cognitive load on the readers' working memory at times, as found in previous research. However, given the enormous variability in research focus and design found in the literature, it remains difficult to pinpoint the extent to which online glossing can prove effective with more complex texts and/or with learners with slightly higher or lower language proficiency than the ones who participated in our study. In other words, more research is needed targeting learners with different levels of proficiency and using more complex texts, such as research articles to validate the findings of this study. Perhaps, another major contribution from this study is design-based research inspired methodology. Such methodology can help address many of the shortcomings of previous studies, which could ultimately invalidate their findings. One cannot test the effectiveness of glossing accurately without a rigorous selection of participants in terms of language proficiency and reading habits and a well-informed selection of texts, based on painstaking linguistic analysis, in an attempt to reduce the explanatory variables to a minimum.

6 Conclusion and future directions

This paper has illustrated how design-based research can be put to good use in testing the effectiveness of technology-enhanced learning innovations aimed at improving reading instruction for academic purposes. Our findings indicate that tools such as the glossary employed in our online reading platform can make complex texts more transparent among readers with language proficiency below the CALP threshold. Such a glossary can allow these readers to grasp the meaning of unknown words instantly, thus speeding up syntactic parsing and enabling at times other higher-order reading processes such as summarizing and inferencing at paragraph and whole-text levels.

Reading instruction can be further enhanced thanks to a better understanding of how factors such as text complexity, from a linguistic and a cognitive point of view, can influence effective reading comprehension. The quantitative and qualitative analyses across a variety of text types carried out in this study have revealed clear trends in text complexity at various levels: lexical, syntactic, and discourse levels. Thanks to this insight, we are now in a better position to enhance academic reading instruction online with pedagogical strategies and additional tools that can help us minimize the impact of text complexity across a variety of text types. For instance, in addition to having tools such as the paragraph gist and note-taking questions, already included in our online reading platform, we could make good use of strategies such as pre-reading questions and vocabulary games to activate reading schemata and pre-teach key lexis. Another option would be to develop online tasks that could train readers to unpack information in heavily nominalized texts into more congruent forms and design post-reading tasks to help learners consolidate the academic vocabulary previously introduced during reading. Last but not least, explicit instruction on how a research space is created rhetorically in journal article introductions could be provided by including on-screen annotations highlighting moves and steps and/or by using questions promoting readers to label moves and steps on the text.

On the whole, the insights gained from this study can be easily incorporated into the design of an improved online platform dedicated to providing reading instruction to speakers of English as a second language, for instance in Content and Language Integrated learning programs. The platform could also be employed at the university level in advanced EAP courses, or in self-access reading activities aimed at training reading for study skills or research purposes.

Research funding: This work was supported by a teaching development grant from the Technological and Higher Education Institute of Hong Kong (grant number SG171818).

References

Anderson, Terry & Julie Shattuck. 2012. Design-based research: A decade of progress in educational research? *Educational Researcher* 41(1). 16–25.

Baron, Naomi S. 2017. Reading in a digital age. *Phi Delta Kappan* 99(2). 15–20.

Bland, Susan K., James S. Nobbitt, Susan Armington & Geri Kay. 1990. The naive lexical hypothesis: Evidence from computer-assisted language learning. *The Modern Language Journal* 74. 440–450.

Boers, Frank. 2022. Glossing and vocabulary learning. *Language Teaching* 55(1). 1–23.

Brown, Dale. 2012. The frequency model of vocabulary learning and Japanese learners. *Vocabulary Learning and Instruction* 1(1). 20–28.

Chan, Alice Yin Wa. 2011. The use of monolingual dictionary for meaning determination by advanced Cantonese ESL learners in Hong Kong. *Applied Linguistics* 33(2). 111–140.

Collins, Allan, Diana Joseph & Bielaczyc Katerine. 2004. Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences* 13(1). 15–42.

Cummins, James. 1976. The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypothesis. *Working Papers on Bilingualism* 9. 1–43.

Cummins, James. 1981. The role of primary language development in promoting educational success for language minority students. In California State Department of Education (ed.), *Schooling and language minority students: A theoretical framework*, 3–50. Los Angeles: California State Department of Education.

Cummins, James. 1983. Language proficiency and academic achievement. In John Oller (ed.), *Issues in language testing research*, 108–130. Rowley, Mass: Newbury House.

Cummins, James. 1985. The construct of language proficiency in bilingual education. In James E. Alatis & John J. Staczek (eds.), *Perspectives on bilingualism and bilingual education*, 209–231. Washington DC: Georgetown University Press.

De Ridder, Isabelle. 2002. Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension and the reading process? *Language, Learning and Technology* 6(1). 123–146.

Dede, Chris. 2005. Why design-based research is both important and difficult? *Educational Technology* 45(1). 5–8.

Duffelmeyer, Frederick A. 1979. The effect of rewriting prose material on reading comprehension. *Reading World* 19(1). 1–11.

Ellis, Rod. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.

Eveland, Willian P. & Sharon Dunwoody. 2001. User control and structural isomorphism or disorientation and cognitive load? Learning from the Web versus print. *Communication Research* 28(1). 48–78.

Fang, Zhihui. 2004. Scientific literacy: A systemic functional perspective. *Science Education* 89. 335–347.

Garralda Ortega, Ángel. 2018. A case for blended EAP in Hong Kong higher education. *Asian EFL Journal* 20(92). 6–34.

Gettys, Serafima, Lorens A. Imhof & Joseph O. Kautz. 2001. Computer-assisted reading: The effect of glossing format on comprehension and vocabulary retention. *Foreign Language Annals* 34. 91–106.

Grabe, William. 2014. Key issues in L2 reading development. In Xudong Deng & Richard Seow (eds.), *Proceedings of the 4th CELC symposium for English language teachers*, 8–18. Singapore: Centre for English Language Communication. [https://www.nus.edu.sg/celc/research/books/4th%20Symposium%20proceedings/2\).%20William%20Grabe.pdf](https://www.nus.edu.sg/celc/research/books/4th%20Symposium%20proceedings/2).%20William%20Grabe.pdf) (accessed 24 August 2021).

Grabe, William P. & Fredricka L. Stoller. 2013. *Teaching and researching: Reading*, 2nd edn. Abingdon: Routledge.

Halliday, Michael A. K. 1993. Some grammatical problems in scientific English. In Michael A. K. Halliday & James R. Martin (eds.), *Writing science: Literacy and discursive power*, 159–180. London: Falmer.

Hermida, Julian. 2009. The importance of teaching academic reading skills in first-year university courses. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1419247 (accessed 25 August 2021).

Hirvela, Alan R. 2016. *Connecting reading & writing in second language writing instruction*, 2nd edn. Detroit: University of Michigan Press ELT.

Hu, Marcella Hsueh-Chao & Paul Nation. 2000. Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13(1). 403–430.

Hulstijn, Jan H. 1992. Retention of inferred and given meanings: Experiments in incidental vocabulary learning. In Pierre J. L. Arnaud & Béjoint Henry (eds.), *Vocabulary and applied linguistics*, 113–125. London: Macmillan.

Hyland, Ken & Polly Tse. 2009. Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies* 9(2). 111–129.

Jacobs, George M. 1994. What lurks in the margin: Use of vocabulary glosses as a strategy in second language reading. *Issues in Applied Linguistics* 5. 115–137.

Kaufman, Geoff & Mary Flanagan. 2016. High-low split: Divergent cognitive construal levels triggered by digital and non-digital platforms. In Jofish Kaye, Druin Allison, Cliff Lampe, Dan Morris, Juan P. Hourcade, Loren Terveen & Scooter Morris (eds.), *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2773–2777. New York, NY: Association for Computing Machinery.

Kelly, Anthony E. 2004. Design research in education? Yes, but is it methodological? *The Journal of the Learning Sciences* 13(1). 115–128.

Ko, Myong Hee. 2005. Glosses, comprehension and strategy use. *Reading in a Foreign Language* 17. 125–143.

Krashen, Stephen. 2004. *The power of reading: Insights from the research*. Portsmouth, NH: Heineman.

Kroll, Judith F. & Gretchen Sunderman. 2003. Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representations. In Catherine J. Doughty & Michael H. Long (eds.), *The handbook of second language acquisition*, 104–129. Malden, MA: Blackwell.

Lee, Hansol & Jang Ho Lee. 2013. Implementing glossing in mobile-assisted language learning environments: Directions and outlook. *Language, Learning and Technology* 17(3). 6–22.

Lee, Ho, Hansol Lee & Jang Ho Lee. 2016. Evaluation of electronic and paper textual glosses on second language vocabulary learning and reading comprehension. *The Asia-Pacific Education Researcher* 25. 499–507.

Lyman-Hager, Mary-Ann & James F. Davis. 1996. The case for computer-mediated reading: Une vie de boy. *French Review* 69(5). 775–792.

Marefat, Hamideh, Abbas Ali Rezaee & Farid Naserieh. 2016. Effect of computerized gloss presentation format on reading comprehension: A cognitive load perspective. *Journal of Information Technology Education: Research* 15. 479–501.

Milton, James. 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mohsen, Mohammed Ali & Mohammed Balakumar. 2011. A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL* 23(2). 135–159.

O'Donnell, Mary E. 2013. Second language learners' use of marginal glosses. *Foreign Language Annals* 45(4). 543–563.

Presley, Michael. 2006. *Reading instruction that works*, 3rd edn. New York: Guilford Press.

Reeves, Thomas C. 2005. Design-based research in educational technology: Progress made, challenges remain. *Educational Technology* 45(1). 48–52.

Roessingh, Hetty, Pat Kover & David Watt. 2005. Developing cognitive academic language proficiency: The journey. *TESOL Canada Journal* 23(1). 1–27.

Schmitt, Norbert, Xiangying Jiang & William Grabe. 2011. The Percentage of words known in a text and reading comprehension. *The Modern Language Journal* 95(1). 26–43.

Schugar, Jordan T., Heather Schugar & Christian Penny. 2011. A nook or a book? Comparing college students' reading comprehension level, critical reading, and study skills. *International Journal of Technology in Teaching and Learning* 7(2). 174–192.

Standing Committee on Language Education and Research (SCOLAR). 2003. Action plan to raise language standards in Hong Kong: Final report of language education review. https://scolarhk.edb.hkedcity.net/sites/default/files/media/ActionPlan-Final_Report%28E%29_with%20cover.pdf (accessed 15 July 2022).

Taylor, Alan M. 2002. *A meta-analysis on the effects of L1 glosses on L2 reading comprehension*. West Lafayette: Purdue University dissertation.

Taylor, Alan M. 2009. CALL-based versus paper-based glosses: Is there a difference in reading comprehension? *CALICO Journal* 27(1). 147–160.

To, Vinh. 2018. Linguistic complexity analysis: A case study of commonly-used textbooks in Vietnam. *Sage Open* 8(3). 1–13.

Vanderlinde, Ruben & Johan van Braak. 2010. The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal* 36(2). 299–316.

Wang, Feng & Michael J. Hannafin. 2005. Design-based research and technology-enhanced learning environments. *Educational Technology Research & Development* 53(4). 5–23.

Wang, Min & Keiko Koda. 2007. Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning* 57(s1). 71–98.

Wright, Tanya S. & Gina N. Cervetti. 2017. A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly* 52(2). 203–226.

Zou, Di. 2016. Comparing dictionary-induced vocabulary learning and inferencing in the context of reading. *Lexikos* 26(1). 372–390.