Jiahao Yan and Qing Ma*

Theory-supported corpus pedagogy for ESL pre-service teachers: using Parallel EAP Corpora for language learning

https://doi.org/10.1515/jccall-2024-0016
Received August 1, 2024; accepted November 5, 2024; published online January 8, 2025

Abstract: Corpus technology (CT) as an effective technology for language learning is slowly being adopted by in-service language teachers. However, there is limited focus on pre-service language teachers, who first need to experience CT and gain a basic corpus literacy as learners. Therefore, this research aimed to demonstrate the learning benefits of CT to pre-service teacher and explore their intentions for future use of CT. A corpus pedagogy supported by constructivism, socio-cultural theory, and noticing hypothesis, was designed to help three-classes of pre-service ESL teachers in Chinese Hong Kong to improve citation skills using Parallel EAP Corpora. A mixed-method design was adopted by collecting data via tests, surveys, classroom interactions and interviews. The results revealed that the participants learned to use reporting verbs for citations, acquired adequate corpus literacy, positively evaluated the theory-supported pedagogy, and expressed strong intentions to use CT for learning vocabulary and academic writing, conducting research to complete course assignments, and facilitating future teaching. The findings highlight the significance of designing corpus-based learning activities through theory-supported pedagogy and reveal that introducing CT to pre-service ESL teachers who have multiple identities as language learners, novice researchers, critical thinkers, and future teachers, can have important pedagogical implications.

Keywords: theory-supported corpus pedagogy; constructivism; socio-cultural theory; noticing hypothesis; Parallel EAP Corpora; future intentions

***Corresponding author: Qing Ma**, Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong, China, E-mail: maqing@eduhk.hk. https://orcid.org/0000-0003-3125-3513
Jiahao Yan, Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong, China, E-mail: lukeyan@s.eduhk.hk. https://orcid.org/0000-0002-5256-9915

Open Access. © 2024 the author(s), published by De Gruyter and FLTRP on behalf of BFSU. CC BY This work is licensed under the Creative Commons Attribution 4.0 International License.

1 Introduction

Johns (1990) coined the term data-driven learning (DDL) which refers to the use of corpora, the structured collections of authentic language data, to facilitate language learning. Ma et al. (2024) highlighted the pedagogical value of corpora and corpus tools, considered them as a typical technology within the context of computer-assisted language learning, and referred to them as "corpus technology" (CT). CT is defined as "the use and application of technology associated with corpus linguistics and corpora for language learning and teaching" (Ma and Rui et al. 2024, p. 462), which encompasses various approaches of direct, undirect, and paper-based DDL. Meta-analyses (Boulton 2021; Boulton and Vyatkina 2021; Chen and Flowerdew 2018; Lee et al. 2019) and individual articles (e.g., Fang et al. 2021; Ma and Fang et al. 2024; Yang et al. 2024; Yang and Mei 2024) have shown the effectiveness of CT in augmenting learners' inductive and autonomous learning. For example, Boulton (2021) reviewed 351 published studies and confirmed CT's effectiveness in various areas, including vocabulary and grammar, academic writing and error correction.

Despite much research highlighting CT's pedagogical value, its actual classroom application is somewhat limited, resulting in a research-practice gap (Boulton and Vyatkina 2021; Chambers 2019). To bridge this gap, several researchers have recently begun to offer CT-focused teacher-training programmes to in-service teachers and pre-service teachers (Chen et al. 2019; Frıgınal et al. 2020; Ma et al. 2022, 2023; Ma and Rui et al. 2024), and recent studies have documented some in-service teachers' applications of CT in their actual teaching practices (e.g., Crosthwaite et al. 2023; Ma and Rui et al. 2024). However, there are very limited studies which either focus on introducing CT to pre-service ESL teachers or investigate their attitudes and intentions to use CT in their future classroom practices.

To implement CT in teaching, pre-service teachers are expected to experience CT as learners first (Breyer 2009). If pre-service ESL teachers recognise the effectiveness of CT in their own language learning, it is highly likely that they will incorporate CT into their future teaching practices (Breyer 2009; Chambers 2019; Ma et al. 2023; Ma and Rui et al. 2024). The survey study by Ma et al. (2023) on 183 pre- and in-service teachers established a strong correlation between teachers' corpus literacy and their likelihood of utilising CT in their future teaching endeavours. Therefore, this study will develop a CT training programme focusing on addressing the learning needs of pre-service teachers who are language learners and future teachers, as well as equipping them with necessary corpus literacy, which includes 1) understanding of corpora, 2) corpora search skills, 3) analysis of corpus data, 4) advantages of (using) corpora, and 5) limitations of (using) corpora, as outlined by Ma et al. (2023).

To demonstrate the pedagogical value of CT and facilitate participants' corpus literacy, a theory-supported corpus pedagogy was adopted by referring to constructivism, socio-cultural theory and the noticing hypothesis. Constructivism emphasises the learner's role in building knowledge through their experiences and reflective practices (Bereiter 1994), while socio-cultural theory asserts that knowledge is a collaborative creation that is internalised and mediated through language (Williams 2008). Schmidt's (1990, 2001) Noticing Hypothesis states that drawing consistent attention to linguistic features can enhance the acquisition of linguistic input, while the juxtaposition of expert and learner data can enhance learners' awareness of target forms (Schmidt 2010). These theories provide theoretical underpinnings for CT (Crosthwaite and Boulton in press; O'Keeffe 2021). However, Boulton and Vyatkina (2021) noted that justifications linking CT to such underlying theories typically occurred post-intervention, rather than during the pedagogical interventions. Thus, Boulton and Vyatkina (2021) and O'Keeffe (2021) advocate for more empirical studies in CT that are driven by theoretical frameworks, aiming to reinforce the connections between CT and language learning theories.

Drawing on the theory-supported corpus pedagogy, this study focuses on a group of pre-service ESL teachers in Chinese Hong Kong to learn reporting verbs, a crucial but challenging linguistic feature in students' thesis writing (Huang 2022; Kwon et al. 2018). After graduation, these pre-service teachers will work as ESL teachers in primary and secondary schools in Chinese Hong Kong and the mainland of China. Focusing on pre-service teachers can help to promote the use of CT to a wide student population. The aims of the study were to equip pre-service teachers with the necessary literacy of using corpus for their own learning, to explore their evaluations of the pedagogy, and to investigate their intentions to use CT in the future. The data were extracted from the Parallel EAP Corpora, which incorporates an expert corpus and a learner corpus, and relevant learning materials were designed to support the pre-service teachers' learning about academic writing, particularly the use of reporting verbs to improve their citation skills. The following research questions (RQs) guided this study:

RQ1: What are pre-service teachers' learning outcomes after participating in the CT training?

RQ2: What are pre-service teachers' evaluations of the theory-supported corpus pedagogy?

RQ3: What are pre-service teachers' intentions to use CT?

2 Literature review

2.1 Constructivist learning and socio-cultural theory

Constructivist learning has gained attention amongst educational researchers because of benefits of improved learning outcomes and the promotion of independent learning (Bada and Olusegun 2015). It considers learners as active agents in knowledge acquisition and requires them to engage in discovery learning to activate higher-order cognitive skills, such as inferencing and hypothesising. However, it may not be ideal for all learners because its underlying learning mechanism places considerable cognitive loads on learners (Kirschner et al. 2006; O'Keeffe 2021). Hence, integrating socio-cultural theory into constructivist learning approaches is recommended (Brown and Palincsar 1989; Kirschner et al. 2006; Palincsar 1998).

Vygotsky's (1978) socio-cultural theory posits that knowledge is co-constructed through collaborative dialogues and social interactions with guidance and mediated support in the form of scaffolding. This theory, which emphasises the importance of providing extensive support and mediation during the learning process, has gained wide acceptance in educational technologies (Feyzi and Yasrebi 2020).

Constructivist learning and socio-cultural theory are also mainstream theoretical approaches for increasing CT's effectiveness in language learning (O'Keeffe 2021). CT provides abundant, authentic language data and intensive exposure for learners to construct their second language (L2) knowledge independently (Flowerdew 2015; Lee et al. 2019), resulting in increased recall of word meanings and collocations (Lee et al. 2019), while teacher intervention helps to alleviate the unsatisfactory effects of individual exploration (Johansson 2009), and peer collaboration enhances learners' academic performances and engagement (Huang 2011). However, there is a lack of pre-designed learning activities based on these theories (O'Keeffe 2021). The social constructivist learning theory (Brown and Palincsar 1989; Palincsar 1998), which acknowledges the overlaps and benefits of constructivist learning and socio-cultural theory, was adopted to support the CT-based pedagogical design in this study.

2.2 The noticing hypothesis and CT

Schmidt's (1990, 2001) noticing hypothesis asserts that noticing precedes understanding and facilitates the conversion of input into intake. This hypothesis is closely intertwined with the frequency of input. As Ellis (2006) concludes, language processing is largely based on frequency and probabilistic knowledge. Schmidt (2010)

also suggests that learners could notice the gaps in their language use by consciously comparing their output to input in the target language.

CT fosters learners' awareness of target forms because inductive learning, a prominent approach in CT, relies heavily on noticing language features (Flowerdew 2015). CT provides ample data pertaining to target forms, including collocations, concordance lines, contexts and frequency information. By inferring language use patterns from authentic language data, learners can internalise statistical information about language use with the assistance of CT (Crosthwaite and Boulton in press). Specifically, Boulton (2011) found that corpus training facilitated learners' development of noticing skills, and Gaskell and Cobb (2004) revealed that learners could apply noticing strategies to concordance lines for self-correction in writing.

However, existing CT studies have mainly focused on native speaker or expert writer corpora, with few having used learner corpora in classrooms (see Boulton 2021). Learner corpora, which consist of authentic language data obtained from ESL learners, give teachers and researchers valuable insights into the challenges that these learners encounter in academic writing (Granger 2004). Comparing expert and learner corpora can help students to overcome their linguistic challenges and can increase their motivation (Granger and Tribble 2014); their learning outcomes can be significantly improved (Ackerley 2017; Cotos 2014) by providing both expert and learner corpus data to enhance learners' noticing of gaps in their language use (Schmidt 2010).

2.3 Applications of CT and reporting verbs

CT for academic writing and vocabulary learning has been widely introduced to learners. For example, Chen and Flowerdew (2018) focused on 37 studies in the academic writing context, and stated that CT was effective for learners (including Chinese students) in terms of error correction, vocabulary acquisition and learning the rhetorical features of academic writing. Lee et al. (2019) reviewed 29 studies of vocabulary learning, and concluded that corpus use benefitted in-depth and long-term vocabulary acquisition; moreover, learners' corpus consultations promoted the self-driven construction of vocabulary knowledge. However, the use of CT to teach the use of reporting verbs for citations in academic writing has received little attention.

Reporting verbs help to attribute content to another source, and to present, evaluate and comment on scholars' claims, as well as to indicate whether these claims should be accepted (Hyland 2002). A review of the literature revealed several of the ESL learners' challenges when reporting citations, including selecting appropriate tenses, learning sentence patterns to integrate citations, and

understanding the evaluative nature and discourse functions of individual words (e.g., Huang 2022; Kwon et al. 2018; Lee et al. 2018; Manan and Noorizah 2014). For example, Lee et al. (2018) revealed that L2 learners tend to overuse the pattern *according to*, instead of the formal and professional reporting structure of *X + verb + that clause* (e.g., *Bigot (1992) states that …*). Huang (2022) found that undergraduate students overuse the verbs to express thinking activities (e.g., *think*, *realise*), and underuse the verbs for evidence-based argumentation (e.g., *identify*), thereby rendering their writing less convincing.

To assist learners to make appropriate lexical, syntactic and rhetorical choices with regard to reporting verbs, Callies (2016) suggested designing classroom activities that allowed learners to explore high-frequency reporting verbs independently by examining expert and learner corpora. In this paper, we discuss the implementation stage by introducing CT, including both expert and learner corpora, to pre-service teachers to teach them reporting verbs.

3 Theoretical framework for the theory-supported corpus pedagogy

A theory-supported corpus pedagogy (see Figure 1) was adopted by integrating social constructivist learning theory and the noticing hypothesis to design corpus-based learning activities. The social constructivist paradigms consist of three types of learning activities; the first type represents socio-culturally focused activities or tasks mediated by teachers. Teachers provide appropriate scaffolding, mediate learners' learning processes using curated data, and design instructional materials. The second type involves peer collaboration, which helps learners enhance the acquired knowledge and internalise corpus use. The last type involves individual

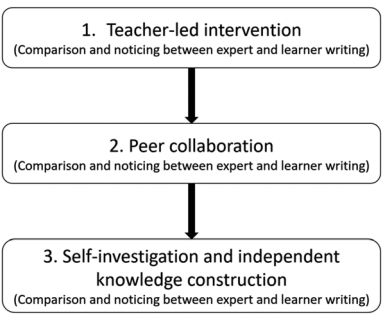

Figure 1: The theory-supported corpus pedagogy.

activities in which learners have more autonomy and independence when interacting with CT to construct language knowledge. The learning activities included both expert and learner corpora to guide the learners to notice the use of target forms and the gaps in their language use by comparing experts' and learners' academic writing.

4 Methodology

4.1 Participants and research context

In total, 88 Year 1 and Year 2 undergraduate ESL pre-service teachers of English Language Education (ELE) attending a university in Chinese Hong Kong were recruited for this study. They mainly came from Chinese Hong Kong and the mainland of China, with the age range of 17–20; some may have had one-year English tutoring experience. The majority of the participants were female students since ELE students were primarily female in that university. Their writing skills were equivalent to IELTS 5.5 or above, meeting the admission requirement of the university. The participants were enrolled in a university vocabulary course that focused on learning the meanings and contextual usage of words. Their course coordinator allocated the students to three classes, with 31, 33 and 24 students, respectively. As mentioned by their course coordinator and confirmed during the tutorial classes and interviews, they lacked prior knowledge about the use of reporting verbs for citations, a crucial linguistic feature necessary for completing many of their course assignments.

4.2 Designing features of Parallel EAP Corpora

This study incorporates recommendations from Granger and Tribble (2014) and Gilquin (2022) on integrating learner corpora into CT pedagogy. Their recommendations include representing students from a specific educational background, establishing norms for relevant genres or subject areas, and focusing on specific types of texts. Therefore, this study used the free-access Parallel EAP Corpora (Ma 2014) (interface shown in Figure 2), as it offers three unique functions corresponding to their three recommendations. Firstly, the Corpus function enables users to search for words in academic texts written not only by experienced researchers, but also by L2 university students who share the similar educational background of the participants of this study. Secondly, by searching using the Subject function, participants can choose the genre English Language Teaching (ELT) Research, which

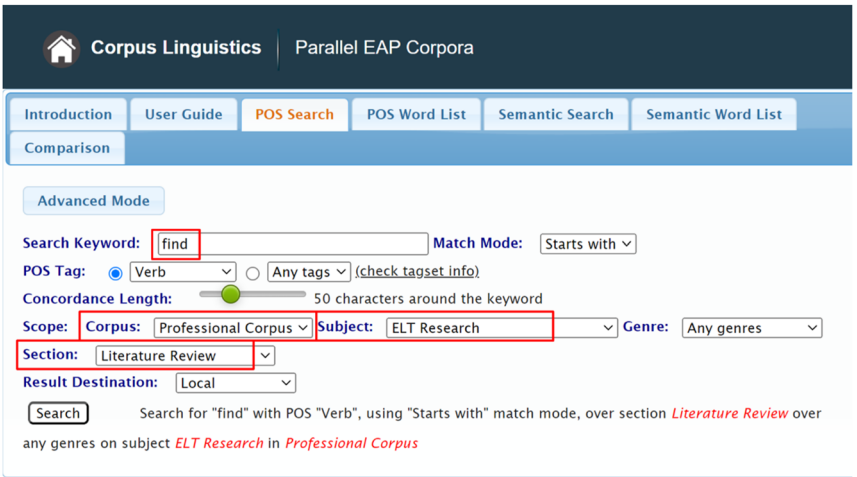

Figure 2: Interface of Parallel EAP Corpora.

is relevant to their academic domain. Thirdly, the Section function assists users to focus on the Literature Review section, which typically contains the highest number and density of citations in a paper. In addition, Parallel EAP Corpora also involve the basic functions of corpus tools, including the generation of concordance lines and collocation lists, thus enabling detailed analyses of the lexico-grammatical features of the target words.

4.3 The instructional design

A two-hour online tutorial via Zoom was designed and conducted by the first author, once for each of the three classes of pre-service teachers. Guided by the proposed theory-supported corpus pedagogy, the instructional design of the learning activities was primarily designed by the authors following the guidance provided in CT research on academic writing and development of corpus literacy (e.g., Ackerley 2017; Callies 2016; Cotos 2014; Gaskell and Cobb 2004; Johansson 2009; Ma 2024; Ma and Mei 2021; Ma et al. 2023). In addition, suggestions of the course coordinator were incorporated to make the learning materials more suitable for the pre-service teacher participants in learning reporting verbs. Figure 3 presents the sequential three-stage training based on socio-cultural and constructivist learning theories, and elaborates the integration of both the expert and the learner corpora in the Parallel EAP Corpora to enhance participants' noticing of the target items of reporting verbs.

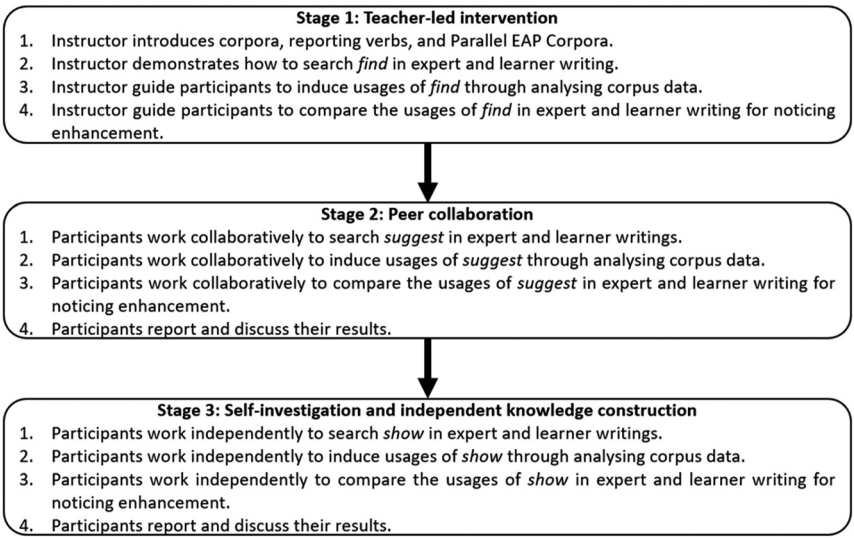

Figure 3: The three-stage training model.

Throughout the tutorial, participants were asked to directly work with Parallel EAP Corpora to learn the use of reporting verbs and develop all the components of corpus literacy, including 1) understanding of corpora, 2) corpora search skills, 3) analysis of corpus data, 4) advantages of corpora, and 5) limitations of corpora. The three most frequent reporting verbs in ELT research, *find*, *suggest* and *show* (Yan and Ma in progress), were incorporated sequentially into the three stages of teacher-led interventions, peer collaborations and individual investigations, to give participants insights into tense and voice choices, sentence patterns, and the discourse purposes underlying the use of verbs for citations when writing a literature review.

In the first stage, the instructor introduced reporting verbs and Parallel EAP Corpora. Since the participants had little prior knowledge of corpus, the instructor elaborated several key terms of corpus in this stage, including definitions of corpus, concordance lines, keyword in context, word tags, etc., to facilitate participants' understanding of corpora. This stage also involved teachers' clear demonstration and step-by-step guidance of searching the word *find* and analyse its usages based on the concordance lines and collocation lists in expert and learner genres, through which participants can develop the next two components of corpus literacy, i.e., corpus search skills and analysis of corpus data.

The participants then engaged in collaborative activities to reinforce their understanding of reporting verbs and abilities of corpus search and analysis. They worked with their groupmates to search the word *suggest*, and analyse the

concordance lines and collocation list of *suggest* to notice the usages of reporting verbs and compare the different usages between expert and learner writing to enhance their notice of the target word. This stage enhanced participants' development of corpus search skills and abilities to analyse corpus data through collaborative activities, and encouraged them to discuss the benefits and limitations of corpus with their peers.

In the third stage, participants were allowed to independently search the word *show* through Parallel EAP Corpora and analyse its usages in expert writing and learner writing. This stage helped participants to examine whether they had developed appropriate understanding of corpora and sufficient corpus search and analysis skills to implement independent learning activities. Clear guiding questions were also provided in case some participants might encounter any challenges in searching and analysing the target word. Through independent interaction with corpus, they were further encouraged to reflect on the benefits and limitations of corpus.

Furthermore, after the tutorial, supporting materials were also distributed to participants, which consists of instructions about the use of Parallel EAP Corpora to learn other reporting verbs, such as *investigate*, *note*, and *think*, and the use of using other popular corpora and corpus tools, such as COCA and AntConc.

4.4 Data collection and analysis

Five types of data (Table 1) were collected in the sequence of

(1) classroom observations,
(2) Likert-scale questionnaires,

Table 1: Data collection methods to answer the RQs.

Data types	RQ1	RQ2	RQ3
	– What are the learning outcomes?	– What are the evaluations of the theory-supported corpus pedagogy?	– What are the intentions to use CT?
Quantitative data	– Test for reporting citations – Likert-scale questionnaire (items 1–6)	–	– Likert-scale questionnaire (items 7–9)
Qualitative data	– Interview transcripts	– Classroom observations – Open-ended questions – Interview transcripts	– Interview transcripts

(3) open-ended question,
(4) interview transcripts, and
(5) tests concerning the reporting of citations.

A trained research assistant conducted the classroom observations during the online tutorials to document how the learning activities facilitated the pre-service teachers' learning progress. Using the similar template in Ma and Rui et al. (2024), the research assistant independently observed all the three tutorials, each of which lasts for 2 h.

4.4.1 Survey

The questionnaire consisted of nine 4-point Likert-scale items with three components: Items 1–3 addressed the participants' perceived learning outcomes with regard to reporting citations; Items 4–6 explored their learning outcomes following the use of Parallel EAP Corpora, and Items 7–9 investigated participants' general intentions to use CT. A total of 50 participants (41 female and 9 male students) voluntarily responded to these questions immediately after each tutorial. Nonetheless, the reliability of the survey items for the three components measured by Cronbach's α was 0.900, 0.960, and 0.923, respectively, indicating high reliability of the questionnaire.

4.4.2 Open-ended question

At the end of the questionnaire, an open-ended question, "*What do you like best about this tutorial?*" was included, and 35 participants provided their responses.

4.4.3 Interview

Due to the voluntary nature of participation, only 7 pre-service teachers (6 female and one male participants) accepted the invitation for the 40 min semi-structured individual interviews and the 20 min test on reporting citations after the tutorials. Table 2 presents detailed ethnographic information about these participants, showing that most of them aspired to become English teachers after graduation. The interviews included key questions asking (1) pre-service teachers' gains through the training, (2) their evaluations of the pedagogy, and (3) their specific intentions regarding the use of CT.

Test. Following the interviews, the participants completed the test regarding the use of reporting verbs to revise citations written by their peers. A pre-test was not administered due to the participants' lack of prior knowledge about the use of

Table 2: Ethnographic information about the interview participants.

Participants	Gender	Year of study	Career plan
S1	F	2/4	English teacher or further studies
S2	F	2/5	English teacher or other challenging job
S3	F	2/5	English teacher
S4	F	2/4	Further research in linguistics or ELT
S5	F	1/5	English teacher or government job
S6	F	1/4	English teacher or research in linguistics
S7	M	2/4	Translation or teaching

reporting verbs for citations, as confirmed during the tutorial classes and interviews. Before attending the tutorials, the participants had not received any formal instruction on reporting verbs and citations. At the beginning of the workshop, the instructor asked the participants to revise some citations similar to those in the test, but no participant provided correct answers, which indicated that they had limited knowledge of the target reporting verbs (*find*, *show*, and *suggest*) and the sentence structures for reporting (e.g., *Verb + that clause*). The use of the five types of data to address the RQs is elaborated on in the following paragraphs.

The test results, the questionnaire Items 1–6, and interview data could be used to answer RQ1. The test consisted of five tasks that required the participants to revise their peers' problematic citations, which were extracted from the learner component of the Parallel EAP Corpora. The test was designed after consulting two experienced EAP teachers and then piloted with 5 undergraduate students who did not participate in the study. Each question was graded using a scale of three points, resulting in a maximum total score of 15. The participants received three points if they addressed all the knowledge points covered in the tutorial, including vocabulary choice (e.g., *find*, *show*, and *suggest*) which express different discourse purposes, tense and voice issues, and sentence structures. One point was given for partially addressing the knowledge points. The test results were examined by the first author and checked by the research assistant. Descriptive analyses were applied to questionnaire items to examine the self-reported learning outcomes. With regard to the qualitative data analysis to examine the pre-service teachers' learning outcomes, the first author and the research assistant coded the interview data independently and resolved disagreements via discussions (Yin 2009).

The two independent coders coded the three types of qualitative data that were used to address RQ2, which investigated the participants' evaluations of the theory-supported corpus pedagogy, were classroom observation notes, open-ended data and the interview transcripts regarding the participants' evaluations of the

pedagogy. Similar to Chang (2014), the three types of data were used as references for each other, and the codes were combined for the data triangulation. Disagreements were resolved via discussions between the coders.

To answer RQ3, descriptive analyses on Items 7–9 and thematic analyses on interview data were used to investigate the participants' general and specific intentions to use CT. Two independent coders coded the interview data and reached agreements through subsequent discussions.

5 Results

5.1 Pre-service teachers' learning outcomes

Two major learning outcomes were summarised: (1) using reporting verbs for citations, and (2) acquiring necessary corpus literacy.

5.1.1 Use reporting verbs for citations

The mean score for the seven interviewees' tests regarding the reporting citations was 11.71 out of 15 (*SD* = 1.11) (Table 3). All the participants scored at least seven, meaning that they generally acquired sound knowledge of the use of the reporting verbs for citations through the tutorial.

Table 4 presents samples of the pre-service teachers' completion of the test tasks, showing that they had acquired the use of reporting verbs for citations. Task 1 requires participants to revise a sample which failed to report the citation "*(Levis 2005, p. 375)*" through a direct quotation; S6 successfully revised the citation by using the past tense of *find* and the *Verb + that* clause; the reporting verb *find*

Table 3: Test scores for reporting citations.

Participant	Score (max = 15)
S1	13
S2	11
S3	10
S4	11
S5	12
S6	12
S7	13
Mean	11.71 (*SD* = 1.11)

serves the discourse function to objectively report the prior research findings. In Task 4, S2 successfully used *suggests* to report the suggestions by "*Hyland (2004)*", in which the reporting verb *suggest* indicates that the writers may adopt the useful suggestions by "*Hyland (2004)*".

The results of Items 1–3 (Table 5) revealed that the pre-service teachers believed that the tutorial developed their knowledge about reporting citations (Item 1, M = 3.30, SD = 0.61), helped to improve their citation abilities (Item 2, M = 3.24, SD = 0.56), and was useful for their English learning and academic writing (Item 3, M = 3.18, SD = 0.56).

All interviewees mentioned that they acquired the language points: "*We can know the tense of the verb and word class*" (S6) and "*I learned when to use the present tense and past tense of suggest and find*" (S7). In summary, we are confident that the pre-service teachers successfully acquired language points regarding the use of reporting verbs for citations in literature review writing.

Table 4: Sample test results for reporting citations.

Sample test tasks	Revisions by the participants
– Task 1: accent is also intertwined with race in determining professional identity. (Levis 2005, p. 375).	1. Levis (2005) found that "accent is also intertwined with race in determining professional identity" (p. 375) (S6).
– Task 4: instruction of a genre-based approach offers writers an explicit understanding of the structure of the target texts and why they are written in the ways they use. (Note: this information should be cited from Hyland (2004).)	– Hyland (2004) suggests that instruction of a genre-based approach offers writers an explicit understanding of the structure of the target texts and why they are written in the ways they use (S2).

Table 5: Pre-service teachers' learning outcomes for reporting citations.

Questionnaire items	*N*	*M*	*SD*
1. The tutorial provides me with the basic knowledge of using reporting verbs for citations in literature review.	50	3.30	0.61
2. The tutorial provides me with guidance of how to improve citation/reporting in academic writing.	50	3.24	0.56
3. Overall, this tutorial is useful for my English learning and academic writing.	50	3.18	0.56

5.1.2 Acquiring necessary corpus literacy

Table 6 presents the results for Items 4–6, which measured the pre-service teachers' learning outcomes regarding the use of Parallel EAP Corpora. The results revealed that the tutorial designed for language learning imparted them an understanding of using Parallel EAP Corpora for learning academic writing (Item 4, $M = 3.28$, $SD = 0.61$), as well as giving the pre-service teachers additional insights into analysing academic language (Item 5, $M = 3.18$, $SD = 0.60$) and teaching vocabulary/grammar (Item 6, $M = 3.14$, $SD = 0.64$). The results demonstrated the various benefits of CT training, such as using CT for language learning, not only for studying and teaching language.

Interview transcripts revealed that the participants acquired all the aspects of corpus literacy. Firstly, all interviewees mentioned that they understood the concepts of CT and mastered search functions of Parallel EAP Corpora, as S4 noted that, "*at the beginning, I found the interface a bit messy because I didn't know the terms, like concordance, POS tag, and POS search. After the training, I found it easy to use the corpus to get the concordance lines.*"

Secondly, they learned how to analyse the search results, especially how to perform this corpus website to compare the use of expert writing and learning writing, as S5 pointed out that, "*previously, I ignored the significance of this comparison process, but now, I have learned how to use corpus to compare expert and learner writings, which is very beneficial for my learning.*"

Thirdly, they discovered the advantages of CT in helping to verify language use independently and to take ownership of their learning process, stating that they could "*take more initiatives*" (S3) and be "*highly motivated to get exposed to the academic language system*" (S1). They could "*check language use by themselves*" (S2) and search for "*what they actually want to know*" (S1). S7 further expressed that independent learning could lead to better learning outcomes: "*We can actually make*

Table 6: The learning outcomes regarding the use of CT.

	Questionnaire item	*N*	*M*	*SD*
4.	The tutorial provides me with an understanding of how to use Parallel EAP Corpora for learning academic writing.	50	3.28	0.61
5.	The tutorial provides me with an understanding of how to use Parallel EAP Corpora for analysing academic language.	50	3.18	0.60
6.	The tutorial provides me with insights of how to use Parallel EAP Corpora for vocabulary/grammar teaching.	50	3.14	0.64

our own exploration of the word, not just getting it fed to us. In this process, I believe the knowledge of the word would be more memorable".

Furthermore, S2 and S6 compared CT to other learning tools. When comparing CT to Grammarly, a website to check language use in writing, S2 said "*when only using Grammarly, you do not know why you're wrong*". S6 compared CT to dictionary use: "*dictionary only tells us the meanings, but corpus gives me many examples*", helping them learn how the word is used in context.

Fourthly, they also perceived the limitations regarding the use of CT for learning and teaching. For example, S3 expressed the concern that "*there were lots of numbers and sometimes I may just get lost*". S1 pointed out that, although they "*took some time to summarise the main sentence structures*", they were "*not so sure whether the conclusion was correct or accurate*". S7 was concerned that "*there was always difficulty in analysing the collocations and concluding usages because there were many exceptions*". Accordingly, more training is needed to enhance their corpus literacy to help them overcome the limitations associated with CT use.

5.2 Pre-service teachers' evaluations of the theory-supported corpus pedagogy

Thematic analysis of the qualitative data, including classroom observations, open-ended data and interview transcripts, revealed that the pre-service teachers enjoyed these learning activities under the theory-supported corpus pedagogy and assessed them positively. The analysis identified 4 main themes: (1) teacher scaffolding, (2) peer collaboration, (3) independent corpus search, and (4) the usefulness of Parallel EAP Corpora.

5.2.1 Teacher scaffolding

Of the 35 participants who responded to the open-ended question about their favourite tutorial activities, 16 mentioned teacher scaffolding. They appreciated the "*demonstration of using corpora*", "*clear instructions and exercises to work on*" and the "*steps in using Parallel EAP Corpora shown in the tutorial*".

The participants may have enjoyed teacher scaffolding because it addressed their difficulties in using Parallel EAP Corpora. The research assistant observed that "*some students found it difficult to use Parallel EAP Corpora and encountered many challenges*" at the beginning of the tutorial. The instructor adopted the strategy of "*demonstrating the search processes of the word find and providing clear search steps to the students*" in response, as shown in Figure 4.

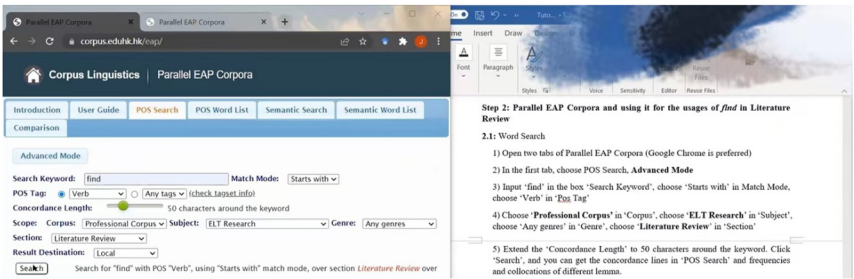

Figure 4: A screenshot that the research assistant captured during the third tutorial.

Similarly, the interview data showed the participants' concerns that using Parallel EAP Corpora might be challenging. Of note, S6 expressed that "*the practical step-by-step guidance and the very specific practice offered in the tutorial helped us to understand how to use the corpus, and I would like to apply these strategies in my future teaching*" (S6).

5.2.2 Peer collaboration

During the tutorial, group work was implemented to search for the word *suggest* to increase the participants' corpus literacy. According to the observation notes, "*the instructor entered all eight breakout rooms (Tutorial 2) and all groups heatedly discussed the usages of suggest*". In addition, five participants who answered the open-ended question said that they enjoyed peer collaboration the most, as they could collaboratively search through Parallel EAP Corpora and discuss the findings.

The interview transcripts revealed that peer collaboration was crucial for participants to build on the knowledge covered in the tutorial. S1, S3, S4 S5, and S6 noted that the collaborative activities enabled them to exchange ideas, discuss their findings and support each other's learning. S6 explained that peer discussions in breakout rooms help them "*be facilitated with the relevant skills.*"

5.2.3 Independent corpus search

Individual work was implemented during the tutorial for participants to conduct hands-on corpus search of the word *show*; as the observation notes revealed, the "*students became quicker in reporting their search results*".

According to the open-ended data, seven participants stated that their favourite activity was the independent search using a corpus. One participant stated explicitly that independent searching "*ensures we really know how to use corpora*".

The interviews indicated other possible reasons for the participants' enjoyment of independent corpus searches, which were considered to be a "*more active way of learning*" (S3) and "*an important way to internalise the knowledge and get familiarised with the use of corpus on our own*" (S1).

5.2.4 Usefulness of Parallel EAP Corpora

The observation notes further revealed how the instructor used certain strategies and allowed the participants to compare learners' writing to experts' writing through the use of Parallel EAP Corpora, which could "*motivate students by allowing them to explore the issues found in their peers' writing*". The observation notes also revealed that the pre-service teachers were most attentive when comparing expert and learner corpora, where they "*actively shared their thoughts by unmuting themselves and using the chat box*".

Similarly, the open-ended data demonstrated the participants' appreciation and enjoyment of the comparison of experts' and learners' writing. As one participant stated, "*I like the part comparing learners and experts' usage of some reporting verbs because we can avoid making the same mistakes.*"

All seven interviewees emphasised that using Parallel EAP Corpora to compare expert and learner corpora was extremely useful for learning academic writing, as it helped them to "*see the differences between the scholars and students*" (S1), to "*know how to write in a way more like a scholar*" (S1), and to "*gain a very deep impression*" about the "*accurate collocations and correct grammar*" (S5). S2 even stated that "*if I had learned this corpus earlier, I think my writing ability would be better*".

In summary, the participants enjoyed the learning activities in the pedagogy and evaluated them positively, as the first three themes pertained to learning activities based on the social constructivist learning theory and the last theme reflected the integration of both expert and learner corpora data in Parallel EAP Corpora for enhanced noticing.

5.3 Pre-service teachers' intentions regarding the use of CT

The results for Items 7–9 (Table 7) showcase pre-service teachers' three general intentions regarding the use of corpora, which were for future learning (M = 3.12, SD = 0.59), research (M = 3.10, SD = 0.54) and teaching (M = 3.16, SD = 0.55) purposes. These results highlight the tremendous value achieved through a single tutorial, which increased the pre-service teachers' interest and demonstrated that corpora can be utilised not only for learning purposes, but also for conducting language analyses and facilitating language teaching.

Table 7: Pre-service teachers' general use intentions.

Questionnaire item	*N*	*M*	*SD*
7. I am willing to use Parallel EAP Corpora or similar corpora for learning academic writing.	50	3.12	0.59
8. I am willing to use Parallel EAP Corpora or similar corpora for future language analyses.	50	3.10	0.54
9. I am willing to use Parallel EAP Corpora or similar corpora for future vocabulary/ grammar teaching.	50	3.16	0.55

The thematic analysis of the interview transcripts further revealed three specific intentions regarding the use of CT:

(1) improving vocabulary learning and academic writing,
(2) conducting research to complete course assignments, and
(3) facilitating future teaching.

The participants also mentioned their concerns about the use of CT in learning and future teaching, and expressed a strong desire for more CT tutorials.

5.3.1 Improving vocabulary learning and academic writing

Inspired by the valuable learning outcomes, all the interviewees expressed strong willingness to further explore the benefits of CT for learning vocabulary and improving their academic writing. Before participating in the tutorial, they were unsure whether they could use the items that they had learned correctly; as S3 explained, they were "*unable to identify these errors*". However, by using corpus resources, they "*could look for some accurate collocations and the correct grammar*" (S1). CT could help to facilitate their academic writing when they "*were not so sure how to write in a proper way or use a certain kind of word in a certain kind of academic context*" (S1). S4 further underscored CT's benefits of making learning and writing more effective, and said "*using corpus can shorten the time and make things more effective because we don't need to wait for the teachers' responses*".

5.3.2 Conducting research to complete course assignments

The participants recognised that the process of learning target linguistic features via CT was inherently intertwined with analysing the language data embedded in the corpora; as S3 stated that "*the way that I learned to write my academic essay is*

actually like doing linguistic research" and S4 stressed that she "*had the feeling that I was proud that I made contributions to the analysis*".

Inspired by this, they expressed their intention to employ corpora to conduct linguistic research for their research reports, course assignments and their honours projects. The specific purposes included the analyses of word use or sentence structures (S2 and S4), conducting discourse analyses of conversations (S7), investigating language variation and change (S3, S6), and even conducting comparative studies across different languages (S6). In particular, S1 mentioned the willingness to write an essay discussing how to use CT to improve students' learning: "*I will use corpora and write essays about how to help students with their learning by using this corpora system*". This finding is particularly encouraging, as we are clearly nurturing future CT researchers.

5.3.3 Facilitating future teaching

As pre-service teachers, most of the participants will become teachers in educational institutions in Chinese Hong Kong and the mainland of China. They had previously learned the term CT, but did not fully grasp its pedagogical value; as S5 claimed, "*when I first encountered the corpus, I thought it was only a tool for linguistics, but now I know it is also a tool for teachers for language teaching*". After experiencing the effectiveness of CT in their own learning during the tutorial, they expressed strong intentions to use CT in their future teaching to help their students to experience "*another way to learn English*" (S2) because CT is "*a very interesting method for them*" (S3).

Furthermore, both S2 and S5 expressed that CT could enhance their target students' learning motivation in the future. S5 observed that students are passive learners in traditional classrooms, waiting for the "*teachers to tell them the right structure*", while CT is "*a more active way of learning*". S2 observed that Chinese Hong Kong students were "*passive to learn English*" while corpus can help them "*learn independently and be more responsible for themselves*".

Many of the pre-service teachers proposed specific contexts in which they could use CT to facilitate their future teaching. S5 believed that teachers could use CT to design fill-in blank tasks based on target words, and "*ask students to fill in the blank using the correct form*". S4 mentioned that she would like to use corpus in future academic writing classes, while S3 believed that CT was effective for helping learners to correct the "*common mistakes made by primary and secondary students*". Both S3 and S6 indicated that the frequency data produced by CT could "*help students be more aware of the possible mistakes*" (S6), which is in line with Schmidt's (2001) noticing hypothesis. Furthermore, S4 and S6 said that comparing expert and learner

corpora could allow teachers to "*predict what problems students may encounter in writing*" (S4) and "*students' writing errors*" (S6).

S1, S2 and S3 began to think critically about whether CT was suitable for their future students; they believed that students with higher English proficiency, such as tertiary and senior secondary students, could engage "*actively and deeply*" (S2) with CT, while younger learners may experience difficulties. For example, S1 mentioned that "*for senior secondary classrooms, I think it is good to include corpus in their lessons, but for low-level students, it's quite hard*". S3 thought that it would be useful to introduce corpora in tertiary education, but "*for secondary or primary education, corpus might be too complicated for students*". Specifically, S2 wondered whether there were any corpora that were suitable for lower-level students.

5.3.4 Desire for additional tutorials

At the end of the interviews, all the participants expressed eagerness to learn more about aspects of CT that were relevant to their individual learning, research or teaching purposes. They expressed strong willingness to participate in more tutorials and hoped that their teachers would provide more opportunities or resources for using corpus. For example, some participants wanted to learn about the functions of different corpora and corpus tools, such as COCA, AntConc, and even Chinse corpora. It was highly encouraging to see these pre-service teachers become extremely interested in CT and the various applications of corpus linguistics.

6 Discussion

6.1 Effectiveness for learning the use of reporting verbs and developing corpus literacy

The tutorial addressed pre-service ESL teachers' learning needs regarding the use of reporting verbs for citations, following the long-standing recommendation that source work should be the focus of writing instruction for early year undergraduates (Kwon et al. 2018). This study demonstrated the effectiveness of employing both expert/native and learner corpora to learn reporting verbs, and reinforced what established language learning theories can offer to CT-based pedagogical practices.

Furthermore, the participants developed their corpus literacy through participating in the CT training supported by the theory-supported corpus pedagogy. They not only learned several basic concepts of corpora and how to conduct

search functions of Parallel EAP Corpora, but also were equipped with the skills of analysing corpus data and developed a deeper understanding of the advantages and limitations of CT. As shown in Ma et al. (2023), all of these are crucial aspects of developing corpus literacy, which further influences pre-service teachers' future intention to adopt CT in classroom teaching.

6.2 Significance of the theory-supported corpus pedagogy

Corpus linguists propose that constructivist learning, socio-cultural theory and the noticing hypothesis can be linked to CT to make students' learning more effective (Crosthwaite and Boulton in press; Flowerdew 2015; O'Keeffe 2021). In this study, we successfully integrated these language learning theories into corpus-based learning designs, marking a significant advancement over many prior studies in corpus linguistics that often lacked robust, theory-supported pedagogical frameworks (Boulton and Vyatkina 2021; O'Keeffe 2021). This integration highlights our contribution to enhancing the theoretical grounding and practical application of corpus-based language education in several ways. First, our study shows that the social constructivist learning theory can help participants learn corpus use and target items in a sequential and well-organised process through teacher-led intervention, peer collaboration, and independent learning. Our research demonstrates how active learning combines the advantages of constructivist learning – which focuses on students' active engagement in inferencing and hypothesizing, as outlined by Bada and Olusegun (2015) and Brown and Palincsar (1989) – with the principles of socio-cultural theory as well as noticing hypothesis. The sociocultural theory emphasizes the critical role of teachers and peers in mediating and facilitating the construction of knowledge, as described by Jaramillo (1996). Together, these theories underpin our approach to integrating interactive, collaborative elements into the learning process, enhancing student learning outcomes. Second, through combining expert and learner corpus data into the activities under social cultural and constructivist learning theories, it provides evidence of how learner corpus data can be integrated for noticing enhancement, as claimed by Schmidt (2010) that providing data generated by learners themselves can augment them to notice the gaps of their own language use. Our results further revealed that the learners evaluated the experience positively and engaged actively in the learning activities, thus highlighting the significance of this theory-supported corpus pedagogy.

The successful implementation of the theory-supported corpus pedagogy is firmly grounded in the data and functionalities provided by the Parallel EAP Corpora. This resource offers abundant writing samples of literature reviews,

enabling participants to explore the use of reporting verbs for citations in the Literature Review section. Its user-friendly interface minimizes the technical demands on teachers when preparing clear scaffolding materials and facilitates students' ability to conduct independent corpus searches and analyses efficiently, thereby helping them to construct the language use patterns. Additionally, Parallel EAP Corpora allows users to compare both expert and learner academic writing samples, which fosters their noticing of the target items. Given the positive feedback from the participants, it is recommended that future EAP studies and practices incorporate Parallel EAP Corpora or design similar corpora to enhance students' advanced academic writing skills.

6.3 The benefits of introducing CT to pre-service teachers

After recognizing the benefits of Corpus Technology (CT) for their own learning, acquiring the necessary corpus literacy, and observing the importance of theory-supported corpus pedagogy, the pre-service teachers began to critically reflect on how and whether CT could be implemented in classroom teaching to enhance their future students' language learning. Römer (2006) suggests that incorporating CT into more educational contexts should focus on teacher education, and should target initial teacher training at universities, practicing teachers and advanced learners. However, despite this on-going call, CT has had limited influence on language teachers' classroom practices (Boulton 2017; Chambers 2019; Ma et al. 2022). The findings of this study provide new insights to address this issue by demonstrating that integrating CT into pre-service teachers' language learning can also serve the purposes of pre-service teacher training.

Furthermore, this study revealed that CT could facilitate pre-service teachers' academic development as novice researchers by expanding their methodological toolbox for language research (Paquot and Gries 2021), thus enabling them to contribute to the fields of applied linguistics and language education (Cheng et al. 2003; Mussetta and Vartalatis 2018).

In summary, many pre-service ESL teachers have multifaceted identities as language learners, future teachers and novice researchers, shaping their experiences and interactions with CT. They are also critical thinkers who reflect on the extent to which CT can support their own language learning and their future teaching in different contexts. Focusing on these pre-service teachers facilitates the future development of CT in the domains of language research and education.

7 Pedagogical implications

This study suggests two main pedagogical implications based on the theory-supported corpus pedagogy. Firstly, activities involving teacher-led interaction, peer collaboration and independent student exploration are recommended in CT educational practices. Well-defined instructions and scaffolding materials supplied by teachers can prevent students from becoming overwhelmed by the vast corpus data, enable the efficient utilisation of CT, and equip students with the necessary skills for autonomous knowledge construction (Cobb and Boulton 2015; Flowerdew 2015). Peer collaboration enables learners to share knowledge and fosters collaborative knowledge construction (Zheng et al. 2021). It also motivates students to use corpora for learning, familiarises them with CT, enhances their language analysis skills, and improves their acquisition of target items (Cheng et al. 2003; Huang 2011; O'Keeffe 2021). Independent work using CT to construct language knowledge promotes students' independent learning abilities and long-term learning outcomes (Flowerdew 2015; Lee et al. 2019; O'Keeffe 2021).

Secondly, this study revealed the value of using both expert and learner corpora, such as the Parallel EAP Corpora, to enhance learners' noticing of target forms, to make the target features more prominent, to capture the learners' attention, and to make the target forms salient (O'Keeffe 2021). In line with Ackerley (2017); Cotos (2014); Yan and Ma (2024), the findings of this study suggest that the combination of expert and learner corpora helps learners to gain a deeper understanding of vocabulary use, collocations, sentence structures and the discourse conventions of reporting verbs. Therefore, we encourage more corpus resources integrating both expert and learner corpora for learning purposes, and recommend that educators should include these corpora in their teaching practice.

8 Conclusions

This study reported on the design and implementation of a theory-supported corpus pedagogy for early year undergraduate pre-service ESL English teachers in Chinese Hong Kong to use Parallel EAP Corpora to learn the use of reporting verbs for citations when writing literature reviews. The pedagogy involved a three-stage training model based on the social constructivist learning theory and the integration of both expert and learner corpora to increase noticing. The results proved participants' enhanced corpus literacy, and revealed the pre-service teachers' positive attitudes towards CT and the use of Parallel EAP Corpora, their enjoyment of the instructional activities, and their multiple intentions to use CT for future learning,

teaching and research. The study addressed pre-service teachers' academic development as language learners, future teachers, novice researchers and critical users of CT. The findings provide significant pedagogical implications for designing learning activities based on socio-constructivism and integrating expert and learner corpora for enhanced noticing, thus laying the foundation for future research to explore the effectiveness of CT in the development of advanced academic writing skills.

However, this study has limitations due to its predominantly female pre-service teacher participants and a small sample size of only seven students who completed the test, attributed to the voluntary nature of participation. To enhance the generalisability of the findings, future studies should aim to recruit a larger and more diverse group of participants, including more male students. Additionally, longitudinal research is recommended to track the actual usage of CT in learning, teaching, and research activities. Beyond lexico-grammatical knowledge, proficient literature review writing also demands mastery of discourse functions and critical thinking skills. Future research should therefore focus on these advanced aspects of academic writing.

Acknowledgement: We are grateful for the Research Grants Committee of Hong Kong SAR (GRF 18600123) and The Education University of Hong Kong (CRAC 04A32) which provided funding for supporting this research.
Author contributions: Jiahao Yan: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Visualization; Roles/Writing – original draft; and Writing – review & editing. Qing Ma: Formal analysis; Funding acquisition; Methodology; Resources; Software; Supervision; Validation; Visualization; and Writing – review & editing.
Research funding: This article is supported by the CRAC project (04A32) funded by the Education University of Hong Kong and GRF 18600123 funded by Research Grants Committee, Hong Kong SAR.

References

Ackerley, Katherine. 2017. Effects of corpus-based instruction on phraseology in learner English. *Language, Learning and Technology* 21. 195–216.

Bada, Steve Olusegun & Steve Olusegun. 2015. Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education* 5(6). 66–70. http://doi.10.9790/7388-05616670.

Bereiter, Carl. 1994. Constructivism, socioculturalism, and Popper's world 3. *Educational Researcher* 23(7). 21–23.

Boulton, Alex. 2011. Language awareness and medium-term benefits of corpus consultation. In Ana Gimeno Sanz (ed.), *New trends in corpus assisted language learning: Working together*, 39–46. Madrid: Macmillan ELT.

Boulton, Alex. 2017. Corpora in language teaching and learning. *Language Teaching* 50(4). 483–506.

Boulton, Alex. 2021. Research in data-driven learning. In Pascaul Pérez-Paredes & Geraldine Mark (eds.), *Beyond the concordance: Corpora in language education*, 9–34. Amsterdam/Philadelphia: John Benjamins.

Boulton, Alex & Nina Vyatkina. 2021. Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language, Learning and Technology* 25(3). 66–89.

Breyer, Yvonne. 2009. Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning* 22(2). 153–172.

Brown, Ann L. & Annemarie S. Palincsar. 1989. Guided, cooperative learning and individual knowledge acquisition. In Lauren Resnick (ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, 393–451. Hillsdale, NJ: Erlbaum.

Callies, Marcus. 2016. Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. In Rolf Kreyer, Steffen Schaub & Barbara Ann Güldenring (eds.),*Angewandte Linguistik in Schule und Hochschule*, 391–415. Frankfurt, Germany: Peter Lang.

Chambers, Angela. 2019. Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching* 52(4). 460–475.

Chang, Ji-Yeon. 2014. The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL* 26(2). 243–259.

Chen, Meilin & John Flowerdew. 2018. A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics* 23(3). 335–369.

Chen, Meilin, John Flowerdew & Laurence Anthony. 2019. Introducing in-service English language teachers to data-driven learning for academic writing. *System* 87. 102148.

Cheng, Winnie, Martin Warren & Xu Xun-Feng. 2003. The language learner as language researcher: Putting corpus linguistics on the timetable. *System* 31(2). 173–186.

Cobb, Thomas & Alex Boulton. 2015. Classroom applications of corpus analysis. In Douglas Biber & Randi Reppen (eds.), *Cambridge handbook of corpus linguistics*, 478–497. Cambridge, UK: Cambridge University Press.

Cotos, Elena. 2014. Enhancing writing pedagogy with learner corpus data. *ReCALL* 26(2). 202–224.

Crosthwaite, Peter & Alex Boulton. In press. Expanding the boundaries of data-driven learning. In Henry Tyne, Mireille Bilger, Laurie Buscail, Maï Leray, Niall Curry & Carmen Pérez-Sabater (Dir.) (eds.), *Discovering language: Learning and affordance*. Frankfurt, Germany: Peter Lang.

Crosthwaite, Peter, Luciana & David Wijaya. 2023. Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning* 36(7). 1392–1420.

Ellis, Nick C. 2006. Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27. 164–194.

Fang, Liuqin, Qing Ma & Jiahao Yan. 2021. The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students. *Journal of China Computer-Assisted Language Learning* 1(1). 80–109.

Feyzi, Behnagh, Reza & Sepideh Yasrebi. 2020. An examination of constructivist educational technologies: Key affordances and conditions. *British Journal of Educational Technology* 51(6). 1907–1919.

Flowerdew, Lynne. 2015. Data-driven learning and language learning theories: Whither the twain shall meet. In Alex Boulton & Agnieszka Leńko-Szymańska (eds.), *Multiple affordances of language corpora for data-driven learning*, 15–36. Amsterdam, Netherlands: John Benjamins.

Frıgınal, Eric, Peter Dye & Matthew Nolen. 2020. Corpus-based approaches in language teaching: Outcomes, observations, and teacher perspectives. *Boğaziçi Üniversitesi Eğitim Dergisi* 37(1). 43–68.

Gaskell, Delian & Thomas Cobb. 2004. Can learners use concordance feedback for writing errors? *System* 32(3). 301–319.

Gilquin, Gaëtanelle. 2022. Written learner corpora to inform teaching. In Reka Jablonkai & Eniko Csomay (eds.), *The Routledge handbook of corpora and English language teaching and learning*, 281–295. London: Routledge.

Granger, Sylviane. 2004. Computer learner corpus research: Current status and future prospects. In Ulla Connor & Thomas A. Upton (eds.), *Applied corpus linguistics*, 123–145. Leiden, Netherlands: Rodopi.

Granger, Sylviane & Chris Tribble. 2014. Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In Sylviane Granger (ed.), *Learner English on computer*, 199–209. London: Routledge.

Huang, Li-Shih. 2011. Corpus-aided language learning. *ELT Journal* 65(4). 481–484.

Huang, Yueyue. 2022. A corpus-based study on the semantic use of reporting verbs in English majors' undergraduate thesis writing. *Journal of Language Teaching and Research* 13(6). 1287–1295.

Hyland, Ken. 2002. Activity and evaluation: Reporting practices in academic writing. In John Flowerdew (ed.), *Academic discourse*, 115–130. London: Longman.

Jaramillo, James A. 1996. Vygotsky's sociocultural theory and contributions to the development of constructivist curricula. *Education* 117(1). 133–141.

Johansson, Stig. 2009. Some thoughts on corpora and second-language acquisition. In Karin Aijmer (ed.), *Corpora and language teaching*, 33–44. Amsterdam, Netherlands: John Benjamins.

Johns, Tim. 1990. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria* 10. 14–34.

Kirschner, Paul, John Sweller & Richard Clark. 2006. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist* 41(2). 75–86.

Kwon, Monica Heejung, Shelley Staples & R. Scott Partridge. 2018. Source work in the first-year L2 writing classroom: Undergraduate L2 writers' use of reporting verbs. *Journal of English for Academic Purposes* 34. 86–96.

Lee, Joseph J., Chris Hitchcock & J. Elliott Casal. 2018. Citation practices of L2 university students in first-year writing: Form, function, and stance. *Journal of English for Academic Purposes* 33. 1–11.

Lee, Hansol, Mark Warschauer & Jang Ho Lee. 2019. The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics* 40(5). 721–753.

Ma, Qing. 2014. Parallel EAP Corpora specialising in English language studies/education and their implications. In *Paper presented at second asia pacific corpus linguistics conference (APCLC 2014), March 7–9*. Hong Kong: The Hong Kong Polytechnic University.

Ma, Qing. 2024. Corpus-based language pedagogy for pre-service and in-service teachers: Theory, practice, and research. In Peter Crosthwaite (ed.), *Corpora for language learning*, 174–186. London: Routledge.

Ma, Qing, Ming Ming Chiu, Shanru Lin & Norman B. Mendoza. 2023. Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL* 35(1). 19–39.

Ma, Qing & Fang Mei. 2021. Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning* 1(1). 177–190.

Ma, Qing, Fang Mei & Bojie Qian. 2024. Exploring EFL students' pronunciation learning supported by corpus-based language pedagogy. *Computer Assisted Language Learning*. 1–27. https://doi.org/10.1080/09588221.2024.2432965.

Ma, Qing, Jinlan Tang & Shanru Lin. 2022. The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning* 35(9). 2731–2760.

Ma, Qing, Rui Yuan, Lok Ming Eric Cheung & Jing Yang. 2024. Teacher paths for developing corpus-based language pedagogy: A case study. *Computer Assisted Language Learning* 37(3). 461–492.

Manan, Nor Azma & Mohd Noor Noorizah. 2014. Analysis of reporting verbs in Master's theses. *Procedia-Social and Behavioral Sciences* 134. 140–145.

Mussetta, Mariana & Andrea Vartalatis. 2018. Writing across the curriculum in ELT training courses: A proposal using data-driven learning in disciplinary assignments. *International Journal of Teaching and Learning in Higher Education* 30(2). 300–307.

O'Keeffe, Anne. 2021. Data-driven learning–a call for a broader research gaze. *Language Teaching* 54(2). 259–272.

Palincsar, Annemarie S. 1998. Social constructivist perspectives on teaching and learning. *Annual Review of Psychology* 49(1). 345–375.

Paquot, Magali & Stefan Th Gries (eds.). 2021. *A practical handbook of corpus linguistics*. Cham, Switzerland: Springer Nature.

Römer, Ute. 2006. Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik* 54(2). 121–134.

Schmidt, Richard. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11(2). 129–158.

Schmidt, Richard. 2001. Attention. In Peter Robinson (ed.), *Cognition and Second Language Instruction*, 3–32. Cambridge, UK: Cambridge University Press.

Schmidt, Richard. 2010. Attention, awareness and individual differences in language learning. In *Proceedings of CLaSIC 2010*. Singapore: National University of Singapore.

Vygotsky, Lev Semenovich (ed.). 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Williams, Jessica. 2008. The speaking-writing connection in second language and academic literacy development. In Diane Dewhurst Belcher & Alan Hirvela (eds.), *The Oral/Literate Connection: Perspectives on L2 Speaking, Writing and Other Media Interactions*, 10–25. Ann Arbor MI: University of Michigan Press.

Yan, Jiahao & Qing Ma. 2024. Developing advanced citation skills: A mixed-methods approach to corpus technology training for novice researchers. *Journal of English for Academic Purposes* 72. 101451.

Yan, Jiahao & Qing Ma. In progress. *Reporting verbs for citation practices: A corpus-based study on research writing in English language teaching*.

Yang, Yingying, Lin Chen & Xumin Tian. 2024. Student perceived effectiveness of task-based instructional design of data-driven synonym learning featuring "mini-lecture". *Journal of China Computer-Assisted Language Learning* 4(1). 74–114.

Yang, Jing & Fang Mei. 2024. Promoting critical reading instruction in higher education: A three-step training scheme facilitated by using corpus technology. *Journal of China Computer-Assisted Language Learning* 4(1). 115–142.

Yin, Robert K. 2009. How to do better case studies. In Leonard Bickman & Debra J. Rog (eds.), *The SAGE handbook of applied social research methods*, 254–282. Thousand Oaks, CA: SAGE Publications Ltd.

Zheng, Lanqin, Zhong Lu, Jiayu Niu, Miaolang Long & Jiayi Zhao. 2021. Effects of personalized intervention on collaborative knowledge building, group performance, socially shared metacognitive regulation, and cognitive load in computer-supported collaborative learning. *Educational Technology & Society* 24(3). 174–193.

Bionotes

Jiahao Yan
Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong, China
lukeyan@s.eduhk.hk
https://orcid.org/0000-0002-5256-9915

Jiahao Yan is a PhD candidate of the Department of Linguistics and Modern Linguistics Studies, The Education University of Hong Kong, Hong Kong, China. His research interests include English for Academic Purposes, corpus linguistics, computer-assisted language learning, and GenAI in language learning.

Qing Ma
Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong, China
maqing@eduhk.hk
https://orcid.org/0000-0003-3125-3513

Dr Qing Ma is Associate Professor at the Department of Linguistics and Modern Linguistics Studies, The Education University of Hong Kong, Hong Kong, China. Her main research interests include second language vocabulary acquisition, corpus linguistics, corpus-based language pedagogy (CBLP), computer-assisted language learning (CALL) and mobile-assisted language learning (MALL).