ReCALL (2026), 38: 1, 94–110
doi:10.1017/S0958344025100347

RESEARCH ARTICLE

TPACK for corpus technology: TESOL student teachers' growth and self-efficacy in independent language learning and teaching

Qing Ma
The Education University of Hong Kong, Hong Kong SAR, China

Jiahao Yan
The Education University of Hong Kong, Hong Kong SAR, China

Peter Crosthwaite
The University of Queensland, Australia

Chi-Kin John Lee
The Education University of Hong Kong, Hong Kong SAR, China

Abstract

The adoption of corpus technology in school classroom settings remains limited, largely due to insufficient technological pedagogical content knowledge (TPACK) training for pedagogical corpus use. To address this gap, we investigated how teacher education in corpus-based language pedagogy (CBLP), a subdomain of TPACK for corpus technology tailored to language teachers, influenced student TESOL teachers' self-efficacy for independent language learning and teaching. Employing a mixed-methods approach, including a CBLP training intervention ($n = 120$), survey data ($n = 96$), and interviews ($n = 8$) with student teachers at a university in Hong Kong SAR, China, the research validates a theoretical model through confirmatory factor analysis and structural equation modelling. Results demonstrate that corpus literacy (CL) is foundational for effective CBLP implementation and development of independent learning self-efficacy, which in turn fosters innovative, resource-rich instructional strategies. CBLP also enhances teachers' self-efficacy for student engagement, fostering more interactive and motivating classrooms. These findings emphasise the value of embedding CL and CBLP within TESOL teacher-education programmes to prepare future language teachers for self-efficacy within dynamic, technology-enhanced classrooms.

Keywords: TESOL student teachers; TPACK; corpus-based language pedagogy; independent language learning self-efficacy; independent language teaching self-efficacy

1. Introduction

A key learning challenge for learners of a foreign language is limited exposure to authentic language data input. To address this gap, applied linguists have sought to employ language

Corresponding author: Qing Ma; Email: maqing@eduhk.hk

Cite this article: Ma, Q., Yan, J., Crosthwaite, P. & Lee, C-K.J. (2026). TPACK for corpus technology: TESOL student teachers' growth and self-efficacy in independent language learning and teaching. *ReCALL* 38(1): 94–110. https://doi.org/10.1017/S0958344025100347

© The Author(s), 2025. Published by Cambridge University Press on behalf of EUROCALL, the European Association for Computer-Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

corpora and corpus technology as pedagogical aids, now widely recognised for their potential to foster inductive, discovery-based and, ultimately, independent language learning, commonly known as data-driven learning (DDL) (Boulton, 2017; Johns, 1991). However, while corpus tools are increasingly prevalent in language research, and despite a wealth of research on pedagogical corpus use in higher education, corpora remain unfamiliar to many primary and secondary school teachers given the limited focus on their use in teacher-education programmes (Boulton, 2017; Callies, 2019; Chambers, 2019).

Existing research (Crosthwaite, Luciana & Wijaya, 2023; Ma, Tang & Lin, 2022; Ma, Yuan, Cheung & Yang, 2024a; Schmidt, 2023) confirms language teachers' insufficient technological pedagogical content knowledge (TPACK; Mishra & Koehler, 2006) for integrating corpus technology into their ongoing professional development, with teachers struggling to understand why and how to structure corpus activities within their lesson planning. TPACK extends Shulman's (1987) concept of pedagogical content knowledge (PCK) by integrating technology. In the context of corpus technology, TPACK involves combining language pedagogy and technological competence in corpus use, forming what Ma *et al.* (2022) described as corpus-based language pedagogy (CBLP). This approach emphasises the importance of corpus literacy (CL; Mukherjee, 2004) involving technological content knowledge essential for operating corpus tools.

Another key concept for successful CBLP is teachers' teaching ability, or teachers' self-efficacy (TSE; Tschannen-Moran, Hoy & Hoy, 1998). TSE refers to teachers' beliefs in their ability to organise and conduct teaching successfully and significantly influences teaching effectiveness and student learning outcomes (Caprara, Barbaranelli, Steca & Malone, 2006; Tschannen-Moran & Hoy, 2001a).

However, there remains a gap in understanding pre-service TSE for independent teaching involving CBLP. Given that TPACK and TSE are potentially correlated (Joo, Park & Lim, 2018; Lee & Tsai, 2010), this study aims to explore how student TESOL teachers' TPACK for corpus technology relates to their self-efficacy in independent language learning and teaching – a link not yet fully examined in the DDL or CBLP literature. This study specifically focuses on non-native TESOL pre-service teachers from Hong Kong who were both language learners and prospective language teachers.

2. Corpora and language learning

Corpora and corpus technology offer rich language resources that enable students to explore language inductively and address their own language queries in situ (Boulton, 2017; Johns, 1991). The benefits of using corpora and corpus technology in language learning are now widely recognised via DDL, as advocated by Johns (1991) and others (Boulton, 2011, 2017; O'Keeffe, 2021). DDL promotes inductive, discovery-based, and autonomous language learning (Boulton, 2017; Crosthwaite & Boulton, in press; Lee, Warschauer & Lee, 2019). Research suggests that corpus consultation supports the development of various language skills such as vocabulary knowledge (Huang & Ma, 2025; Lee *et al.*, 2019), academic writing (Chen & Flowerdew, 2018), rhetorical skills (Yan & Ma, 2024), and syntactic features (Lee *et al.*, 2019). As O'Keeffe (2021) noted, "the long-held and widespread consensus is that the core pedagogical benefit of corpus use lies in its potential to encourage learners to construct their L2 knowledge independently by exploring the linguistic data from corpus input" (p. 261). However, empirical research on the impact of corpus technology on learners' ability to learn independently remains limited (Boulton, 2011; Crosthwaite & Boulton, in press).

3. Applications of corpora in language teaching

Despite a wealth of studies on corpora for language learning, evidence of ordinary teachers' use of corpora in mainstream classroom settings remains limited (Boulton, 2017; Callies, 2019;

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

Crosthwaite *et al*., 2023). Several barriers hinder the adoption of corpora in classroom teaching. First, few teachers or student teachers have received adequate corpus technology training, leaving them unaware of its classroom potential (Boulton, 2017; Callies, 2019; Chambers, 2019; Karlsen & Monsen, 2020). Second, even after training, some teachers lack the pedagogical skills to use corpora effectively when teaching (Chung, Crosthwaite, Cao & de Carvalho, 2024); many teachers view corpora solely as a research tool rather than a valuable classroom resource (Karlsen & Monsen, 2020). Third, most corpus research is conducted in tertiary contexts with a focus on English for academic or specific purposes, often by researchers with advanced expertise in corpus-based research (Chen & Flowerdew, 2018). Moreover, limited attention has been paid to developing corpora for younger learners attending primary and secondary schools (Meunier, 2019), further discouraging teachers from adopting corpus-based approaches (Boulton, 2017; Callies, 2019).

Scholars have sought to address these challenges by promoting CL amongst pre-service and in-service teachers. Mukherjee (2004) defined CL as the ability to understand what a corpus is, its affordances and limitations, how to analyse concordance lines, and how to interpret corpus data effectively. While this foundational knowledge is essential, it mainly prepares teachers to understand corpus tools, not necessarily to teach with them. DDL refers to the direct use of corpora by learners themselves to discover language patterns independently (Johns, 1991), which can be seen as an application of CL (Farr & Leńko-Szymańska, 2024), while the pedagogical integration of corpus technology into classroom teaching is less emphasised. In this regard, DDL alone does not equip teachers with the pedagogical strategies required to scaffold corpus use in diverse classroom contexts.

To address this gap, Ma *et al*. (2022, 2024a, 2024b) introduced the concept of CBLP, a more comprehensive framework that combines corpus technology with language teaching expertise. CBLP represents a targeted application of TPACK for language education, enabling teachers not only to use corpora but also to integrate them meaningfully into instruction. CBLP thus extends beyond both CL and DDL by focusing on pedagogical integration. Training in CBLP is particularly important for pre- and in-service teachers in school contexts, where corpus tools remain underutilised (Boulton, 2017; Callies, 2019; Chambers, 2019) without sufficient pedagogical support.

4. CBLP: TPACK for corpus technology

The TPACK framework, developed by Mishra and Koehler (2006), extended Shulman's (1987) PCK by including technological knowledge. Traditionally, teacher education has focused on either subject content or general pedagogy, often treating them as separate domains (Shulman, 1986). Shulman's PCK bridged this gap, emphasising the need to blend content and pedagogy for effective teaching. TPACK further integrated technology with PCK, guiding how to use specific technologies to teach particular subjects.

It is contended that TPACK in general domains may not sufficiently prepare teachers to use specific educational technologies for particular subjects (Graham, Borup & Smith, 2012; Lee & Tsai, 2010; Tseng, Chai, Tan & Park, 2022). Thus, specialised training linking specific technologies (such as corpus tools) to specific subjects (such as English language teaching) is required (Graham *et al*., 2012). This specific form of language pedagogy refers to CBLP, defined as the ability to integrate corpus-based technologies for effective classroom language teaching (Ma *et al*., 2022). Corpus technology has been defined as the use and application of technology associated with corpus linguistics and corpora for language learning and teaching (Ma *et al*., 2024a). In this context, corpus technology – representing a specific form of technological content knowledge – is combined with language pedagogy to form a subject-specific TPACK for corpus technology. In practice, a validated two-step CBLP training framework has been developed (Ma *et al*., 2022, 2024b), with the first phase building teachers' CL and the second enhancing their pedagogical skills for effective classroom integration.

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

However, assessing teacher TPACK can be challenging because much of it is tacit and is reflected in actions and reasoning rather than via explicit statements (Sloan, Allen, Bass & Milligan-Mattes, 2018). While Shulman's (1987) model of pedagogical reasoning offers a useful guide, it is necessary to consider how to evaluate teachers' abilities to demonstrate sufficient TPACK for CBLP in order to understand what TESOL student teachers need to know to implement CBLP effectively. For example, attempts have been made to assess teachers' lesson-planning activities involving corpora for TPACK following CBLP training (Chung *et al.*, 2024), but this approach concerns the product rather than the process. Therefore, teachers' ongoing process of learning and implementing CBLP also needs to be considered.

5. TSE and TPACK

TSE is a crucial motivational factor affecting classroom effectiveness and students' outcomes. TSE is defined as "a teacher's belief in her or his ability to organise and execute the courses of action required to successfully accomplish a specific teaching task in a particular context" (Tschannen-Moran *et al.*, 1998: 223). Thus, TSE is task- and context-specific, as it is influenced by subject matter, teaching environment, and students' characteristics. High TSE is associated with positive classroom behaviours, including enhanced teaching goals, planning, organisation, and effort (Tschannen-Moran & Hoy, 2001a; Tseng *et al.*, 2022).

TPACK and TSE are two key constructs for evaluating teachers' instructional capacity, and they share attributes linked to teaching performance. Tseng *et al.* (2022) concluded that teachers with high TSE were more skilful in integrating technology into classroom pedagogy. Studies also indicate that TPACK development is related to TSE (Joo *et al.*, 2018; Lee & Tsai, 2010; Tseng *et al.*, 2022). For example, Joo *et al.* (2018) showed that Korean student teachers with higher TPACK reported higher TSE.

However, non-native English-speaking student teachers may find it particularly difficult to develop TSE, as their linguistic abilities are often compared to native norms in teaching situations (Wyatt, 2016). Understanding the factors that contribute to their TSE is important because, in comparison to research on experienced teachers, there is limited focus on TESOL student teachers' self-efficacy in independent teaching in general and concerning CBLP in particular. Sevimel and Subasi (2018) highlighted the lack of research on how student teachers developed self-efficacy for language teaching. Therefore, this study aims to examine how student TESOL teachers progressed towards independent teaching by developing TPACK for corpus technology. Our research explored the co-development of CL, CBLP, and independent learning self-efficacy (ISE), ultimately supporting self-efficacy for independent language teaching.

6. Theoretical model, hypotheses, and research questions

Based on the existing literature, we proposed a theoretical model (see Figure 1) that connected four key constructs of CL, CBLP, ISE, and TSE. ISE and TSE should be treated as distinct constructs in the context of non-native English-speaking student TESOL teachers because the former reflects teachers' confidence in managing their own autonomous language learning, while the latter pertains to their perceived ability to facilitate their target students' learning in the classroom effectively. According to Ma *et al.* (2022), CL – understanding and using corpus technology – is the basis for developing CBLP. Thus, **Hypothesis 1** (**H1**) states that CL will have a positive influence on CBLP.

Scholars (e.g. Boulton, 2017; Crosthwaite & Boulton, in press; Lee *et al.*, 2019) have suggested that engaging with corpora fosters independent and autonomous language learning, although robust empirical support remains limited. To address this gap, our model examines how pre-service TESOL teachers' TPACK for corpus technology (encompassing both CL and CBLP) affects

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

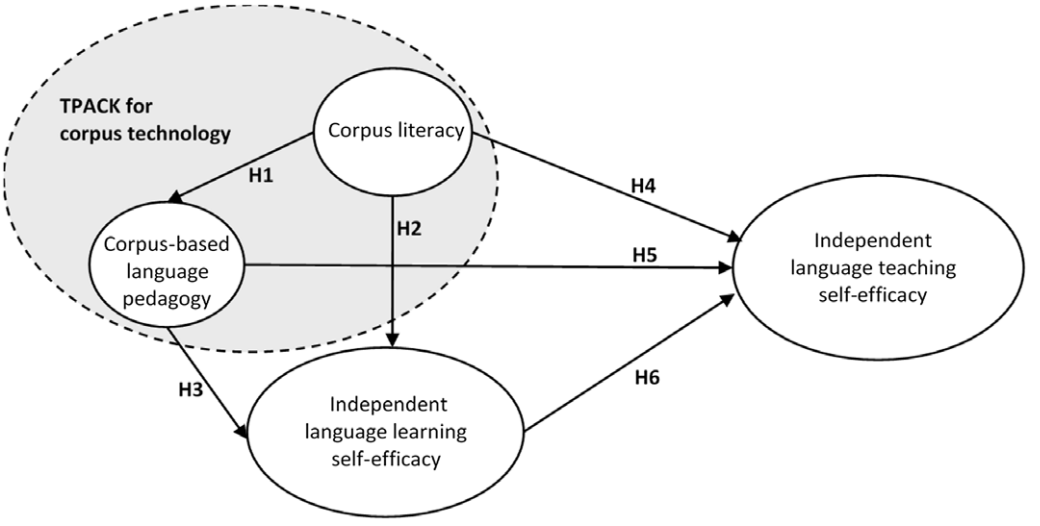


Figure 1. Theoretical model of TPACK for corpus technology and self-efficacy in independent language learning and teaching.

their ISE, defined as their perceived capacity for autonomous language learning. Given that most of our participants (from Hong Kong and/or mainland China) were non-native English speakers, it was essential to explore how they independently enhanced their subject knowledge for future teaching. Accordingly, **Hypothesis 2** and **Hypothesis 3** (**H2–H3**) state that both CL and CBLP will have positive influences on ISE.

Finally, we proposed that three of the components – CL, CBLP, and ISE – would contribute to self-efficacy for independent language teaching (**Hypotheses 4–6; H4–H6**). Therefore, the present study aims to validate this theoretical model in light of these hypotheses via an empirical study that included a CBLP intervention, a survey, and interview data.

Using the model, we investigated the following six hypotheses:

1. H1: CL will have a positive influence on CBLP.
2. H2: CL will have a positive influence on self-efficacy in independent language learning.
3. H3: CBLP will have a positive influence on self-efficacy for independent language learning.
4. H4: CL will have a positive influence on TSE for independent language teaching.
5. H5: CBLP will have a positive influence on TSE for independent language teaching.
6. H6: ISE will have a positive influence on TSE for independent language teaching.

Specifically, our study was guided by the following two research questions (RQs):

RQ1. How does student teachers' TPACK development for corpus technology influence their self-efficacy for independent language learning and teaching, respectively?

RQ2. How do student teachers perceive their TPACK learning experience for corpus technology?

7. Methodology

7.1. Context and participants

We used convenience sampling and invited 120 senior-year, pre-service TESOL student teachers (third year or above) from a university in Hong Kong to participate in the study. Our pre-survey

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

showed that all participants had very limited experience of corpus technology, since it was not included in their prior education programme training. All participants were non-native English speakers with dual identities as English learners and prospective English teachers.

First, the participants were invited to attend a four-week training programme in TPACK for corpus technology. At the end of the training programme, they were invited to complete a survey covering CL, CBLP, ISE, and TSE. As a follow-up procedure, eight participants agreed to participate in interviews to understand their perceptions and perspectives on their learning experience. Ethical procedures were followed by obtaining consent from all participants who completed the survey and interview, guaranteeing their understanding through clear communication and ensuring data privacy and anonymity.

7.2. Training in TPACK for corpus technology (CBLP)

The first two weeks of training focused on developing foundational corpus knowledge, while the next two weeks focused on practical pedagogical applications. In Week 1, the participants were introduced to the core concepts in corpus linguistics, including the nature and functions of corpora. Week 2 provided hands-on experience with a range of free, user-friendly online corpus tools, such as Lextutor (https://www.lextutor.ca/), COCA (https://www.english-corpora.org/coca/), SKELL (https://skell.sketchengine.eu/#home?lang=en), and Versatext (https://versatext.versatile.pub/) to equip them with the essential corpus query skills and strategies. The participants briefly explored the pedagogical potential of each platform, learned how to extract authentic language data, and designed some short, corpus-based learning activities. Weeks 3 and 4 shifted the focus to learning CBLP. The participants developed skills in creating corpus-based teaching materials, guided by the two-step CBLP training model (Ma *et al.*, 2022, 2024b), where Step 1 focused on improving the teachers' CL and Step 2 their pedagogical skills for classroom integration. The training assisted the student teachers to adapt and transform teaching materials to suit their learners' needs. They also studied sample CBLP lesson materials in texts and videos hosted on the CAP website (https://corpus.eduhk.hk/cap/), which was developed by the lead author. Trainees were guided in the design of CBLP materials, both individually and in collaboration with their peers.

7.3. Research instruments

7.3.1. Survey

The measurement survey used a six-point Likert scale and consisted of 25 items across four dimensions: CL, CBLP, ISE, and TSE. The *CL* section included five items that were aligned with the five core components – understanding, searching, analysis, advantages, and limitations – as described by Ma, Chiu, Lin and Mendoza (2023). The *CBLP* dimension featured five items that were adapted from the TPACK surveys by Chai, Koh and Tsai (2011). For *ISE*, five items were adapted from the Independence of Learning survey by Macaskill and Taylor (2010). The *TSE* section consisted of 10 items that were selected and adapted from the Ohio State Teacher Efficacy Scale (Tschannen-Moran & Hoy, 2001b), with the first five items measuring teacher self-efficacy for instructional strategies (TIS) and the remaining five focusing on teacher self-efficacy in student engagement (TSEN). See Part A in the supplementary materials for the survey questions.

7.3.2. Interview

Following the survey, eight participants agreed to participate in the semi-structured interviews, which consisted of three parts: (1) general background information, (2) CL and independent language learning, and (3) CBLP and independent language teaching. During the interviews, the student teachers reflected on how the training in TPACK for corpus technology influenced their

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

self-efficacy in independent language learning and teaching, as well as the potential links between these two forms of self-efficacy. Sample interview questions included:

1. *To what extent do you think that learning about corpora has helped you to become an independent language learner?* (CL)
2. *To what extent do you think that learning about corpus technology has helped you to learn how to design a corpus-based English lesson independently?* (CBLP and independent language teaching)

Some self-reflection questions were also included; for example, *Reflecting on your own language learning, what do you think a language teacher should do to engage students inside or outside of the classroom?*

7.4. Data analysis

After data cleaning and accounting for missing values, 96 valid survey responses were retained for the analysis. Following Kline (2015), the survey data were processed using reliability analysis, confirmatory factor analysis (CFA), and structural equation modelling (SEM) analyses using the R statistical environment (R Core Team, 2021) to confirm the five constructs, to test the theoretical model, and to verify the hypotheses (see Figure 1).

The analysis of the interview data involved several iterations of reading and coding to identify themes and patterns (Miles, Huberman & Saldaña, 2020). Using the constant comparative method (Denzin & Lincoln, 2000), all the data sources were coded and analysed independently for each selected student TESOL teacher. Two researchers coded the data independently and achieved an interrater reliability of 0.91; any disputed cases were resolved following a discussion and reaching mutual agreement. The in-depth analyses were then reviewed by each corresponding student teacher to allow for member checking (Miles *et al*., 2020) and to ensure accurate analyses.

8. Results

8.1. Validation of the theoretical model (RQ1)

The reliability analysis of the five constructs indicated strong internal consistency, with all Cronbach's α and composite reliability (CR) values exceeding 0.9 (see Part B in the supplementary materials). ISE and CL attained the highest mean scores (4.80 and 4.73 out of 6.00, respectively), reflecting greater self-perceived efficacy in these areas. By contrast, CBLP had the lowest mean (4.14), suggesting that the teachers considered developing pedagogical skills to be more challenging than acquiring subject knowledge or technological skills. In summary, these findings confirmed that all the constructs were measured reliably, thus supporting the robustness of the instrument that was used in this study. However, the high reliability coefficients for CBLP and TSEN (α and CR exceeding 0.95) indicated potential redundancy issues. To address this, adding covariances between several items may be necessary, as discussed in the subsequent paragraphs.

With regard to the CFAs for the five constructs, several fit indices for the original CBLP and the teachers' TSEN did not meet the recommended thresholds (see Part C in the supplementary materials). This indicated the need for model modifications, a standard practice in SEM analyses to enhance the model fit and is widely adopted in empirical studies (Kline, 2015). After adopting the recommendations generated by the fit modification function in R, several error covariances were added to CBLP (CBLP4 ↔ CBLP5) and TSEN (TSEN2 ↔ TSEN3; TSEN4 ↔ TSEN5). These modifications may have reflected the participants' status as pre-service teachers. They understood the ideas behind CBLP implementation and knew theoretically how to engage students, but they did not have real teaching experience in the

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

classroom. This discrepancy is likely to have contributed to the observed covariances between the items in the two constructs. In future study, we will consider merging these similar items.

After the above-mentioned revisions, the CFA results indicated that most fit indices for the five constructs met the recommended thresholds (Kline, 2015). Specifically, $\chi 2/df$ values for all the constructs were below 3.0, and comparative fit index (CFI) and Tucker–Lewis index (TLI) values were all above 0.90, reflecting a strong model fit. The standardised root-mean-square residual (SRMR) values were well below 0.05 for each construct, further supporting an excellent fit. However, the root-mean-square error of approximation (RMSEA) values for CL and TIS were 0.11 and 0.12, respectively, which exceeded the recommended cut-off of 0.08. As Kline (2015) stated, RMSEA is sensitive to sample size and can be decreased by increasing the sample size; the model remains acceptable if other fit indices, such as CFI and SRMR, meet acceptable thresholds. Therefore, the overall CFA results suggested that the measurement models for the constructs were well supported.

The next step was to finalise the model. First, CFAs were used to evaluate two models: a four-construct model with all items under TSE as one construct, and a five-construct model with the items under TSE divided into TSEN and TIS. As shown in Part C in the supplementary materials, the five-construct model demonstrated a better performance, with two of the five indices meeting the recommended standards, including $\chi 2/df$ ratio (2.19; lower than 3.00) and CFI (0.90; higher than 0.90). This suggests that treating TSEN and TIS as independent constructs enhanced the measurement model's goodness of fit, aligning with Tschannen-Moran and Hoy (2001b).

Second, a CFA was conducted on the five-construct model by adding the modifications for CBLP and TSEN, as mentioned previously. The result indicated that this model demonstrated a largely satisfactory goodness of fit and could thus be accepted as the final model. Most indices met the recommended standards: the $\chi 2/df$ ratio was 1.93 (well below the cut-off of 3.00); CFI was 0.92, and TLI was 0.91, both above the recommended 0.90 threshold. The SRMR value was 0.04, reflecting an excellent fit (< 0.05). The sample-sensitive RMSEA value was 0.09, above the recommended value of 0.08, yet below the poor fit threshold of 0.10, which might be resolved through increasing the sample size in the future. Although the RMSEA was slightly elevated, the model showed a good fit for the other indices, supporting the overall adequacy of the revised model (Kline, 2015).

After confirming the goodness of fit of the model through CFA, we conducted SEM analysis to investigate the path coefficients and examine the proposed hypotheses. As shown in Figure 2, several hypothesised paths were rejected due to insignificant coefficients. First, the paths from CL to TSEN and TIS were insignificant because CL only represented technological knowledge about corpus use, which may not have had a direct influence on the strategies for students' engagement and instruction. However, CL exerted strong and significant total effects on TSEN ($\beta = .753$, $p < .001$) and TIS ($\beta = .770$, $p < .001$). This emphasised that CL was a fundamental component in student teachers' TPACK for corpus technology to enhance their self-efficacy of engaging students and promoting instructional strategies. Second, although the hypothesis from CBLP to TIS was rejected, CBLP still exhibited a significant total effect on TIS ($\beta = .475$, $p = .003$) and an indirect effect through TSEN ($\beta = .158$, $p = .032$). This highlighted CBLP's pivotal role in fostering student engagement, which in turn enhances the effectiveness of instructional strategies, suggesting that engagement efficacy mediates the influence of CBLP on instructional strategies. Third, the insignificant path of CBLP to ISE indicated that CBLP related more to teaching self-efficacy than it did to learning self-efficacy. See Part C in the supplementary materials for the direct and total effects for all the paths.

The finalised SEM model, as illustrated in Figure 2, represented a well-supported and interpretable model for TPACK in the context of corpus technology and self-efficacy in language learning. Of the initial 10 hypotheses, six were supported, thus highlighting key pathways within the model. **CL** exerted strong, direct effects on both **CBLP** ($\beta = .867$, $p < .001$) and **ISE** ($\beta = .604$, $p < .001$). This indicates that higher CL not only enhanced the student

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

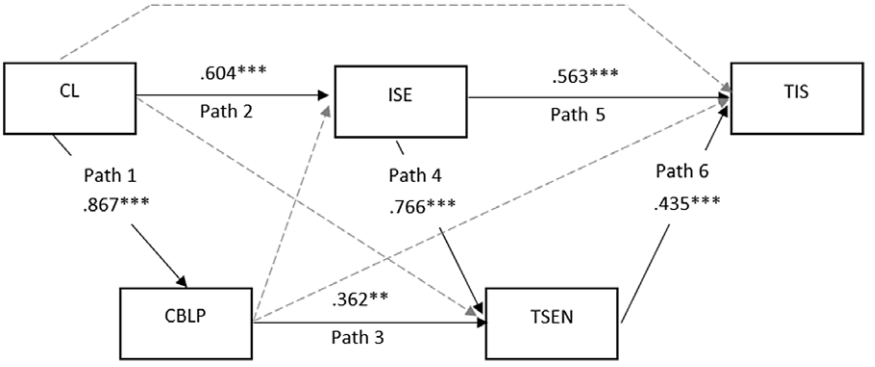


Figure 2. Final model of TPACK for corpus technology and self-efficacy in independent language learning and teaching. ***Note***. CL = corpus literacy; CBLP = corpus-based language pedagogy; ISE = independent language learning self-efficacy; TSEN = teacher self-efficacy for student engagement; TIS = teacher self-efficacy for instructional strategies.
***$p < 0.001$. **$p < 0.01$.

teachers' ability to apply corpus technology in pedagogy but also increased their confidence in independent language learning.

ISE itself played a central role in this model, as it directly predicted both TIS ($\beta = .563$, $p < .001$) and TSEN ($\beta = .766$, $p < .001$). This suggests that the student teachers with higher self-efficacy in language learning tended to be better equipped to engage learners and to implement instructional strategies effectively. CBLP further contributed to TSEN ($\beta = .362$, $p < .01$), signifying that pedagogical skills in corpus-based methods can enhance efficacy in engaging students. In turn, TSEN exerted a significant, direct effect on TIS ($\beta = .435$, $p < .001$), highlighting that strong self-efficacy to engage students is essential for promoting broader instructional language strategies.

Overall, this final model demonstrated that foundational skills in CL and pedagogy laid the foundation for developing instructional efficacy, which ultimately leads to better student engagement. The interconnections highlight the importance of TPACK as a precursor to effective teaching and student-centred outcomes. These findings suggest that professional development for TESOL student teachers or in-service teachers should focus on strengthening both their CL and their self-efficacy to maximise both teaching strategies and students' engagement in technology-enhanced language classrooms.

8.2. Interview results (RQ2)

The analysis of the interview results added further support to the relationships that were identified amongst the different constructs (as shown above), corresponding to the six paths verified in the final model (Figure 2) across six broad themes, which were further divided into 13 subthemes.

8.2.1. Theme 1: CL contributing to CBLP (Path 1)

The participants stated that improved *CL could help teachers teach students to correct their language mistakes* (Subtheme 1.1), particularly in word choice and grammatical accuracy. As one student teacher explained, "I may use corpus tools to address students' confusion regarding word choice in writing. [. . .] allowing them to see which option is more frequently used" [S07]. This demonstrates how CL can empower student teachers to provide data-driven, authentic language examples, thereby making abstract rules more concrete and accessible for students.

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

Moreover, the interviews revealed that *CL could consolidate students' prior knowledge and promote deeper learning* (Subtheme 1.2). As a participant shared, "I believe that using a corpus can deepen students' understanding of what they have learned, enabling them to learn more effectively" [S01]. Such insights illustrate how corpus-based methods can not only correct misunderstandings but also assist students to build on what they know, thus supporting more meaningful and long-lasting language learning.

8.2.2. Theme 2: CL contributing to self-efficacy for independent language learning (Path 2)

Under the subtheme of *improving language accuracy* (Subtheme 2.1), the participants described how corpus tools helped resolve uncertainties about word forms and usage. One participant shared, "If I am not sure about the comparative form of a word, [. . .], I will type them separately and then look at their frequency" [S08]. This approach, made possible by CL, empowered learners to make informed, data-driven decisions in their language use. The same participant emphasised, "Utilising corpora can enhance students' ability to explore language learning independently" [S08], highlighting how corpus tools could foster learner autonomy.

Improving language learning efficiency (Subtheme 2.2) was also mentioned frequently in the interviews. One participant reflected, "I can search for the collocations [. . .], which improves the efficiency of my language learning" [S02]. This demonstrates that CL not only increases accuracy but also streamlines the learning process by allowing learners to access and apply authentic language data more efficiently.

8.2.3. Theme 3: CBLP contributing to TSE for engaging students (Path 3)

Three subthemes emerged from the participants' reflections. First, *providing hands-on opportunities to search corpora* (Subtheme 3.1) empowered the students to take charge of their classroom learning. As one participant noted, "This approach [independent corpus search] empowers students to take ownership of the learning process and promotes deep understanding" [S04]. This suggests that CBLP shifts the role of the teacher to that of a facilitator, enabling students to become active, autonomous learners.

Second, *creating a student-centred classroom* (Subtheme 3.2) was identified as a key benefit. A participant explained, "[T]the classroom can be student-centred in which students discover and summarise language rules by themselves" [S06]. This transformation fosters engagement and participation, as students are more involved in constructing their own knowledge.

Third, *combining corpus use with fun elements* (Subtheme 3.3) was seen as essential for sustaining students' interest. As one student teacher described, "We must incorporate as many engaging elements as possible, such as games and videos, and combine games and videos with corpus usage to create a more interesting classroom environment" [S01]. This approach not only makes learning enjoyable but also enhances students' motivation.

8.2.4. Theme 4: Self-efficacy for independent language learning contributing to teachers' instructional strategies (Path 4)

The first subtheme emphasised the *teacher providing a role model for independent language learning in the classroom and helping learners become independent language learners* (Subtheme 4.1). One participant stated, "The role of a teacher is to empower students to become independent learners" [S03]. This highlights the importance of teachers modelling autonomous learning behaviours, which in turn inspires and equips students to adopt similar strategies for their own language development.

The second subtheme focused on *selecting online resources carefully and converting them into suitable teaching materials* for classroom use (Subtheme 4.2). As one student teacher expressed, "I believe it is a relatively important skill for teachers to search for, filter, and retrieve information

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

from the internet, and transform it into teaching materials in class" [S08]. This demonstrates that teachers with strong self-efficacy in independent learning will be adept at navigating digital resources, customising them, and integrating them into their instructional practices.

8.2.5. Theme 5: Self-efficacy for independent language learning contributing to teachers' strategies to engage students (Path 5)

The first subtheme discussed *using various resources and activities to engage students in classroom learning* (Subtheme 5.1). One student teacher commented, "By utilising various resources and offering diverse activities, students can be motivated and become more engaged in learning" [S04]. This highlights the motivational benefits for students when teachers diversify instructional content. This approach demonstrates how teachers who are confident in their independent learning skills actively seek out and integrate diverse learning resources/activities, resulting in more dynamic and engaging classroom experiences.

The second subtheme was *connecting learning to students' real lives to engage them in classroom learning* (Subtheme 5.2). Another interviewee explained, "We may use vocabulary that students are already familiar with, connecting learning to the students' real-life experiences" [S03]. This strategy not only makes lessons more relevant and relatable but also increases students' interest and participation.

8.2.6. Theme 6: Self-efficacy for student engagement contributing to teachers' instructional strategies (Path 6)

One subtheme involved *using corpus resources to increase students' interaction and enrich students' classroom learning* (Subtheme 6.1). As one participant shared, "To motivate students to actively participate in class, we incorporated pair and group corpus-based activities to enhance interaction among students" [S08]. This illustrates how student teachers can feel confident about engaging students to use corpus activities not just for individual learning but also as a foundation for dynamic, interactive group tasks, fostering a more collaborative classroom environment.

The second subtheme centred on the idea that *CBLP and task-based language teaching (TBLT) could be combined to improve classroom teaching* (Subtheme 6.2). As one participant noted, "I believe corpus-based teaching can incorporate elements of task-based instruction" [S05]. This highlights the value of integrating CBLP with TBLT (or other relevant pedagogical approaches) to provide innovative instructional strategies that enhance students' engagement and learning outcomes.

9. Discussion

In the present study, we sought to empirically verify the linking of TPACK, CBLP, and student teachers' TSE. Following a survey validating and confirming the hypotheses in our theoretical model linking these concepts, interview data further revealed latent subthemes underlying student teachers' responses. The following discussion outlines the study's main contributions in more detail.

9.1. A verified framework of TPACK for corpus technology

The final theoretical model suggests that improved CL may be associated with greater self-efficacy for independent language learning. This finding lends some empirical support to the widely held view that corpus use may contribute to the development of learner autonomy (Boulton, 2017; Lee *et al*., 2019; O'Keeffe, 2021). The model also highlights the crucial role of CBLP in shaping instructional efficacy, which shows that strong pedagogical skills in corpus-based methods enhance self-efficacy regarding student engagement, and support the creation of

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

interactive, student-centred learning environments (Lee *et al*., 2019; Ma *et al*., 2024a; Schmidt, 2023). This is likely due to the inherently participatory and inquiry-driven nature of corpus-based activities, which actively involve students in language exploration (Boulton, 2011; Johns, 1991; Meunier, 2019). Furthermore, CBLP has an indirect effect on self-efficacy for instructional strategies. This underscores that pedagogical expertise in corpus methods strengthens teachers' overall instructional strategies through its positive influence on engagement efficacy. The central role of ISE is also evident in the model; it directly predicts both self-efficacy for instructional strategies and student engagement (Karatas & Arpaci, 2021). This highlights that teachers who are confident about their own autonomous learning through corpus are better positioned to engage students' corpus use and implement a diverse range of instructional strategies. This dynamic chain of influence highlights the essential function of self-efficacy in facilitating effective teaching and learning processes.

Several hypotheses were not supported by the findings. As a pedagogical approach, CBLP may not directly enhance ISE. In addition, while CL encompasses the technological knowledge required to understand and use corpora, it does not directly influence TSE regarding instructional strategies or students' engagement. However, CL may affect self-efficacy indirectly by strengthening both CBLP and ISE. Although the overall effect is significant, the lack of a direct effect of CBLP on TIS is unexpected, which indicates the need for further research on the relationship between CBLP and teachers' instructional strategies.

9.2. Impact of TPACK training for corpus technology on TESOL student teachers

Integrating TPACK training for corpus technology significantly enhances TESOL student teachers' professional development by equipping them with innovative tools and strategies to foster student engagement and independent learning. By receiving targeted training in corpus tools, teachers gain confidence and improve their self-efficacy in designing hands-on activities, thus encouraging their students to actively explore authentic language data (Crosthwaite *et al*., 2023; Ma *et al*., 2024a). The student teachers' appreciation of hands-on corpus searches when enacting CBLP in lesson design also echoes the concept of mastery experiences, a core component in developing self-efficacy (Mohammed, 2021). The emphasis on hands-on activities also promotes a shift from teacher-centred methods towards a student-centred environment where discovery learning is emphasised. Additionally, when corpus use is combined with fun elements, such as games, collaborative searches, or multimedia activities, it can further increase students' interest and motivation (Crosthwaite & Boulton, in press; Meunier, 2019). These approaches help to create a dynamic classroom atmosphere conducive to deeper learning and sustained engagement.

Second, the study highlights that self-efficacy for independent language learning fosters the development of effective teaching strategies. When student teachers themselves are confident in their ability to learn independently using digital resources, they are more likely to integrate diverse online materials and activities into their instruction. This includes the exploration and incorporation of various online learning resources, which will not only enrich the lesson content, but will also expose students to multiple perspectives and authentic language contexts. Importantly, the ability to connect classroom learning to students' real-life experiences, such as through relevant topics or familiar vocabulary, helps bridge the gap between academic content and everyday usage (O'Neill & Short, 2025).

Third, enhancing self-efficacy for student engagement leads to significant improvements in instructional strategies. Student teachers who feel equipped to engage learners are more likely to adopt interactive practices, such as using corpus resources for pair and group work, thereby increasing students' interaction and participation (Liu & Ma, 2025; Ma & Mei, 2021; O'Keeffe, 2021). Furthermore, the combination of corpus-based teaching and TBLT empowers teachers to design lessons that are both data-driven and communicative. This hybrid approach not only

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

supports language learning through authentic tasks but also provides students with opportunities to engage in practical, goal-oriented activities (Zare, Noughabi & Al-Issa, 2024).

Finally, our study provides rich, interview-based evidence regarding student teachers' reflective practice of CBLP, emphasising its crucial role in teacher education, alongside CL and pedagogical skills (Farr & Leńko-Szymańska, 2024). Thematic analysis revealed the significance of reflective engagement with CBLP, where student teachers not only deepened their understanding of corpus tools and language instruction but also developed greater self-efficacy and autonomy in both learning and teaching. The participants described how CBLP empowered them to design more interactive, student-centred lessons, to model independent learning behaviours, and to integrate diverse digital resources effectively. Eventually, the reflective practice of CBLP fosters both teachers' professional growth and more engaging, authentic classroom experiences for their students.

Our study provides valuable guidance for the integration of corpus technology into TESOL teacher education, which should focus on the cultivation of CL and ISE, in addition to explicit training in CBLP. By doing so, TESOL educators can maximise the impact of technological and pedagogical knowledge on student engagement and instructional efficacy. This helps TESOL and teachers of other languages prepare effectively for the demands of technology-enhanced classrooms.

9.3. TESOL student teachers' perceived impact of teacher TPACK for corpus technology on their target language learners

The interview data showed that the TESOL student teachers hold positive perceptions regarding their enhanced TPACK for corpus technology, which they believe can subsequently influence their target students' language learning outcomes in the future. First, CBLP provides student teachers with powerful tools to present language rules in authentic, context-rich ways. By drawing on real-life examples from corpora, teachers can move beyond abstract explanations and rote memorisation and can provide students with concrete data that support and deepen their existing language knowledge (Boulton, 2017; Meunier, 2019). This approach not only clarifies grammatical structures and word usage but also helps learners internalise language patterns through repeated exposure to authentic usage (Crosthwaite & Boulton, in press; Yan & Ma, 2024). Therefore, the use of corpus data bridges the gap between theoretical knowledge and practical application, making language learning more meaningful and effective for students.

Second, the utilisation of corpus tools as a result of TPACK training fosters greater self-efficacy for independent language learning. When the student teachers incorporate corpus technology into their future instruction, they will empower learners to investigate language questions autonomously, such as exploring word frequency and authentic collocations. This process will not only improve the learners' language accuracy but also enhance their learning efficiency, as they will be able to access and verify language patterns quickly and independently.

Third, the influence of TPACK training extends beyond direct instruction to the modelling of effective learning strategies by teachers. In this study, the student teachers who demonstrated independent language learning behaviours, such as searching for, evaluating, and integrating online resources, may serve as role models for their students in the future. This modelling effect is particularly important for promoting autonomy, as learners are more likely to adopt similar strategies when they observe their teachers engaging in independent learning and resource selection. Moreover, when teachers carefully curate and adapt online materials for classroom use, they ensure that learning resources are relevant and accessible, further supporting students' engagement and independent learning (Lasekan, Pachava, Godoy Pena, Golla & Raje, 2024).

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

10. Conclusion

In conclusion, our study establishes an empirically verified framework for integrating TPACK and corpus technology into TESOL teacher education. The model demonstrates that CL can lead to students' autonomy and proficiency in CBLP, significantly increase self-efficacy for student engagement, and empower teachers to create interactive, student-centred, and enjoyable learning environments. When teachers have high self-efficacy in both independent learning and student engagement, they are better equipped to design lessons that leverage online resources and real-world contexts, thereby motivating and engaging students more effectively. These insights highlight the transformative potential of TPACK training for corpus technology in TESOL teacher preparation. Embedding CL, the development of self-efficacy, and CBLP in teacher-education programmes is essential for preparing future language teachers to meet the demands of dynamic, technology-enhanced classrooms. This integration will not only promote pedagogical innovation but also foster a culture of independent, lifelong learning amongst both teachers and students.

Despite the valuable insights offered by this study, there are two main limitations. First, there are minor limitations of the model, such as the relatively high RMSEA. Although student teacher cohorts are usually small internationally, future studies could increase the sample size to improve model fit indices. Second, the study focused exclusively on pre-service TESOL teachers, which may limit the generalisability of the results to in-service TESOL teachers or teachers of other languages. Future research should involve a larger and more diverse participant pool, including in-service TESOL teachers or those teaching different languages to further validate the theoretical model, thereby increasing its robustness and generalisability across diverse educational contexts and professional experience levels.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344025100347

Data availability statements. The data supporting the findings of this study are available within the supplementary materials.

Authorship contribution statement. Qing Ma: Conceptualisation, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualisation, Writing – original draft, Resources, Writing – review & editing. Jiahao Yan: Data curation, Formal analysis, Methodology, Visualisation, Writing – review & editing. Peter Crosthwaite: Investigation, Validation, Writing – review & editing. John Chi-Kin Lee: Validation, Writing – review & editing.

Funding disclosure statement. This article is supported by the GRF grant (18600123), funded by the Research Grants Committee of Hong Kong, and the SoLT project (02A13), funded by the Education University of Hong Kong.

Competing interests statement. The authors declare no competing interests.

Ethical statement. This study was conducted in accordance with the ethical standards of the Education University of Hong Kong and received approval from the Human Research Ethics Committee at the Education University of Hong Kong (approval number: 2022-2023-0110). All participants participated voluntarily in this research, and appropriate ethical procedures were followed to obtain participant consent.

GenAI use disclosure statement. The authors declare no use of generative AI.

References

Boulton, A. (2011) Data-driven learning: The perpetual enigma. In Goźdź-Roszkowski, S. (ed.), *Explorations across languages and corpora*. Lausanne: Peter Lang, 563–580. https://hal.science/hal-00528258/

Boulton, A. (2017) Corpora in language teaching and learning. *Language Teaching*, 50(4): 483–506. https://doi.org/10.1017/S0261444817000167

Callies, M. (2019) Integrating corpus literacy into language teacher education. In Götz, S. & Mukherjee, J. (eds.), *Learner corpora and language teaching*. Amsterdam: John Benjamins, 245–264. https://doi.org/10.1075/scl.92.12cal

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

Caprara, G. V., Barbaranelli, C., Steca, P. & Malone, P. S. (2006) Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, 44(6): 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Chai, C. S., Koh, J. H. L. & Tsai, C. C. (2011) Exploring the factor structure of the constructs of technological, pedagogical, content knowledge (TPACK). *The Asia-Pacific Education Researcher*, 20(3): 595–603. https://hdl.handle.net/10497/4790

Chambers, A. (2019) Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4): 460–475. https://doi.org/10.1017/S0261444819000089

Chen, M. & Flowerdew, J. (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. https://doi.org/10.1075/ijcl.16130.che

Chung, T. T., Crosthwaite, P., Cao, C. T. H. & de Carvalho, C. T. (2024) Walking the walk? (Mis)alignment of EFL teachers' self-reported corpus literacy skills and their competence in planning and implementing corpus-based language pedagogy. *TESOL Quarterly*, 58(3): 1046–1080. https://doi.org/10.1002/tesq.3299

Crosthwaite, P. & Boulton, A. (in press) DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In Tyne, H., Bilger, M., Buscail, L., Leray, M., Curry, N. & Pérez-Sabater, C. (eds.), *Discovering language: Learning and affordance*. Lausanne: Peter Lang.

Crosthwaite, P., Luciana & Wijaya, D. (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. https://doi.org/10.1080/09588221.2021.1995001

Denzin, N. K. & Lincoln, Y. S. (2000) *The handbook of qualitative research* (2nd ed.). Thousand Oaks: Sage Publications.

Farr, F. & Leńko-Szymańska, A. (2024) Corpora in English language teacher education: Research, integration, and resources. *TESOL Quarterly*, 58(3): 1181–1192. https://doi.org/10.1002/tesq.3281

Graham, C. R., Borup, J. & Smith, N. B. (2012) Using TPACK as a framework to understand teacher candidates' technology integration decisions. *Journal of Computer Assisted Learning*, 28(6): 530–546. https://doi.org/10.1111/j.1365-2729.2011.00472.x

Huang, S. & Ma, Q. (2025) A systematic review of data-driven learning research on language learning and teaching for pre-tertiary learners: Balancing qualitative and quantitative research. *Learning and Individual Differences*, 122: Article 102752. https://doi.org/10.1016/j.lindif.2025.102752

Johns, T. (1991) Should you be persuaded: Two examples of data-driven learning. In Johns, T. & King, P. (eds.), *Classroom concordancing. ELR Journal* (Vol. 4). Birmingham: University of Birmingham, 1–16.

Joo, Y. J., Park, S. & Lim, E. (2018) Factors influencing preservice teachers' intention to use technology: TPACK, teacher self-efficacy, and technology acceptance model. *Journal of Educational Technology & Society*, 21(3): 48–59. https://www.jstor.org/stable/26458506

Karatas, K. & Arpaci, I. (2021) The role of self-directed learning, metacognition, and 21st century skills predicting the readiness for online learning. *Contemporary Educational Technology*, 13(3): Article ep300. https://doi.org/10.30935/cedtech/10786

Karlsen, P. H. & Monsen, M. (2020) Corpus literacy and applications in Norwegian upper secondary schools: Teacher and learner perspectives. *Nordic Journal of English Studies*, 19(1): 118–148. https://doi.org/10.35360/njes.500

Kline, R. B. (2015) *Principles and practice of structural equation modeling* (4th ed.). New York: The Guilford Press.

Lasekan, O. A., Pachava, V., Godoy Pena, M. T., Golla, S. K. & Raje, M. S. (2024) Investigating factors influencing students' engagement in sustainable online education. *Sustainability*, 16(2): Article 689. https://doi.org/10.3390/su16020689

Lee, H., Warschauer, M. & Lee, J. H. (2019) The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5): 721–753. https://doi.org/10.1093/applin/amy012

Lee, M.-H. & Tsai, C.-C. (2010) Exploring teachers' perceived self efficacy and technological pedagogical content knowledge with respect to educational use of the world wide web. *Instructional Science*, 38(1): 1–21. https://doi.org/10.1007/s11251-008-9075-4

Liu, J. & Ma, Q. (2025) Examining corpus-based language pedagogy (CBLP) practices in data-driven learning (DDL) for low-proficiency L2 English learners: A meta-analysis. *Educational Technology & Society*, 28(2): 53–76. https://doi.org/10.30191/ETS.202504_28(2).RP04

Ma, Q., Chiu, M. M., Lin, S. & Mendoza, N. B. (2023) Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL*, 35(1): 19–39. https://doi.org/10.1017/S0958344022000180

Ma, Q., Lee, H. T. H., Gao, X. & Chai, C. (2024b) Learning by design: Enhancing online collaboration in developing pre-service TESOL teachers' TPACK for teaching with corpus technology. *British Journal of Educational Technology*, 55(6): 2639–2667. https://doi.org/10.1111/bjet.13458

Ma, Q. & Mei, F. (2021) Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1): 177–190. https://doi.org/10.1515/jccall-2021-2008

Ma, Q., Tang, J. & Lin, S. (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. https://doi.org/10.1080/09588221.2021.1895225

Ma, Q., Yuan, R., Cheung, L. M. E. & Yang, J. (2024a) Teacher paths for developing corpus-based language pedagogy: A case study. *Computer Assisted Language Learning*, 37(3): 461–492. https://doi.org/10.1080/09588221.2022.2040537

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

Macaskill, A. & Taylor, E. (2010) The development of a brief measure of learner autonomy in university students. *Studies in Higher Education*, 35(3): 351–359. https://doi.org/10.1080/03075070903502703

Meunier, F. (2019) A case for constructive alignment in DDL: Rethinking outcomes, practices, and assessment in (data-driven) language learning. In Crosthwaite, P. (ed.), *Data-driven learning for the next generation*. Abingdon: Routledge, 13–30. https://doi.org/10.4324/9780429425899-2

Miles, M. B., Huberman, A. M. & Saldaña, J. (2020) *Qualitative data analysis: A methods sourcebook* (4th ed.). Thousand Oaks: SAGE Publications.

Mishra, P. & Koehler, M. J. (2006) Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6): 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mohammed, A. (2021) Students' speaking proficiency and self-efficacy theory. *Advances in Social Sciences Research Journal*, 8(2): 318–325. https://doi.org/10.14738/assrj.82.9660

Mukherjee, J. (2004) Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In Connor, U. & Upton, T. A. (eds.), *Applied corpus linguistics: A multidimensional perspective*. Amsterdam: Rodopi, 239–250. https://doi.org/10.1163/9789004333772_014

O'Keeffe, A. (2021) Data-driven learning – a call for a broader research gaze. *Language Teaching*, 54(2): 259–272. https://doi.org/10.1017/S0261444820000245

O'Neill, G. & Short, A. (2025) Relevant, practical and connected to the real world: What higher education students say engages them in the curriculum. *Irish Educational Studies*, 44(1): 23–40. https://doi.org/10.1080/03323315.2023.2221663

R Core Team. (2021) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

Schmidt, N. (2023) Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, 35(1): 40–57. https://doi.org/10.1017/S0958344022000106

Sevimel, A. & Subasi, G. (2018) The factors affecting teacher efficacy perceptions of Turkish pre-service English language teachers. *The Journal of Language Teaching and Learning*, 8(1): 1–17.

Shulman, L. S. (1986) Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2): 4–14. https://doi.org/10.3102/0013189X015002004

Shulman, L. (1987) Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1): 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Sloan, K., Allen, A., Bass, K. M. & Milligan-Mattes, E. (2018) Evaluation of PCK in STEM residency programs: Challenges and opportunities. In Uzzo, S. M., Graves, S. B., Shay, E., Harford, M. & Thompson, R. (eds.), *Pedagogical content knowledge in STEM: Research to practice*. Cham: Springer, 157–173. https://doi.org/10.1007/978-3-319-97475-0_9

Tschannen-Moran, M. & Hoy, A. W. (2001a) Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7): 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M. & Hoy, A. W. (2001b) *Ohio State Teacher Efficacy Scale (OSTES)* [Database record]. *APA PsycTests*. https://doi.org/10.1037/t11400-000

Tschannen-Moran, M., Hoy, A. W. & Hoy, W. K. (1998) Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2): 202–248. https://doi.org/10.3102/00346543068002202

Tseng, J.-J., Chai, C. S., Tan, L. & Park, M. (2022) A critical review of research on technological pedagogical and content knowledge (TPACK) in language teaching. *Computer Assisted Language Learning*, 35(4): 948–971. https://doi.org/10.1080/09588221.2020.1868531

Wyatt, M. (2016) "Are they becoming more reflective and/or efficacious?" A conceptual model mapping how teachers' self-efficacy beliefs might grow. *Educational Review*, 68(1): 114–137. https://doi.org/10.1080/00131911.2015.1058754

Yan, J. & Ma, Q. (2024) Developing advanced citation skills: A mixed-methods approach to corpus technology training for novice researchers. *Journal of English for Academic Purposes*, 72: Article 101451. https://doi.org/10.1016/j.jeap.2024.101451

Zare, J., Noughabi, M. A. & Al-Issa, A. (2024) The impact of data-driven learning form-focused tasks on learners' task engagement: An intervention study. *ReCALL*, 36(3): 306–323. https://doi.org/10.1017/S0958344024000120

About the authors

Qing Ma is a Professor at the Department of Linguistics and Modern Language Studies and Associate Dean (Research) at the Faculty of Humanities, the Education University of Hong Kong. Her research focuses on second language vocabulary acquisition, corpus linguistics, corpus-based pedagogy and literature, CALL, MALL, and AI in language education.

Jiahao Yan is a Postdoctoral Fellow at the Department of Linguistics and Modern Language Studies, the Education University of Hong Kong. His main research interests include English for academic purposes, corpus linguistics, corpus-based language pedagogy, and generative AI in language learning and learning.

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press

Peter Crosthwaite is an Associate Professor in the School of Languages and Cultures at the University of Queensland. His areas of research include corpus linguistics and the use of corpora for language learning (i.e. data-driven learning), as well as computer-assisted language learning, and English for general and specific academic purposes.

John Chi-Kin Lee is the President and Chair Professor of Curriculum and Instruction of the Education University of Hong Kong. His research interests are curriculum and instruction and geographical and environmental education. He was named among the top 2% most-cited scientists in the world in terms of career-long impact released by Stanford University.

https://doi.org/10.1017/S0958344025100347 Published online by Cambridge University Press