Applied Corpus Linguistics 1 (2021) 100003

Contents lists available at ScienceDirect

Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp

Voices from the periphery: Perceptions of Indonesian primary vs secondary pre-service teacher trainees about corpora and data-driven learning in the L2 English classroom

Peter Crosthwaite [a,*], Luciana [b], Martin Schweinberger [c]

[a] School of Languages and Cultures, University of Queensland
[b] Universitas Katolik Indonesia Atma Jaya, Jakarta, Indonesia
[c] UiT, The Arctic University of Norway

A B S T R A C T

Data-driven learning (DDL), or the use of language corpora for the purposes of language learning and teaching, has seen a marked increase in research interest within ICT-rich WEIRD (Western, Educated, Industrialised, Rich and Democratic) contexts. However, less is known about its adoption in nations such as Indonesia where ICT/CALL training is underdeveloped, a situation indicative of a potential "digital divide" between the haves and have-nots when it comes to adoption of DDL worldwide. The present study reports on a series of DDL workshops and lesson planning activities within a CALL training program for pre-service L2 English-as-a-foreign-language teachers in Indonesia at both primary and secondary school levels. Training included an online course in DDL for academic writing, DDL expert's comments on trainees' lesson plans, and a series of online workshops. Trainees' perceptions of the training and the potential implementation of DDL within the Indonesian L2 English teaching context were explored through questionnaire and interview data. The findings painted a mixed picture of pre-service trainees' appreciation for the potential of corpora to greatly assist the practice of language learning and teaching, tempered by acknowledgement that integrating corpora into classroom practice in the Indonesian context would be a considerable challenge. Primary school teachers were also significantly less likely than secondary school teachers to be willing adopt corpora in their future teaching practice. We address the source of these concerns while offering suggestions for future DDL training involving pre-tertiary educators.

1. Introduction

"Data-driven learning" (DDL, Johns, 1997) is a pedagogical approach to technology-enhanced language learning involving either teacher-printed concordances of pre-selected corpus data for language learner mediation or learners' direct use of corpus query software to aid language acquisition. DDL has been used for the teaching and learning of second language (L2) vocabulary, grammar, collocation, and the resolution of L2 errors, with the act of corpus consultation facilitating usage-based learning processes that emphasise the role of frequency and contingency during L2 acquisition (Ellis, 2006). This emphasis is achieved through DDL via constructivist, student-led focus-on-form achieved through students' individual, autonomous corpus consultation (Cobb, 1999), which may be enhanced through teacher- and peer-mediated collaborative focus-on-forms, as suggested under socio-cultural / socio-constructivist accounts of L2 learning (O'Keeffe, 2020). Claims regarding DDL's efficacy in these areas have found support in over 250 related empirical studies (Boulton and Cobb, 2017).

While a number of studies have explored DDL from a teacher education perspective at the tertiary level (e.g. Chen et al., 2019), we are now seeing an increase in studies featuring DDL training for pre-tertiary English as an additional language/dialect (EAL/D) or English as a Foreign Language (EFL) teachers (Crosthwaite, 2019). The typical outcomes of such training include short-term positive perceptions of corpora and DDL by such teachers (e.g. Tyne, 2012), but which may also be affected by certain 'rational fears' (Boulton, 2009) teachers hold about corpora and/or DDL that might delay or ultimately prevent integration into mainstream classroom practice (Latif, 2020). These 'fears' include concerns about whether and how DDL actually promotes learning, a lack of availability and/or functionality of DDL resources suitable for classroom implementation, and (of importance for the present study) whether teachers and learners across cultures, proficiencies or learning contexts are willing to (or even able to) integrate DDL into actual classroom practice. Schaeffer-Lacroix (2019) suggests trainees' poor knowledge of corpus tool features and corpus data exploration techniques are "first order" barriers, which can be resolved fairly simply with appropriate training. However, "second-order" barriers, such as trainees' individual doubts about DDL as an approach or cultural arguments against its adoption, are harder to overcome, and can result in teachers' outright rejection of DDL as viable for younger learners. A general lack of Technological Pedagogical Content Knowledge

* Corresponding author at: Room 510 Gordon Greenwood Building, University of Queensland, St. Lucia, QLD 4072, Australia.
E-mail addresses: p.cros@uq.edu.au (P. Crosthwaite), luciana@atmajaya.ac.id (Luciana), m.schweinberger@uq.edu.au (M. Schweinberger).

https://doi.org/10.1016/j.acorp.2021.100003
Received 25 November 2020; Received in revised form 14 January 2021; Accepted 16 February 2021
Available online 2 March 2021

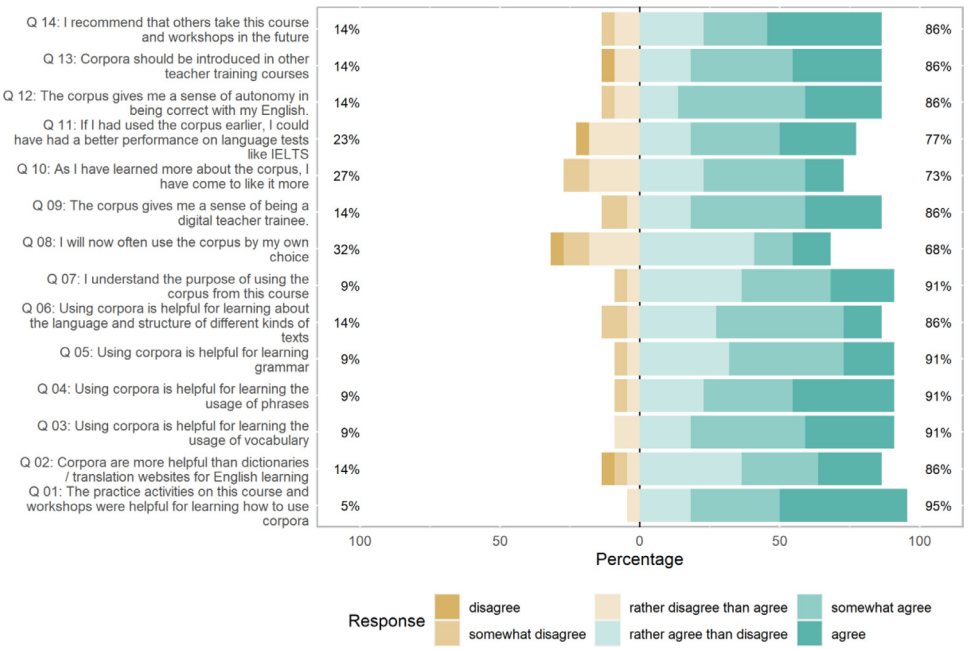

Fig. 1. Participant's responses regarding opinions of corpus training.

teachers differ across school types while the analysis did not confirm significant differences between questions (see Fig. 3).

6.2. *Participants' perceptions regarding their corpus literacy following training*

Fig. 4 explores all trainees' perceptions to Part 2 of the survey, dealing with trainees' self-perceptions of corpus literacy following DDL training.

When asked which corpus platforms the trainees preferred for learning, SKELL was the most popular, followed by Versatext, SketchEngine and Linggle respectively. Overall, over half of the respondents reported difficulty in terms of understanding corpus output and analysing and interpreting concordance and collocation data. In particular, 82% of participants agreed that reading concordances was difficult. Much of this difficulty may stem from the use of the BAWE in SketchEngine for much of the DDL training, which may be beyond the L2 English capabilities of many of the participants. 86% of participants also felt that it took too much time and effort to analyse the corpus data to draw meaningful conclusions, which could also be attributed to L2 English concerns, or the complexity of the SketchEngine platform. Nonetheless, 64% of respondents believed the act of formulating corpus queries was easy to learn, while 86% agreed that they had mastered the basics of corpus consultation by the time the training was completed, despite the aforementioned difficulties.

Fig. 5 describes participants' responses to this section of the survey divided across primary/secondary school trainees.

100% of secondary school trainees believed they had mastered the basics of corpus consultation by the end of the training period as compared to only 75% of the primary school trainees, while only 58% of primary school trainees felt they were able to generate appropriate corpus queries compared with 70% of secondary school trainees. Primary school trainees reported a higher degree of difficulty than secondary school trainees particularly for Q7 (difficulty due to cut-off sentences), Q8 (difficulty analysing concordance output), Q12 (difficulty determining which corpus to use) and Q13 (difficulty with complexity of corpus data). However, an additional Conditional Inference Tree analysis did not confirm any significant partitioning of the data based on school type or questions.

6.3. *Participants' perceptions regarding future integration of corpora into the classroom*

Fig. 6 explores all trainees' perceptions to Part 3 of the survey, dealing with trainees' perceptions of future corpus integration into classroom practice.

While the trainees were very positive regarding the potential efficacy of DDL for improving students' general (95%) and specific (100%) writing abilities, and that corpora would increase students' confidence in learning (77%) and writing (86%) in English, trainees were less enthusiastic about adopting corpus use within their teaching context. Notably, trainees' responses were at chance or even negative toward regarding adopting corpus use in their own context or designing DDL-focused teaching materials for younger learners. Trainees were split on whether DDL would be useful for teaching younger learners, and negative towards learners replacing their current language reference resources with corpora. Importantly, 82% of respondents agreed that corpora and DDL would be very useful resources for their teaching. However, 91% of trainees reported that there would be difficulties implementing DDL in Indonesia, a situation explored in more detail in the discussion section.

Fig. 7 describes participants' responses to this section of the survey divided into those from primary/secondary school trainees.

Unfortunately, 100% of primary trainees responded there would be difficulties implementing corpora and DDL in the Indonesian context, compared with (a still high) 80% for secondary school trainees. Secondary school trainees were also more positive in their overall reaction to the usefulness of corpora/DDL as a resource for their language teaching in Q13. Interestingly, although primary school trainees did not go on to complete the lesson planning activities, the data suggest that the secondary school trainees were less positive than primary school trainees on their responses to Q4 ('corpora improve confidence in English learning'), Q5 ('corpora improving confidence in English writing'), Q7 ('I could use corpora to design teaching materials for younger learners'), Q8 ('corpora are useful for teaching younger learners'), and Q10 ('learners will replace current resources with corpora'). However, on a more positive note, there was little difference in both groups' positive perceptions of the value of corpora for improving students' general and specific writing, and for improving students' confidence in learning and writing. The issue for both groups appears to be that of implementation into classroom practice, and more specifically implementation in

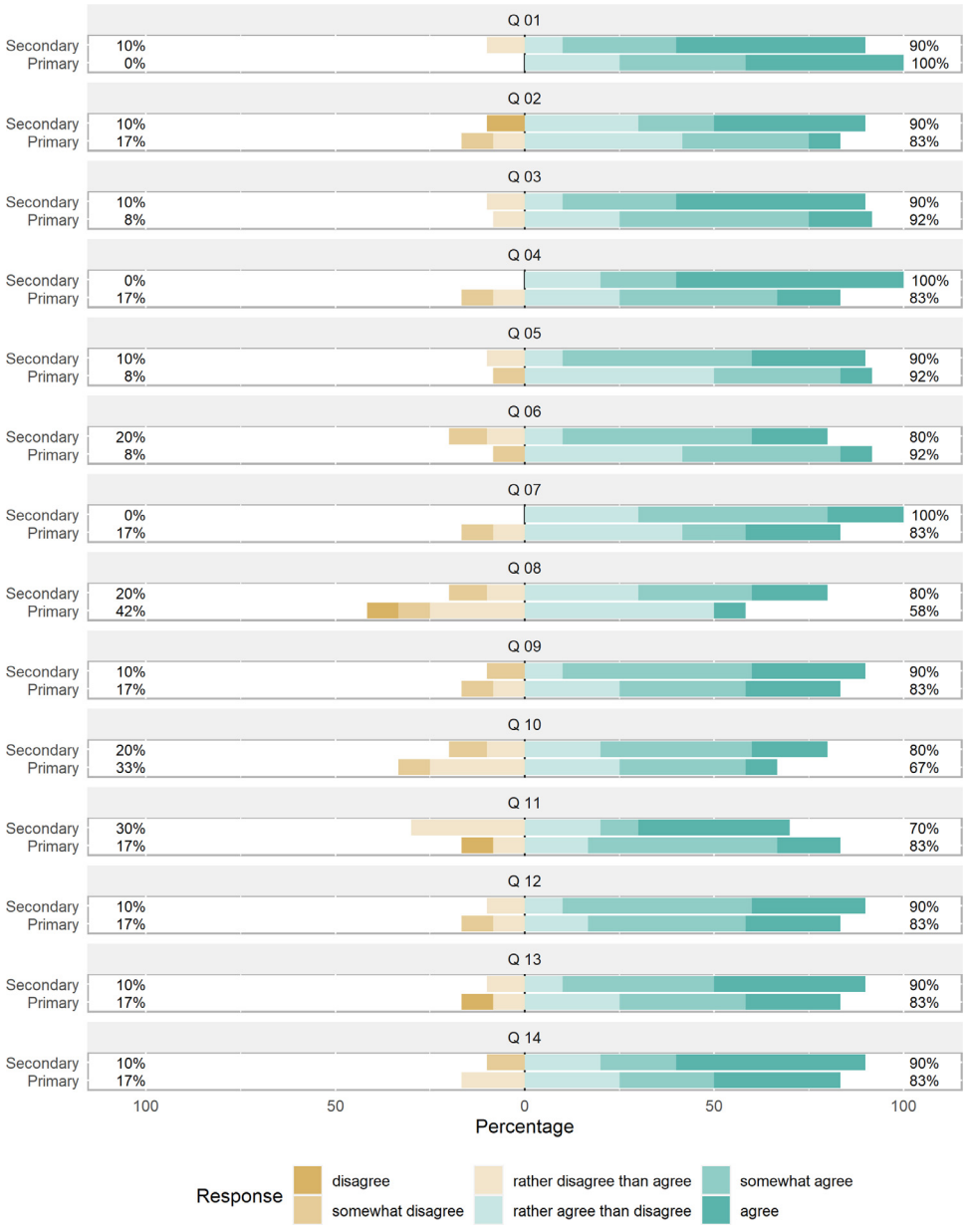

Fig. 2. Primary vs. Secondary school participant's responses regarding corpus training.

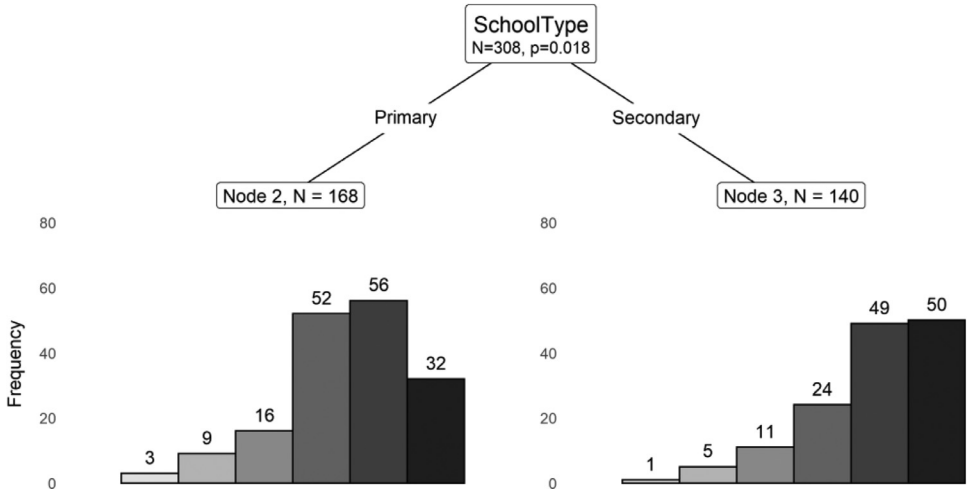

Fig. 3. Results of the Conditional Inference Tree analysis that was applied to the attitudinal data.

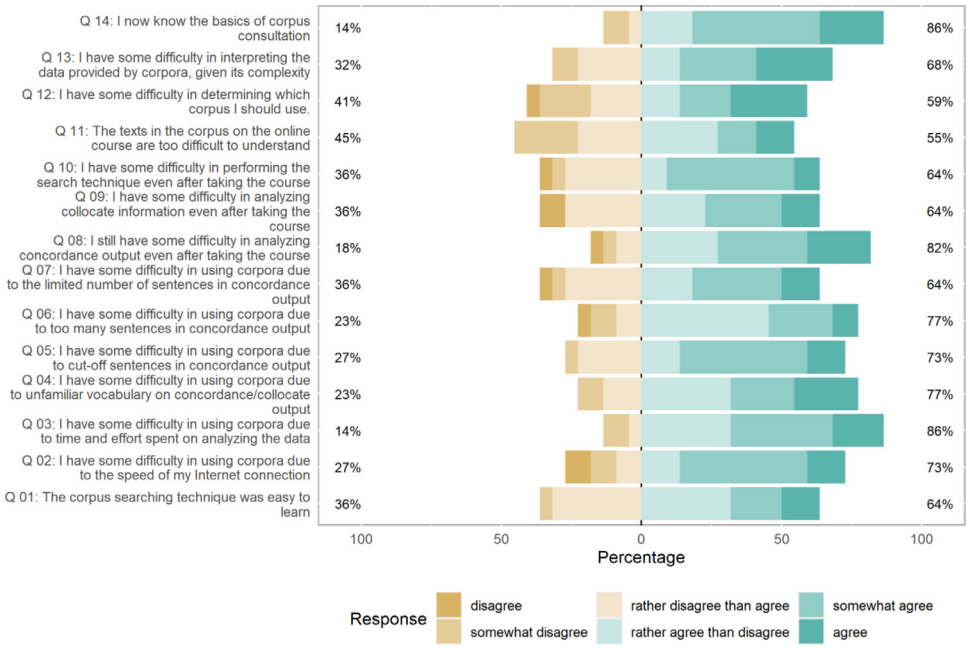

Fig. 4. Participants' self-perceptions of corpus literacy following DDL training.

the Indonesian context. An additional Conditional Inference Tree analysis found teachers' ratings differed by school type and question, with more negative responses being given to questions 6, 7, 8, and 10 and by primary school teachers (see Fig. 8).

6.4. Interviews

When asked whether the practice training activities were helpful for learning how to use corpora, both primary and secondary trainees tended to speak highly of the training they received in terms of how learning about corpora benefitted them personally as a reference resource:

> Trainee J (Primary): *The course helped me to learn how to use corpora. I know how check synonyms, collocations, and also […] find many examples.*

> Trainee C (Secondary): *Speaking personally, the practice activities especially the one in the workshops with [the first author] were really helpful for me to learn how to use corpora, because before this, I wasn't really sure of how to use it.*

However, when asked whether the training had changed their future teaching approach in any way, responses were mixed. Negative responses suggested trainees' individual difficulties with corpus use would prevent them from risking allowing their learners to experiment with corpora:

> Trainee R (Primary): *Speaking based on my experience during the training, I would say that it hasn't really changed my approach in my teaching because there are certain things that I am still not familiar with. At the same time, I'm so worried. If I keep continuing the approach which I'm still unfamiliar with, the students will also get even more confused by it.*

> Trainee F (Secondary): *I don't think it is suitable. The way to use it is complicated. How to access it …not easy. I myself sometimes do not know which one to use. Many functions to click ….so the students may be confused.*

Positive responses appeared to come from trainees who had not reported difficulties with corpus use following training. Rather, such trainees felt that learning about corpora had opened up new possibilities for their teaching practice and made lesson preparation easier:

> Trainee A (Secondary): *Some corpus like that can be used to learn something, and it makes the lesson somehow easier to understand, because of the [language] features that can be easily accessed.*

> Trainees S (Secondary): *Yeah, I can use corpus to find example sentence and use in context, if I were still using dictionary, the students will get bored[…]. So, it's better to use a technological assistance[sic].*

Trainees were then asked whether they felt corpora would be useful for the teaching of younger learners, again with mixed results. Positive responses tended come from secondary school trainees, and focused on how younger learners were already adept users of technology, or how corpora could help with common language-related questions younger learners frequently ask in the classroom:

> Trainee A (Secondary): *It's possible for young learners to use the corpus, because the corpus itself is very easy to use and operate even for young learners, and as we know that even younger learners can operate any device these days like smartphones or laptop/computers.*

> Trainees S (Secondary): *It's useful to answer kids' questions like "Why do verb 2 of 'go' is not 'goed'?" other than saying 'udah dari sananya' it's better to show them real life examples from corpora.*

Negative responses tended to come from primary school trainees, who suggested that the corpus platforms used during training were too complex for younger learners, the corpus data is too complex, or that, while useful, that teachers currently lack strategies for classroom implementation:

> Trainee J (Primary): *Yet, I am not certain whether I will use corpora for younger learners since they may not be familiar with the software. I am afraid that they will be confused. And I guess the teacher may […] consider [corpora] too sophisticated for children.*

> Trainee M (Primary): *To teach primary school students, the language in the corpus is not common for the primary school. For children, they may grasp when dealing with everyday activities, easier to remember[sic]. But in the corpus, the examples are seldom used. That's why it is not suitable.*

> Trainee R (Primary): *I hope it would be useful to teach them, as long as the teacher has […] I mean a creative and effective way to implement it, so that students won't get puzzled.*

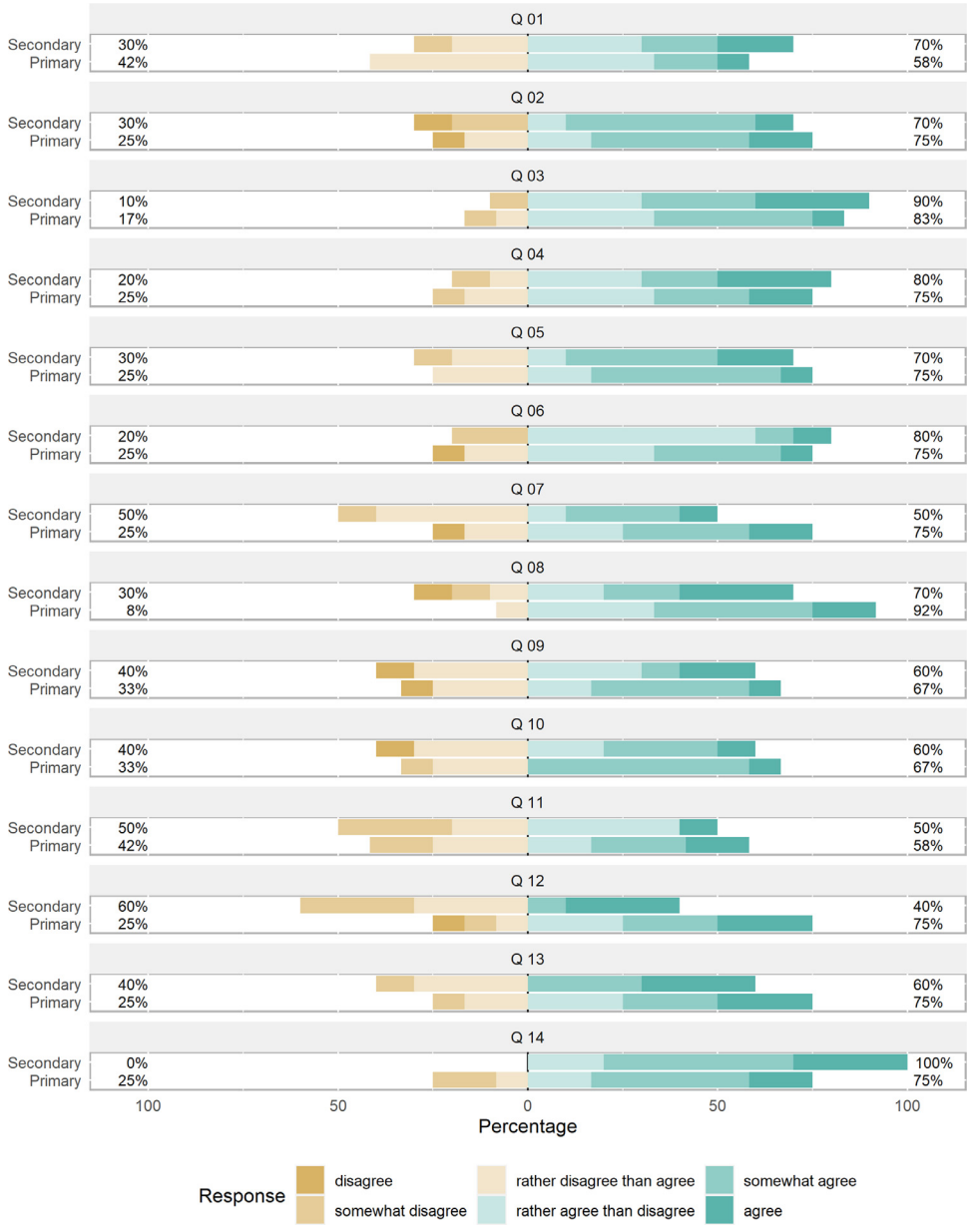

Fig. 5. Primary vs. Secondary school participant's responses regarding corpus literacy.

Trainee A (Primary): *From the training, I know how to use [corpora][…] But for the lesson plans, it does not help me. I feel it is more difficult for me to design the activities using the corpus. To adjust the LPs with corpus is challenging. I have to redesign my LPs. I am not sure what activities to design in my LPs.*

When asked how corpora might be useful for teaching, a number of trainees discussed the usefulness of DDL for teaching general grammar and vocabulary, although some trainees mentioned specific language features that they felt could be better targeted through DDL:

Trainee J (Primary): *Corpora is useful for teaching grammar since they provide a lot of examples that can benefit the students for understanding grammar. That is highly useful because I often have difficulties if I have to make my own sentences.*

Trainee F (Primary): *I can use the corpus for grammar but for the primary students …the grammar is not important. If I have to use the corpus, I may use it to show synonyms and antonyms because the examples are more accurate. Though I do not think it is useful for the primary school students, if I have to use it, I will use it for synonyms and antonyms.*

Trainee S (Secondary): *We can integrate corpora to teach, for instance, linking adverbials and show them the use in context, other than creating our own example.*

Trainees were then asked what younger learners would think of corpora, again with mixed results. Positive responses focused on DDL making learning more attractive, while negative responses focused on the complexity of DDL as a potential turn-off for younger learners:

Trainee A (Secondary): *In my opinion, young learners will find it fun to learn about the lesson using the corpus, because we know that young learners really like to learn with devices.*

Trainee O (Primary): *Some young learners may respond excitedly as they find something new and useful. However, there will be some who may not be familiar with the use of computer. It becomes challenges for them [sic].*

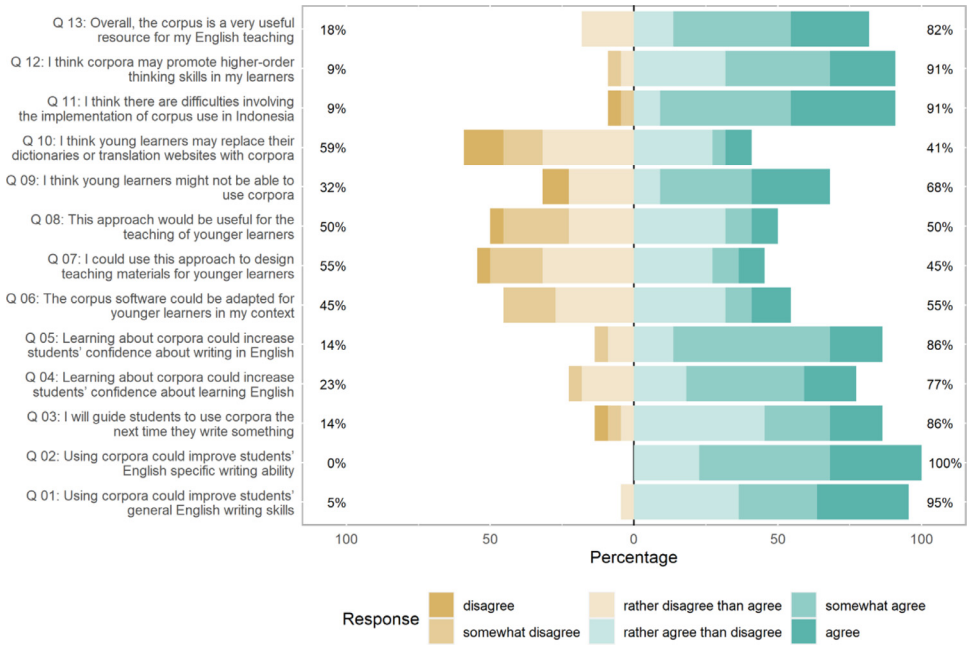

Fig. 6. Participants' perceptions of future integration of DDL into classroom practice.

Some trainees acknowledged that their own difficulties navigating corpora for DDL may not necessarily translate into younger learners experiencing the same difficulties:

> Trainee C (Secondary): *Young learners might think corpora is kind of hard to understand, to try, and way more complex than printed dictionaries or digital dictionaries, but might be their curiosity is more than the adult learners [sic], and they will find out something that [they] didn't explain before by themselves.*

> Trainee R (Primary): *In my opinion at first they will be overwhelmed because it is new, but with continuous adjustment from the instructor, I think they will slowly try to adapt themselves. They like to try something new.*

Finally, trainees were pressed on the potential issues Indonesian teachers may be faced with when trying to implement DDL in their teaching practice, as well as the potential barriers to DDL integration that might arise at the school and curricula levels in Indonesia. Four main issues were prominent in the data, namely a lack of available devices or policies preventing the use of devices in the classroom; a lack of internet connectivity; a perceived lack of space for CALL within the curriculum; and a lack of training and technical knowledge among Indonesian teachers. Regarding the first two, trainees suggested:

> Trainee A (Secondary): *In some area of the school there will be lack of some device such as computer or tablet that students can use to access the corpus […] there might be some places with a lack of signal so they will have difficulty to access the corpus.*

> Trainee C (Secondary): *Teachers are going to need the same number of devices as they have students, and all devices should work well. Not only the device but also the speed of the internet.*

> Trainee R (Primary): *One of the barriers that I could probably think of is […] that most of the schools here, particularly public schools don't really […] allow [students] to use their cell phones during the class.*

Regarding finding a space for DDL within the Indonesian curriculum, trainees commented:

> Trainee O (Primary): *I think the Indonesian curriculum may not fully support corpus use.*

> Trainee R (Primary): *For the curriculum which has already been standardized, [DDL] doesn't meet the requirement there.*

> Trainee A (Primary): *The school may have objections because they have to think of technology facility for the students. The curriculum …may be not ready because it is still based on the traditional system.*

However, with the exception of Trainee A's reference to the 'traditional system', they did not provide further specifics as to exactly why this might be the case, which is interesting as these are pre-service trainees still lacking specific knowledge of their teaching context. Regarding teachers' CALL capabilities, teachers' age was seen as a barrier to DDL integration in a number of comments:

> Trainee M (Primary): *The teachers' age and position may affect the decision to use the corpus. The senior ones may not want to try it.*

Trainees' also tended to make negative assumptions about the TPACK of Indonesian teachers in general, as seen in these comments:

> Trainee A (Primary): *The main barrier is the way teachers think. They still see it as too complicated for them to try.*

> Trainee S (Secondary): *I do believe there will be a lot of issues, but 'gaptek' issue probably is the most prominent one. A lot of teachers in Indonesia don't know to utilise technology deeper. Well, they might know how to Google or search videos in Youtube, but if it were a different type of software, and complicated enough like SketchEngine, they would think its troublesome to teach with it, so they don't incorporate corpus into their teaching.*

The latter comment is particularly interesting given that Trainee S's submitted DDL lesson plan demonstrated the highest possible TPACK, involving the careful, scaffolded integration of corpus use into the lesson objectives and an innovative 'game' on an interactive quiz website consolidating knowledge gained from corpus consultation. In this comment, Trainee S is admitting that while they were readily able to incorporate DDL into their teaching practice following training, they did not believe their abilities to be representative of the teaching population of Indonesia at large.

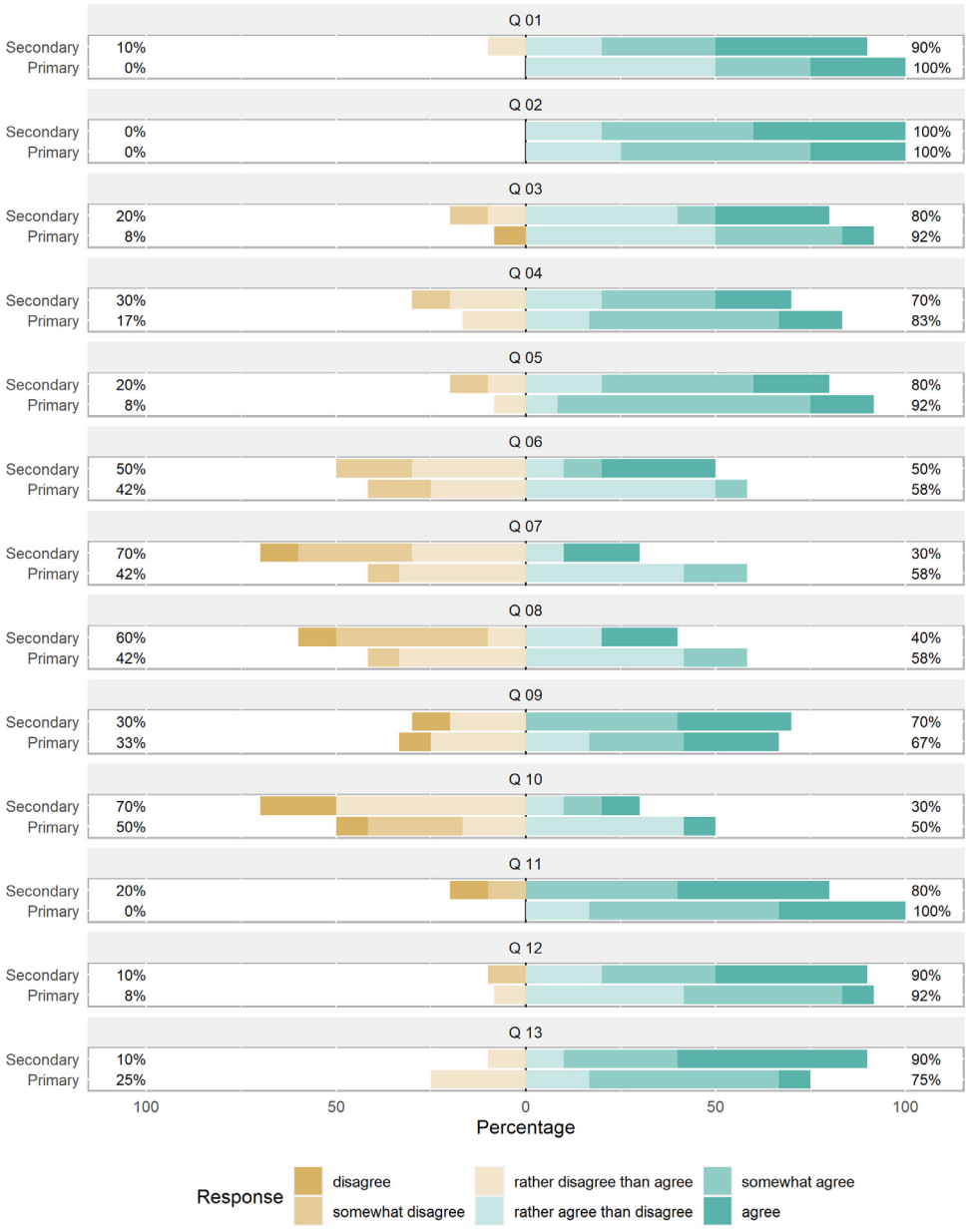

Fig. 7. Primary vs. Secondary school participant's responses regarding future classroom integration of DDL.

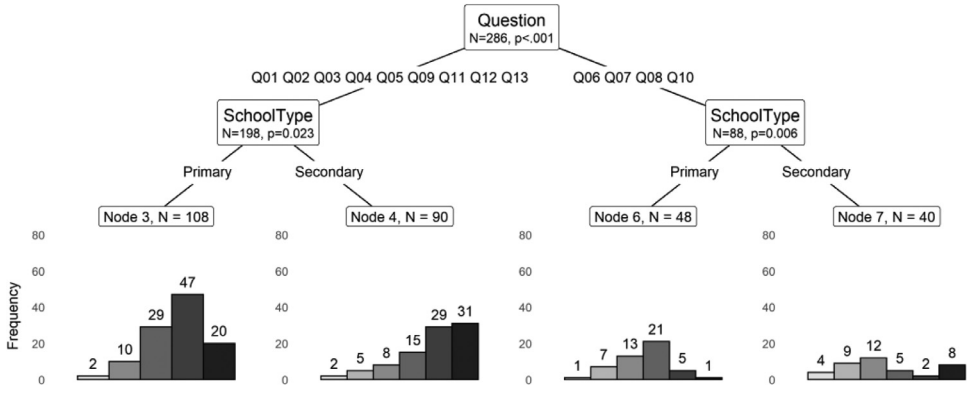

Fig. 8. Results of the Conditional Inference Tree analysis that was applied to the attitudinal data.

7. Discussion

The present study has provided a range of 'voices from the periphery' regarding the potential for integrating corpus use into classroom practice. The overall findings suggest that following a comprehensive (if short) online DDL training regimen, pre-service trainees at both primary and secondary levels generally appreciated the *potential* for corpora and DDL to improve the practice of language teaching and learning for pre-tertiary learners. However, this appreciation of what *could* be was tempered by trainees' acknowledgement – particularly trainees entering primary education - that integrating corpora into classroom practice in the Indonesian context would be a considerable challenge.

Regarding RQ1 ('perceptions of pre-service trainees about integrating DDL into classroom practice in Indonesia'), our predictions about trainee's willingness to adopt DDL based on previous related research and our review of the Indonesian CALL literature were largely supported in the data. In fact, the perceptions of these Indonesian trainees regarding the affordances and limitations of DDL following training also share many similarities with those of trainees in WEIRD contexts. Namely, trainees were appreciative of the training in terms of developing their own personal use of corpora, saw the potential value of corpora for lesson preparation and attracting learners' interest, believed that corpora could be useful for teaching both grammar and vocabulary, and that corpora could be particularly beneficial when used in conjunction with L2 writing lessons. Trainees however generally reported difficulty with the act of consulting a corpus and found analysing corpus data difficult, and, based on their own experiences, tended to believe that younger learners might experience the same problems. This was one of the central 'rational fears' preventing integration of DDL into mainstream classroom practice outlined in Boulton (2009), and this sentiment has also been found in related teacher DDL training studies in Western contexts, such as Schaeffer-Lacroix (2019).

Participants' perceptions as specifically related to the Indonesian context are particularly troublesome for the potential future implementation of DDL (if not CALL generally) in this region. Of course, one potential reason for trainees' unwillingness to adopt DDL in classroom practice may be that as pre-service teachers, they are as yet not fully aware of what is needed in their future professional careers, hence rejecting corpora before they fully understand how and when they could meaningfully be used. Literature in language teacher education is laden with studies highlighting the mismatch between pre-service teachers' expectations and teaching reality (see Zein, 2015 for such a study in the Indonesian context). Moreover, these trainees are likely unprepared and unready to shift their understanding of what constitutes learning as a result of their training on the pre-service program. They are fully aware of the urgency of adopting ICT in their teaching, but appear unable at present to consider how to integrate it as part of the curriculum. In addition, as pre-service teachers, while generally satisfied with the technical knowledge they gained from the training, they may still lack the pedagogical and content knowledge required to integrate DDL into classroom practice. An interesting finding was that while this study focused on pre-service trainees, some had already characterised the general Indonesian in-service teaching population as lacking the relevant TPACK required to successfully integrate DDL into classroom practice. This negative characterisation of colleagues in the profession is possibly formed from their previous experiences as students in the Indonesian public school system, including their interactions with in-service teachers working in that system. In a survey of in-service English teacher trainees across seven Indonesian provinces, Zein (2016) suggests current professional development lacks support for low proficiency in English, overemphasises theoretical over practical concerns for everyday teaching matters, and overemphasises urban vs. rural areas for professional development opportunities. Each concern potentially leads to weak pedagogical and content knowledge for language teaching among Indonesian teachers, further confounded by limited opportunities to develop the technological knowledge required for TPACK. If this is the 'status quo' of Indonesian teacher education, it is little wonder these trainees appear pessimistic about the likelihood of implementing DDL within language classrooms anytime soon.

Regarding RQ2 ('primary vs. secondary trainees' perceptions'), it is readily apparent that primary school trainees' were significantly less enthusiastic about integrating DDL into their future classroom practice in some (but not all) instances. As mentioned in the description of the training, primary trainees had collectively decided prior to the third online DDL workshop that they would not be submitting DDL-focused lesson plans as a final assessment, leaving only the secondary school trainees to do so – a significant and highly disappointing outcome given that in certain cases in the survey and interview data the primary school trainees' perceptions of corpus use were more positive than that of the secondary trainees.

One possible reason for the reluctance of the primary trainees to more fully engage with DDL is that primary school trainees' teacher education is centered around acquiring knowledge and skills relevant to child-oriented pedagogy, focusing on games, songs, child-friendly activities, and culturally appropriate teaching materials. In other words, educators at primary school level may be less concerned with pedagogical innovation than they are with creating a child-friendly learning environment (Rekalidou and Panitsides, 2015). Negative attitudes may also stem from their perceptions of possible disapproval from their school regarding adopting DDL for teaching, as the school is the main sociocultural environment the trainee is enter (Aizen, 1991). It may also be the case that these trainees do not fully understand the cognitive development of child learners in mediation with technology as part of an autonomous constructivist approach to learning. As a result, trainees tend to associate the use of corpora as too far a mental abstraction from their perception of children's ability to develop (Zein et al., 2020), so completely clinging to this perception that they would rather not allow their learners to attempt to use corpora 'for their own safety'.

It is also of course possible that the online training procedure may have asked too much of the primary trainees from a socio-constructivist perspective. While the trainees were provided scaffolded advice on their initial lesson plans and had the opportunity to ask questions to the corpus expert during the Zoom workshops, COVID-19 restrictions on in-class interaction meant that trainees with low TPACK were unable to receive hands-on assistance, scaffolding and support from their teachers or (perhaps more importantly) their more knowledgeable peers (Kaufman, 2004). The DDL approach to learning is itself considered from both cognitive-constructivist and socio-constructivist perspectives (O'Keeffe, 2020), and trainees who lack experience in the former and opportunities for interaction for the latter would likely have found the training regimen difficult, potentially leading to negative perceptions of corpora/DDL. In particular, having a more knowledgeable peer onboard to influence others' perceptions and decision-making in a particular direction has been shown to greatly influence trainees' attitudes towards DDL in similar studies (Schaeffer-Lacriox, 2019). The secondary school trainees were generally more active in all forms of online discussion and had two to three students who submitted final DDL lesson plans demonstrating the highest standards of TPACK, therefore these trainees' abilities and experiences may have positively influenced other group members to see DDL more favourably.

8. Conclusion

The present study has shown that, like those in WEIRD contexts, teacher trainees from a non-WEIRD context learning about corpora for the first time see the potential of DDL to dramatically improve the teaching and learning of second languages. However, and as also seen in studies conducted in WEIRD contexts, teacher trainees – particularly those entering service in primary education - still struggle to conceive of how to incorporate corpora and DDL into their teaching practice, even following comprehensive training on lesson planning for DDL. Then, adding in context-specific concerns including lack of resources, connec-

tivity, and a general lack of ICT/CALL support in teacher education, it becomes readily apparent that the field of DDL still has a long way to go before mainstream acceptance in pre-tertiary language education. Convincing language teacher trainees that DDL can – and will – significantly improve the authenticity of teacher-created input materials, develop their learners' linguistic problem-solving and technological capabilities, and lead to gains in vocabulary, grammar and phrasal knowledge in the L2 is still a considerable challenge.

However, even if in this case we appear to have lost the battle, this does not mean that we have lost the war. Recommendations for further research and language policy arising is that further cooperation between corpus linguists and teacher educators in Indonesia (and globally) is required if DDL is to find favour among mainstream pre-tertiary educators. Importantly for Indonesia, successful implementation in key urban environments such as Jakarta or government training agencies nationwide may then filter down to rural or less privileged environments (Zein, 2016). It is also increasingly apparent that targeted training and associated software innovations addressing how corpora/DDL can improve primary literacy (in L1 or L2) are required if primary school teachers are to consider implementing DDL in classroom settings. This study has shown that current DDL approaches/tools have not adequately addressed the knowledge base of primary school teacher trainees, have not acknowledged the classroom challenges facing such trainees, nor adequately reflected children's needs and the way they learn (Rekalidou and Panitsides, 2015). Further training also needs to determine how best to "mine the inevitable emotionality of novice teachers in purposeful and systematic ways that respond to the individual concerns of each teacher" (Golombek and Doran, 2014, p.110), if we are to learn how to reconcile novice teachers' evolving experiences with DDL and emerging identities arising from its use beyond their simply being expected to grasp the technological aspects only.

An obvious limitation of the study is that all training was conducted online due to COVID-19, leading to potential issues for trainees not used to online learning as mentioned previously. A further limitation of the present study is the short development window between completing the online training regimen and subsequent data collection. There have been a number of recent studies taking a longer longitudinal approach to trainee's perceptions of their training following a complexity theory perspective (e.g. Martin and Dismuke, 2018; Zein, 2016), tracking development over time as new discoveries are made, new experiences are gained, and new problems arise. It is therefore quite apparent that a follow-up study is needed as these trainees enter the profession upon graduating from their teacher education programme. There, once the teachers better understand the needs of their learners and the specifics of their teaching contexts, they may feel more confident to use corpora to enhance their teaching practice, and the learning practices of their students.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50, 179–211.

Alsop, S., Nesi, H., 2009. Issues in the development of the British academic written english (BAWE) corpus. Corpora 4 (1), 71–83. doi:10.3366/E1749503209000227.

Anthony, L., 2019. AntConc (Version 3.5.8) [Computer Software]. Waseda University. Available from, Tokyo, Japan https://www.laurenceanthony.net/software.

Baisa, V., Suchomel, V., 2014. SkELL: Web interface for English language learning. In: Horák, A., Rychlý, P. (Eds.), Proceedings of Recent Advances in Slavonic Natural Language Processing, Karlova Studánka, Czech Republic, pp. 63–70 5–7 December.

Ballance, O.J., Coxhead, A., 2020. How much vocabulary is needed to use a concordance? Int. J. Corpus Linguist. 25 (1), 36–61.

Bernardini, S., 2000. Systematising serendipity: proposals for concordancing large corpora with language learners. In: Burnard, L., McEnery, T. (Eds.), Rethinking Language Pedagogy from a Corpus Perspective. Peter Lang, Frankfurt, pp. 225–234.

Biggs, J., 1996. Enhancing teaching through constructive alignment. Higher Educ. 32 (3), 347–364.

Boontam, P., Phoocharoensil, S., 2018. Effectiveness of English preposition learning through data-driven learning (DDL). Southeast Asian J. Eng. Lang. Stud. 24 (3), 125–141. http://doi.org/10.17576/3L-2018-2403-10.

Boulton, A., 2009. Data-driven learning: reasonable fears and rational reassurance. Indian J. Appl. Linguist. 35 (1), 81–106.

Boulton, A., Cobb, T., 2017. Corpus use in language learning: a meta-analysis. Lang. Learn. 67 (2), 348–393. doi:10.1111/lang.12224.

Carter, R., McCarthy, M., 1995. Grammar and the spoken language. Appl. Linguist. 16 (2), 141–158. doi:10.1093/applin/16.2.141.

Chang, J.S., 2013. Introducing Linggle: from concordance to linguistic search engine. *Sponsors: national science council, executive yuan, ROC institute of linguistics*. Acad. Sinica NCCU Office Res. Devel. 12.

Chen, M., Flowerdew, J., Anthony, L., 2019. Introducing in-service English language teachers to data-driven learning for academic writing. System 87, 102148. doi:10.1016/j.system.2019.102148.

Cobb, T., 1999. Applying constructivism: a test for the learner-as-scientist. Educ. Technol. Res. Devel. 47, 15–33. doi:10.1007/BF02299631.

Crosthwaite, P., 2019. Data Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. Routledge doi:10.4324/9780429425899.

Crosthwaite, P., 2020. Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. Aust. Rev. Appl. Ling. 43 (2), 169–195. doi:10.1075/aral.00031.cro.

Crosthwaite, P., Stell, A., 2019. "It helps me get ideas on how to use my words": Primary school students' initial reactions to corpus use in a private tutoring setting. In: Crosthwaite, P. (Ed.), Data Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. Routledge, pp. 150–170. doi:10.4324/9780429425899-9.

Di Vito, S., 2019. Teaching French to younger learners through DDL. In: Author (Ed.), Data Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. Routledge, London, pp. 171–186. doi:10.4324/9780429425899-10.

Dreyfus, S.E., 2004. The five-stage model of adult skill acquisition. Bull. Sci., Technol. Soc. 24 (3), 177–181.

Ellis, N.C., 2006. Language acquisition as rational contingency learning. Appl. Linguist. 27 (1), 1–24.

Golombek, P., Doran, M., 2014. Unifying cognition, emotion, and activity in language teacher professional development. Teach. Teach. Educ. 39, 102–111. doi:10.1016/j.tate.2014.01.002.

Hirata, E., 2019. The development of a multimodal corpus tool for young EFL learners: A case study on the integration of DDL in teacher education. In: Crosthwaite (Ed.), Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. Routledge, pp. 88–105. doi:10.4324/9780429425899-6.

Hothorn, T, Hornik, K, Van de Wiel, MA, Zeileis, A, 2006. A lego system for conditional inference. Am. Statistic. 60 (3), 257–263.

Hothorn, T, Hornik, K, Zeileis, A, 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Statist. 15 (3), 651–674.

Hothorn, T, Zeileis, A, 2015. partykit: a modular toolkit for recursive partytioning in R. J. Mach. Learn. Res. 16, 3905–3909.

Johns, T., 1997. Contexts: The background, development and trialling of a concordance-based CALL program. In: Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (Eds.), Teaching and Language Corpora. Longman, London, UK, pp. 100–115.

Karras, J.N., 2016. The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. ReCALL 28 (2), 166–186.

Kennedy, C., Miceli, T., 2010. Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. Lang. Learn. Technol. 14 (1), 28–44.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Suchomel, V., 2014. The Sketch Engine: ten years on. Lexicography 1 (1), 7–36.

Kim, H., 2019. The perception of teachers and learners towards an exploratory corpus-based grammar instruction in a Korean EFL primary school context. Primary Eng. Educ. 25 (1), 123–152.

Koehler, M., Mishra, P., 2009. What is technological pedagogical content knowledge? Contemp. Iss. Technol. Teacher Educ. 9 (1), 60–70.

Latif, M.M.A, 2020. Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. ReCALL doi:10.1017/S0958344020000129, (volume, pages pending).

Levy, M., 1997. CALL: Context and Conceptualisation. Oxford University Press, Oxford.

Lin, L.-C., 2014. Learning word meanings from teachers' repeated story read-aloud in EFL primary classrooms. Eng. Lang. Teach. 7 (7), 68–81. doi:10.5539/elt.v7n7p68.

Lozano, A.A., Izquierdo, J., 2019. Technology in second language education: overcoming the digital divide. Emerg. Trends Educ. 2 (3), 52–70. doi:10.19136/etie.a2n3.3250.

Mahmud, K., Basalama, N., 2017. Computer-Assisted Language Learning: The English Education Program for Student-Teachers. ZAHR Publishing, Yogyakarta.

Marshall, C., Rossman, G.B., 2010. Designing Qualitative Research, 5th Ed. Sage, Thousand Oaks, California.

Martin, S.D., Dismuke, S., 2018. Investigating differences in teacher practices through a complexity theory lens: The influence of teacher education. J. Teacher Educ. 69 (1), 22–39.

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: Corpora and DDL for Pre-tertiary Learners. Routledge, pp. 13–31.

Ministry of National Education Indonesia (2013). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia*.

Montag, J.L., Jones, M.N., Smith, L.B., 2015. The words children hear: Picture books and the statistics for language learning. Psychol. Sci. 26 (9), 1489–1496 https://doi.org/10.1177%2F0956797615594361.

O'Keeffe, A., 2020. Data-driven learning–a call for a broader research gaze. Language Teaching 1–14. doi:10.1017/S0261444820000245.

Özbay, A.S., Olgun, O., 2017. The application of DDL for teaching preposition collocations to Turkish EFL learners. Int. J. Res. Teacher Educ. 8 (3), 1–10.

Piaget, J., 1970. Science and Education and the Psychology of the Child. Orion Press, New York, NY.

Rachmawati, U., 2016. Computer assisted language learning (CALL) as EFL teaching and learning media in indonesia: opportunity and challenges. Jurnal Edulingua 3 (2), 47–56.

Rekalidou, G., Panitsides, E.A., 2015. What does it take to be a "successful teacher"? Universities' role in preparing the future early-years workforce. Early Years 35 (4), 333–350. doi:10.1080/09575146.2015.1080231.

Ridwan, A., 2017. CALL (Computer Assisted Language Learning) as a means professional development of EFL teachers in Indonesia: a Vygotsky's sociocultural perspective. SELL J. 2 (2), 162–172.

Roberts, T.A., 2008. Home storybook reading in primary or second language with preschool children: Evidence of equal effectiveness for second-language vocabulary acquisition. Read. Res. Q. 43 (2), 103–130. doi:10.1598/RRQ.43.2.1.

Saeedakhtar, A., Bagerin, M., Abdi, R., 2020. The effect of hands-on and hands-off data–driven learning on low-intermediate learners' verb-preposition collocations. System, 102268.

Schaeffer-Lacroix, E., 2019. Barriers to trainee teachers' corpus use. In: Crosthwaite, P. (Ed.), Data-Driven Learning for the Next Generation: Corpora and DDL for Younger Learners. Routledge, pp. 47–64.

Sealey, A., Thompson, P., 2004. What do you call the dull words?'Primary school children using corpus-based approaches to learn about language. Eng. Educ. 38 (1), 80–91.

Sikki, E.A.A., Rahman, A., Hamra, A., Noni, N, 2013. The competence of primary school English teachers in Indonesia. J. Educ. Practice 4 (11), 139–145.

Sinclair, S., Rockwell, G., 2016. *Voyant Tools* [Computer Software] http://voyant-tools.org/.

Son, J.B., Robb, T., Charismiadji, I., 2011. Computer literacy and competency: a survey of Indonesian teachers of English as a foreign language. Comput.-Ass. Lang. Learn. Electron. J. (CALL-EJ) 12 (1), 26–42.

Soruç, A., Tekin, B., 2017. Vocabulary learning through data-driven learning in an english as a second language setting. Educ. Sci., Theory Practice 17 (6), 1811–1832. doi:10.12738/estp.2017.6.0305.

Syaifudin, M., van Rensburg, H., 2018. Considerations for the development of computer-assisted language learning (CALL) teacher training course: a practical experience from a call course development in Indonesia. *Arab World Eng. J.* (AWEJ) Special Issue on CALL 4. doi:10.24093/awej/call4.7.

Taghizadeh, M., Hasani Yourdshahi, Z., 2019. Integrating technology into young learners' classes: language teachers' perceptions. Comput. Ass. Lang. Learn. 1–25. doi:10.1080/09588221.2019.1618876.

Tyne, H., 2012. Corpus work with ordinary teachers: data-driven learning activities. In: Thomas, J., Boulton, A. (Eds.), Input, Process and Product: Developments in Teaching and Language Corpora. Routledge, London, pp. 707–712.

Vijjala, S., Meurers, D., 2012. On improving the accuracy of readability classification using insights from second language acquisition. In: NAACL HLT '12 Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, Montreal, Canada, pp. 163–173 June 07, 2012.

Wicher, O., 2019. Data-driven learning in the secondary classroom: a critical evaluation from the perspective of foreign language didactics. In: Crosthwaite, P. (Ed.), Data Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. Routledge, London, pp. 31–46.

Yoon, H., Hirvela, A., 2004. ESL student attitudes toward corpus use in L2 writing. J. Second Lang. Writing 13 (4), 257–283.

Zein, S., 2015. Factors affecting the professional development of elementary English teachers. Prof. Devel. Educ. 42 (3), 423–440.

Zein, S., 2016. Government-based training agencies and the professional development of Indonesian teachers of English for Young Learners: perspectives from complexity theory. J. Educ. Teach. 42 (2), 205–223. doi:10.1080/02607476.2016.1143145.

Zein, S., Sukyadi, D., Hamied, F.A., Lengkanawati, N.S., 2020. English language education in Indonesia: a review of research (2011–2019). Lang. Teach. 53 (4), 491–523.