Data-driven learning and young learners: Perceptions and attitudes of Japanese elementary school students toward EFL corpus-based activities

Castledown

Martin Spivey
Akita International University
mart.spiv@gmail.com

There is an increasing amount of research into corpus-based study as a peda-gogical approach in language learning. This is most commonly known as data-driven learning (DDL) and can refer to both computer-based and paper-based activities. Despite the growing interest, the vast majority of published research into DDL investigates university students and there is a serious lack of inquiry into children and adolescents. This paper seeks to add to the limited knowledge that currently exists on pre-adolescent learners. Present research into young learners primarily investigates corpus-supported or paper-based learning; therefore, the aim of this particular study is to compare the latter with direct cor-pus use in order to help improve our understanding of young learner attitudes toward both forms of DDL. A small-scale, mixed-methods study was designed to investigate the perceptions and attitudes of Japanese sixth-grade elementary school students toward EFL corpus-based activities. Five participants took part in a month-long course of four one-hour online sessions, with data collected via a pre-study survey, post-class questionnaires, and post-study interviews. Results show that students display a very favorable attitude toward DDL activities, with a clear preference for paper-based activities over computer-based activities. The findings also highlight the importance of 'scaffolding' in young learner educa-tion and the potential obstacles that must be overcome to ensure that DDL activities can be successfully implemented in the EFL young learner classroom.

Keywords: attitudes, corpora, data-driven learning, Japan, young learners

Introduction

'Data-driven learning' (DDL) is a term coined by Tim Johns to describe "the use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language" (Johns & King, 1991, p. iii, as cited in Boulton, 2021). Concordance lines are extracted from a corpus, which is "a collection of spoken or written texts to be used for linguistic analysis and based on a specific set of design criteria influenced by its purpose and scope" (Weisser, 2016, p. 13). Johns's approach can be described as a constructivist or 'hard' form of DDL that places students at the center of the learning process. As DDL has become more widespread, a socio-culturally focused or 'soft' approach has developed that is more teacher-led (Hadley & Charles, 2017). In the past two decades, research into DDL in language learning education has grown significantly (Boulton, 2012; Chen & Flowerdew, 2018; Pérez-Paredes, 2022). Many of the studies to date have focused on ESP (Boulton, 2012) and EAP (Chen & Flowerdew, 2018), invariably with university students as the subjects under scrutiny.

Pre-tertiary education is an under-researched field and scholarly inquiry into corpus-based study and school-age learners is particularly sparse (Crosthwaite, 2020). There are a few studies into adolescents and DDL, including Braun (2007), Frankenberg-Garcia (2014), and Szudarski (2020). For even younger age groups, while some research does exist on pre-adolescent children learning through corpus-based activities, it is limited to a small number of journal articles and book chapters (Crosthwaite & Stell, 2020; Kim, 2019; MacGregor, 2014; Sealey & Thompson, 2004, 2006, 2007; Takahashi & Fujiwara, 2016).

This paper aims to contribute to the existing literature on DDL and young learners by examining the attitudes of Japanese eleven and twelve-year-olds toward both direct and indirect corpus activities in the EFL context. A greater understanding is sought on how students react to DDL when learning a foreign language and what, if any, contrasting attitudes there might be between learning directly with corpus software or through corpus-based worksheets. This is done via a small-scale study of five sixth-grade elementary school students undertaking a four-week course on simple verb and noun collocations (i.e., *study science, do judo,* and *like pizza*) that feature in the fifth and sixth-grade elementary school English textbooks *New Horizon Elementary 5* (Allen-Tamai et al., 2020a) and *New Horizon Elementary 6* (Allen-Tamai et al., 2020b). Data were obtained through a pre-study survey to gauge the participants' English level and technological experience, four post-class questionnaires to gain immediate feedback on the activity just completed, and a post-study interview to elicit student perspectives of the study as a whole and their own pedagogical preferences. It is hoped that the results (and subsequent analysis) can provide insight for EFL practitioners who may be considering implementing a DDL approach in the young learner classroom. Henceforth, 'young learners' will refer to any child up to and including elementary school age.

Literature review

Data-driven learning

In recent years, meta-analytical studies have found some evidence of statistically significant learning gains as a result of studying language through DDL activities (Boulton & Cobb, 2017; Lee et al., 2019; Mizumoto & Chujo, 2015). At this point, it would be useful to consider what DDL actually entails, and O'Keeffe's (2021a) DDL research framework presents us with a clear visual representation (Figure 1). It shows the two opposing forms of DDL, from *constructivist-focused* on the left side of the cline to a more *socio-culturally focused* form on the right.
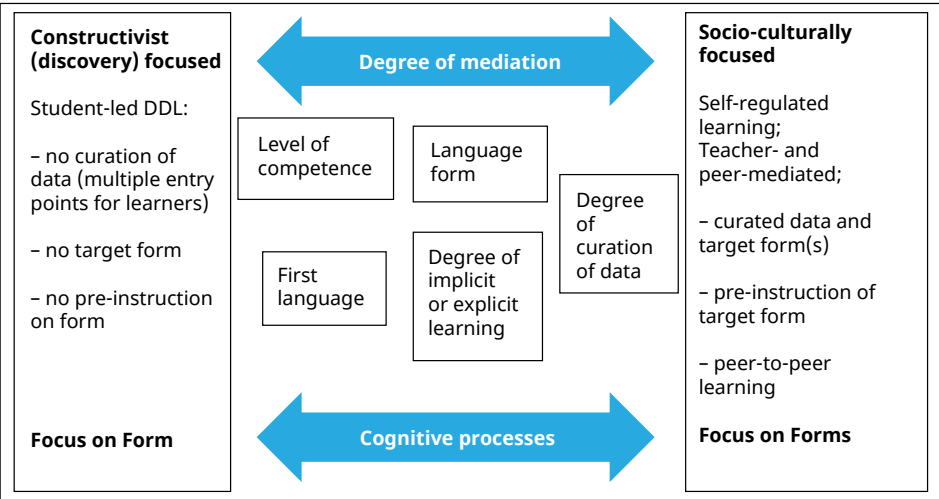
Figure 1. A research framework for DDL (O'Keeffe, 2021a, p. 269)

The upper arrow represents the degree of mediation and the lower arrow represents the cognitive processes involved in the pedagogical approach. Constructivist-focused DDL typically involves students engaging with the corpus (or corpus data) directly to seek answers to language inquiries. In computer-based activities, corpora can be accessed through computer software and learners investigate the data by conducting their own searches and examining concordance lines for evidence of frequent language patterns. Such activities usually require learners to possess a high L2 proficiency level and the ability to learn autonomously with little or no teacher input. Learners act as 'researchers' investigating the language and producing their own hypotheses; therefore, this approach is likely to be most successful with motivated learners who have the intellectual capacity to cope with the high demands of the tasks.

Contrasting with constructivist-focused DDL is a more socio-culturally focused approach that can reduce the cognitive load for learners of a lower language level or age. In this approach, DDL is typically conducted through paper-based activities and students may not actually consult a corpus directly. The teacher might produce a page of concordance lines taken from a corpus

for students to highlight particular words or design a worksheet using concordance lines that have been simplified for ease of understanding. This is a more deductive and teacher-focused form of pedagogy that can still benefit students as they are presented with examples of (mostly) authentic, natural language. A brief background to socio-cultural theory now follows.

Socio-cultural theory

Socio-cultural theory (SCT) originates from Lev Vygotsky, a Russian child psychologist who theorized about how children learn and argued for the importance of mediation in their development (Vygotsky, 1978). His teachings have been applied across the field of SLA, and one of his most pertinent concepts in foreign language instruction is the zone of proximal development (ZPD). The ZPD "represents the distance between what individuals can do on their own and what they can do with assistance from another individual" (Loewen & Reinders, 2011, p. 181). In a classroom context, this can relate to the support a teacher (or peer) may provide to a student that can allow them to do something they could not otherwise do by themselves. Another term is 'scaffolding', which can take various forms. In the context of DDL instruction, activity instructions or a language point explanation may be given in the students' L1 for ease of understanding. It could also involve a pedagogical approach that encourages classmates to support each other in completing a worksheet of simplified corpus data. Such an approach is likely the most effective one to take with young learners as it provides them with the pedagogical and emotional building blocks on which to develop their language skills. A description of some of the current studies on young learners and DDL is presented below.

L2 young learners and DDL

There is a dearth of empirical studies on corpus-based study with young learners in language education. Some research has inquired into the effects of DDL study on L1 learners of elementary school age and found some positive effects on learning (Crosthwaite & Stell, 2020; Sealey & Thompson, 2004, 2006). There are also a few studies on DDL and L2 young learners.

MacGregor (2014) investigated the use of DDL activities with 32 Japanese students aged 9–15 at a private English conversation school. During 35 hours of lessons over 12 months, students were observed while partaking in various activities created from storybook texts. Games were designed to introduce students to the idea of gathering data and calculating frequency and some activities involved data taken from wordlists and concordances. Each student completed a post-activity feedback sheet consisting of yes/no and ranking questions, with space to add more detailed opinions. These sheets were mostly written in the students' L1, ensuring that language was no barrier to data collection. When asked, "Did you enjoy this activity?", an average score of 7.5 was given (1=*no fun*, 10=*lots of fun*). The results of this study are encouraging and suggest that a simple, scaffolded approach increases the chances of young learners

developing positive attitudes toward DDL. However, this study fails to provide insight into *direct* corpus use among elementary-school-age children, which may potentially have a significant impact on these students' perceptions of DDL.

In another study on Japanese learners, Takahashi and Fujiwara (2016) explored the potential of inductive learning through DDL activities with 52 sixth-grade students in an elementary school class. Students were divided into an experimental group (n=23) and a control group (n=24). The experimental group combined oral activities with a DDL-informed worksheet. Results from two post-study quizzes showed a significant increase in scores from the experimental group compared with the control group in the immediate post-test, with a slight score decline in the delayed post-test 17 days later. A post-study questionnaire found that the activities were well-received overall, with most respondents commenting on how they enjoyed the activity as it was in a game format. A number of students stated that they liked working in a group, and some mentioned the role that the teacher played in helping to facilitate understanding. Although a short-length study, the results do tentatively indicate that paper-based DDL can promote language retention and motivation in younger learners.

In the Korean context, Kim (2019) conducted an exploratory, qualitative study on the perceptions of teachers and 12-year-old students learning prepositions via paper-based DDL activities. 18 students completed two sessions and data were collected via questionnaires and post-session semi-structured interviews. Students were required to read through a worksheet of edited concordance lines and find patterns in collaboration with their group, followed by answering questions and producing their own sentences. The students generally found the lessons to be "fun", and some claimed to enjoy working with their peers. There were a few comments on experiencing difficulties due to the challenging vocabulary, and some students suggested that learning through games would make the class more enjoyable. The general student feedback appears to be quite positive toward the use of corpus-based activities; however, this study evaluated children's perceptions of DDL after only two short sessions and a longer research period may have produced different results.

While there is evidently a serious lack of research on elementary-school-age learners and corpus-based language study in both L1 and L2 learning contexts, the aforementioned studies provide some evidence that pre-teen children generally have positive attitudes toward studying English through DDL activities. However, the type of DDL examined in these studies can be placed toward the socio-culturally focused end of the DDL cline (corpus-informed games, curated concordance lines, L1 support, etc.). The current research fails to inform us of the extent to which a computer-based approach may be received by young learners of elementary school age, a gap that the present study seeks to fill to a certain degree.

Research questions

In order to gain a greater understanding of young learner perceptions of both computer-based and paper-based DDL activities, a small-scale, mixed-methods study was designed to answer the following 3 research questions:

RQ1. What are the general attitudes of Japanese young learners toward DDL activities?

RQ2. Is there a significant difference in attitudes toward computer-based activities compared to paper-based activities?

RQ3. What are the students' perceptions of the classroom role of the teacher and peers? What do these suggest about the impact 'scaffolding' can have on the successful implementation of DDL activities with Japanese young learners?

Study

Research design

A weekly course of four 60-minute lessons was held online via the videoconferencing software Zoom in April 2022 due to the uncertainty surrounding the ongoing COVID-19 pandemic. While this required students to participate from their homes rather than face to face in a classroom, it was felt that this would not negatively impact the study significantly. A mixed-methods approach was taken, with both quantitative and qualitative data collected via a pre-study survey and post-class feedback questionnaires using Google Forms and a post-study structured interview with each participant individually. The data triangulation supports the analytical process and enables the results to have greater veracity. The pre-study survey was designed to create a learner profile of each participant that provides the researcher with a rough overview of their English and technological capabilities. The data are used for reference when discussing the results of the post-class questionnaires completed shortly after the DDL session. The post-study interview data can be compared to the post-class responses to see if they correlate or if there are any discrepancies to consider.

Participants

The participants were five 11 and 12-year-olds who had just started their final year of elementary school in Japan and were recruited through the researcher's acquaintances in addition to a call for participants that was placed on social media and the researcher's own school website. For practical purposes, a "convenience sampling" approach to recruitment was taken (Dörnyei, 2007). This enabled the researcher to gather as many willing participants as possible. The main requirements were that students were in the sixth grade of Japanese elementary school and had direct access to a computer and the internet for the study's duration. A pre-study survey was administered to determine the

The
JALT CALL
Journal
vol. 19 no.3

Spivey: Data-driven learning and young learners

399

potential participants' suitability for the course and to collect biodata, such as their English proficiency level, technological experience, and study habits. Table 1 presents a basic profile of each participant.

Table 1. Participant profiles

Student	A	B	C	D	E
Age	12 years	11y 5 months	11y 9 m	11y 11m	11y 7m
Gender	M	M	M	F	F
Started Learning English (age)	9–10	6–8	3–5	0–2	11
English Study (hours per week)	2–3	1–1.5	3+	0.5–1	2–3
EIKEN Test	N/A	Pre-2nd	Jr. Gold	N/A	N/A

Note. M = Male, F = Female

As the table shows, the participants have received varying levels of exposure to English, with some students starting before elementary school and others in later years. Student B has passed the EIKEN Pre-2nd test, an English test administered in Japan that is equivalent to the CEFR A2 level and targeted at senior high school 2nd-year students. Student C has passed Jr. Gold, the highest level of a three-stage English test for Japanese elementary school students. As for the other three participants, Student D began studying English as an infant and Student A started prior to the fifth grade. Student E has the least experience learning English, having started at the age of 11, though she studies between two and three hours a week.

Data collection

Research ethics

Prospective participants were sent an informed consent form providing further information about the research and outlining the ethical guidelines ensuring anonymity, confidentiality, and the right to withdraw from the study at any time. The study participants and their legal guardians were both required to sign a form confirming that they understood its details and gave their full consent to participate. In primary research with children, it is imperative that the participants fully understand what the research involves and explicitly agree to take part (Pinter, 2015); therefore, an explanation in Japanese was also provided by the researcher at the beginning of the first class. It was explicitly stated at that point that students could withdraw at any time during the research study and all of their data would be deleted.

Pre-study survey

Once informed consent was received, a pre-study survey was sent to each participant via Google Forms that requested biodata, such as the participant's full

name (for administrative purposes only), date of birth, and gender, as well as details about previous English language education and experience with technology. Appendix A presents an English version of the set of questions given.

The survey consisted of open-ended questions, multiple-choice questions, and a single Likert-scale question asking participants to rate their elementary school English lessons on a scale of 1–7 (1=*very boring*, 7=*very interesting*). The questions were created and translated into Japanese by the researcher and designed for pre-teen respondents to be able to answer freely and without difficulty, facilitating maximum data collection. The main aim of the survey was to gain an understanding of the students' English proficiency levels, study habits, and previous experience with technology. This would not only inform the results of the research study but also assist the researcher in preparing the class activities. Time limitations meant the survey could not be piloted; however, the quality of the responses suggests that participants had no major comprehension issues.

Post-class questionnaires

After each class, a questionnaire was distributed to participants requesting their feedback on the lesson. This was considered to be the most suitable form of data collection in order to gauge student perceptions of and attitudes toward the activities. As mentioned above, there is a lack of research on young learners and DDL, and previous studies fail to offer suitable templates for question compilation. Therefore, a decision was made to produce original questions that would best suit this study's needs and elicit responses that could produce valuable insight into young learner perceptions of DDL study. The questionnaire was designed for simplicity, collected both qualitative and quantitative data, and consisted of 12 Likert-scale, multiple-choice, and open-ended questions. The questions were written in English (Appendix B), translated into Japanese by the researcher and then checked by a native speaker. Responses were typically received within two hours of the lesson's conclusion, increasing the potential for accurate answers. Following the Likert-scale questions, open-ended questions were then posed to extract more detailed thoughts and opinions on task satisfaction and difficulty as well as teacher and peer support.

Post-study interviews

To strengthen the veracity of the data obtained from the post-class questionnaires, an interview was conducted with each participant individually after course completion (see Appendix C for questions). It has been argued that interviews with children should be conducted in groups to reduce the power differential between the (adult) researcher and the research subjects as well as provide a space for the interviewees to construct meaning with their peers (Eder & Fingerson, 2003). However, it has been found that some children are more willing to talk about themselves in a private interview setting (Lewis, 1992, as cited in Heary & Hennessy, 2006) and, as the interviews were held online

and the participants lacked familiarity with each other (beyond the confines of the study), it was deemed more appropriate to conduct the interviews privately with each student. Within two weeks of the study, a structured interview of approximately 10 minutes was conducted with each participant via Zoom. All interviews were recorded and then transcribed using online transcription software (Sonix, 2023). The replies to the questions were then translated into English by the researcher and checked by a native Japanese speaker for any possible discrepancies.

Nine questions were designed to elicit answers from participants on their attitudes toward the lessons and corpus software, which could be corroborated with and compared to data from the post-class questionnaire. The questions were prepared prior to the interview and presented in the same order for each interviewee in order to enhance the reliability and validity of the data obtained (Rose et al., 2019).

The DDL course

A course of four weekly 60-minute classes was created for this research study. To accommodate as many students as possible, the class was held twice a week and research participants were given the option of attending either session. The same lesson plan was used in both classes. Aside from Class 1A, which only Student A attended, all sessions were group classes. Students A and D attended Session A and Students B, C, and E attended Session B. Four activities were designed and produced by the teacher–researcher, with a balanced mix of computer-based and paper-based activities. Verb–noun collocations were chosen as the language feature under study and lists of verbs and nouns were compiled from the Japanese fifth and sixth-grade elementary school English textbooks *New Horizon Elementary 5* (Allen-Tamai et al., 2020a) and *New Horizon Elementary 6* (Allen-Tamai et al., 2020b).

The websites *Sketch Engine* (2023) and *ddl-study* (Nishigaki, 2022) were selected for the computer-based tasks. Through *Sketch Engine*, access was obtained to the CHILDES English Corpus (22.7 million words of mostly spontaneous child speech). *Ddl-study* is a Japanese website that presents example sentences of English and is designed for child learners.

All the activities were created specifically for elementary school children by the researcher. Simple tasks, such as gap-fills or target word highlighting with colored pencils, were created because of their familiarity and the low level of explanation required. The simplification of concordance lines for easier comprehension was an idea taken from Kim (2019), and the color-coding exercises were inspired by Sealey and Thompson (2004).

Results

RQ1 General attitudes toward DDL

The data from the post-class questionnaires and the post-study interviews show that the research participants hold a very favorable view of the classes undertaken as part of the research study. Figure 2 shows the total scores given by each participant as well as the average score for each class.
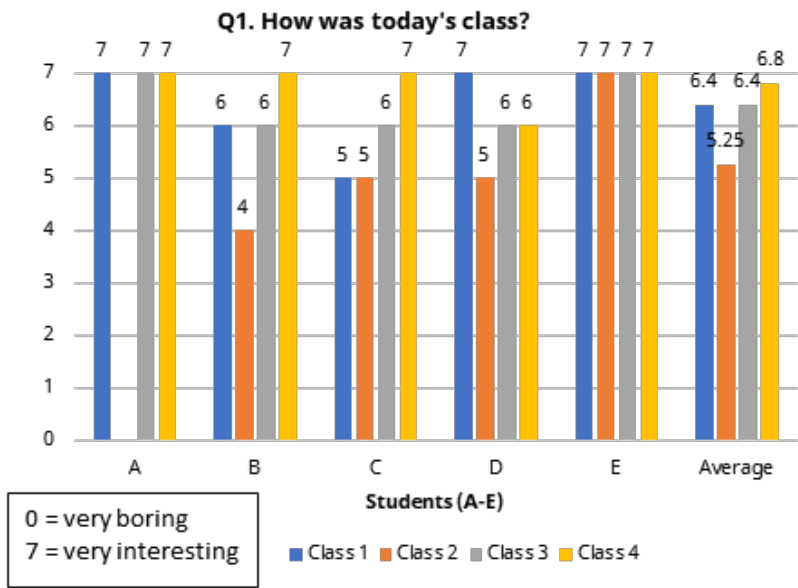
Q1. How was today's class?

Figure 2. Student attitudes toward the DDL classes

Full scores ('very interesting') were given in ten out of nineteen responses (one response is missing as Student A failed to respond for Class 2). The lowest score chosen was by Student B for Class 2 and that was a 'neutral' score of 4 so it cannot be said that any student indicated strong, negative feelings toward any particular lesson. Based on the average score per class, Class 4 was the most popular (6.8) and Class 2 was the least popular (5.25). Interestingly, there is a discrepancy here because when students were asked to rank the activities in order of preference in the post-study interview, four students selected Activity 2 (in Class 2) as their favorite. Table 2 presents the full rankings.

Table 2. Student rankings of the four activities

Activity	Student A	Student B	Student C	Student D	Student E
Most favorite	2	2	4	2	2
2nd favorite	4	4	2	4	4
3rd favorite	3	1	3	1	3
Least favorite	1	3	1	3	1

A more detailed analysis of each activity will be provided in the section on RQ2;

Spivey: Data-driven learning and young learners

however, student feedback suggests that the level of difficulty of Activity 2 may have led to the relatively lower scores given in the immediate post-class questionnaire. Student E commented that "it has been the most difficult activity so far but I was happy when I was able to understand after discussing with the other students," while Student B said "it was difficult but, well, I could understand what the subject predicate was." Figure 3 shows the students' perceptions of the level of difficulty of each class.
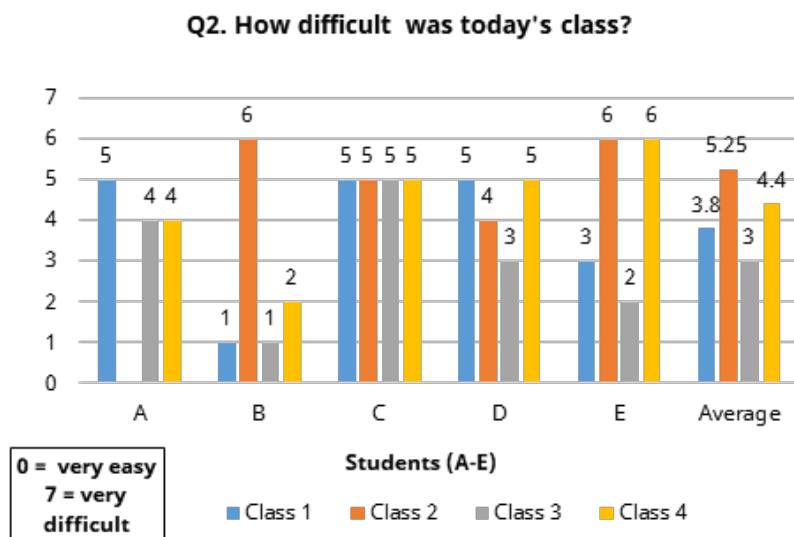
Figure 3. Class difficulty ratings

Student B judged Class 2 to be easily his most challenging class and Student E found it her joint most difficult class. Student B arguably has the highest English ability in the class, whereas Student E has the least amount of English language experience, having started her studies in the fifth grade of elementary school. Curiously, the level of task difficulty does not appear to have had a significant influence on students' enjoyment of the lesson and the discrepancy outlined above may be due to the length of time after lesson completion. Although speculation, in the post-class questionnaire students may have felt tired and demotivated and, therefore, given a lower rating. However, perhaps they could reflect on the experience with a more positive outlook after a few weeks.

The data suggest that students tended to find the classes both interesting and fairly challenging, and so one cannot claim that the learners enjoyed the lessons because the activities were too easy and required little effort. Students A, D, and E specifically mentioned the study of verb and noun collocations as the most memorable aspect of the course in response to Question 1 of the post-study interviews, while Students B and C referred to working with others as a key part of the study for themselves in answer to the same question. Overall, based on participant feedback, the DDL activities were well-received and the level of difficulty actually appears to positively correlate with the students'

level of enjoyment. This suggests that EFL learners in this particular age group can be presented with more challenging DDL tasks and it is unlikely to have a significantly negative impact on how they perceive the activities.

### RQ2 Computer-based versus paper-based activities

The second research question attempts to discover if there are any noticeable differences between student attitudes toward the two types of DDL: computer-based and paper-based activities. Activities 1 and 3 involved the direct use of a corpus, whereas Activities 2 and 4 were predominantly paper-based activities.

The activity rankings made in reply to Question 4 of the post-study interviews show that the students preferred the paper-based DDL activities over the computer-based activities. Activities 2 and 4 were selected as either the favorite or second-favorite activity, while Activities 1 and 3 were the least popular. In answer to Question 1 in the post-class questionnaire, students claimed to enjoy Activity 4 the most with an average rating of 6.8; however, as discussed above, Activity 2 had the lowest average rating (5.25). When asked directly which type of activity they preferred, four out of the five participants chose the paper-based activities. Answering Question 6 in the interview, Student D said that "the worksheets were easier to answer," while, according to Student C, "on the computer, using the search box, there were moments when I didn't know how to use it, but I always use worksheets and they're easy to understand and do and I like writing, so I think I prefer paper[-based activities]." Student B claimed a preference for computer-based activities in Question 6, though it is intriguing that he contradicted himself in Question 4 by choosing the two activities as his least favorite. He stated that he found using the *ddl-study* website in Activity 3 too easy as it was in Japanese and also gave Activity 1 a difficulty score of 1 (one rank above 'very easy') in the post-class questionnaire. This could suggest that, due to his higher English ability, a more challenging activity involving the use of the CHILDES English Corpus on *Sketch Engine* may have been of more interest to him than the tasks in Activity 1.

In such a short-length study, there was limited time in which to build confidence with computer usage as well as direct corpus use; therefore, while the classes with computer-based activities were still rated fairly highly (both 6.4 on average), it is perhaps not too surprising that most students showed a preference for the paper-based activities that used simple task formats they were already familiar with.

RQ3 Student perceptions of teacher and peer roles

In Question 10 of the post-class questionnaire, students were asked to rate the teacher's support in each class. Figure 4 displays the answers to the 7-point Likert-scale question. This was followed by an open-ended question (Question 11) asking for the participants' opinions on what they liked about the support or how they felt improvements could be made.

Q10. How would you rate the teacher's support today?
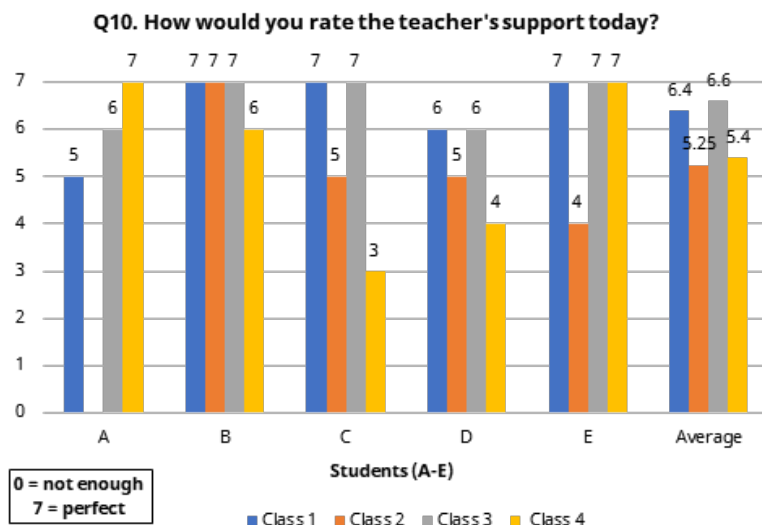
Figure 4. Teacher support ratings

The average rating for each class ranged from 5.25 to 6.6, so a conclusion can be drawn that students were generally happy with the support provided by the teacher in class. It is interesting to note that the lowest scores were given for Activities 2 and 4, despite these being the two most popular activities stated in Question 4 of the post-study interview. Student C gave a low score of 3 for Class 4 and stated that "the explanation in Japanese was easy to understand. I want an easy-to-understand explanation in simple English." Student D only gave a score of 4 for Class 4 but said that there is nothing she would change about the amount of support from the teacher. However, she did claim that "[a] good point was that I could check my answers together with the other students," so perhaps in this particular activity she had a preference for peer support and less teacher intervention.

In the classes, the teacher–researcher communicated in both Japanese and English in order to help ensure better student comprehension. In Question 2 of the post-study interview, the participants were asked for their feelings about the teacher's use of the students' L1 during the classes and all were in agreement that it was helpful. Student D thought that "I think it would have been difficult [if the class had been all in English]," while Student C mentioned parental support too: "…I can later search things that have been said and understand various things after asking my parents." Two students said they preferred using the *ddl-study* website more than *Sketch Engine* and ease of understanding was a factor as the website instructions and menu are in Japanese. Student D commented that she liked both websites; however, she also claimed that *ddl-study* was easier to understand than *Sketch Engine* due to the language factor.

When asked directly about study preferences in the post-study interview, three students said they would rather use corpus software with others in the future, while two said they would prefer to study alone. Student E believed "it's more fun to share ideas and I think it's easier to understand," while Student

C stated that "I can share opinions with my friends and we can tell each other the answers … I think in this kind of lesson I can build deeper relationships with friends … I think this kind of exchange is important." Student B claimed that he would rather work by himself because "when you're by yourself you can think alone about what you want to search for" and Student D said "I can focus more when I study by myself."

These responses indicate that individual learner study preferences have an influence in determining student impressions of DDL activities in a classroom setting. The teacher also seemingly has a key role to play in the effective implementation of DDL activities in the young learner classroom and the data suggest that appropriate L1 use by the teacher is necessary to help 'scaffold' the activities for the students. Peer support is also valued by some students who enjoy checking answers with other classmates and this appears to increase their motivation to learn.

The results presented above show promising signs that both computer-based and paper-based DDL activities can be positively received by young learners of EFL. However, there are clearly various issues that must be addressed in order to ensure the successful implementation of DDL with elementary school learners and these will be discussed below.

Discussion

This section considers the implications of the present study's findings as well as the key issues that can impact the implementation of DDL activities in the young learner EFL classroom, with a particular focus on Japanese EFL learners. Firstly, we shall take a look at young learner attitudes toward DDL and then discuss technology use, the availability of suitable corpora, and the influence of 'scaffolding' in young learner pedagogy.

Young learner attitudes toward DDL

The present study's findings on the perceptions and attitudes of young learners toward DDL study are consistent with those of previous studies. The paper-based activities were generally very well received by participants, as were similar activities undertaken by young learners in Kim (2019), MacGregor (2014), and Takahashi and Fujiwara (2016). While research data are sparse, there are encouraging signs that a DDL-informed approach can motivate elementary-school-age children to learn English in an L2 classroom environment.

Previous studies on young learners and DDL are informative, yet they focus exclusively on paper-based activities or corpus-informed games and fail to investigate direct corpus use through computer-based activities. This study has found tentative evidence of positive attitudes toward learning English through computer-based DDL. The results should be interpreted with caution as only two sessions were observed in which corpus software was prominently used. However, the data do suggest that, with the right computer (and corpus) training, as well as materials and tasks created at an appropriate level of difficulty,

elementary school sixth-grade students can have positive attitudes toward working directly with corpus software to study English. An interesting observation can be made in the responses that show that all five participants expressed a preference for the two paper-based activities over the two computer-based activities. Some post-course interview answers hinted at the fact that a lack of familiarity with the software was an issue. It may be that, when provided with a choice, this particular age group may prefer paper-based DDL regardless of their level of corpus experience, but further investigation is required. The following two sections will consider the particular impediments to the implementation of computer-based DDL with elementary school learners.

Young learners and technology use

Technology education in Japanese elementary schools is continuously improving, with each student now having access to a tablet in the classroom (MEXT, 2021) and computer programming classes for fifth and sixth-grade students since April 2020 (Gougeon & Cross, 2021). Despite this, the findings of the present study raised concerns that Japanese twelve-year-olds appear to have limited experience using personal computers. Prior to the study, only Student E stated that she had experience using a computer and, anecdotally, there were problems in the first session with some students initially struggling with basic functions and needing assistance from their parents or the teacher. This suggests that, despite home access to a computer, actual usage is rather limited.

An OECD report in 2015 claimed that 25% of Japanese people between the ages of 16 and 29 do not have basic computer skills (Cote & Milliner, 2017). Prior to the introduction of direct corpus-based activities in EFL classes, elementary school students may require a period of training in practical computer skills. One might predict that, if the participants in the current study had had a greater degree of familiarity with computers at the start of the course, they may have found the computer-based activities easier to complete. While the enjoyment scores were still fairly high for both classes involving direct corpus use, it can be speculated that they may have stated a preference for these types of activities if their computer skills had been more advanced. Another alternative is to limit direct interaction with corpora or corpus-based activities to software or applications available on tablet computers or smartphones, as young learners tend to have more experience with these devices, although it can restrict the number of corpora or software available.

On the subject of direct corpus use, this leads to another potential concern for educators or parents of young learners and that is the availability of (and access to) suitable corpora.

Appropriate corpora for young learners

When using corpora with young learners, it is important to consider the type of corpus used.

Many corpora are simply not suitable for direct use by young learners due

to their content. General corpora such as the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA) contain files from a range of genres, including novels, academic journals, and magazines, and may include inappropriate language or feature adult topics, and educators would be advised to avoid exposing pre-teen learners to such texts. For the computer-based activities in the present study, the CHILDES English Corpus was used via the *Sketch Engine* online software. The website interface is relatively user-friendly compared with other corpus software that is available and so it was considered to be the most suitable option for elementary school students. Other corpora appropriate for children include the Oxford Children's Corpus and the CLLIP corpus, a BNC sub-corpus made up of 30 child fiction texts (600,000 words), as used in Sealey and Thompson (2006) as well as Kim's (2019) research study. All three corpora are not freely available, however, and require subscriptions to software and, in some cases, special permission for access.

In summary, for the EFL teacher looking to implement computer-based DDL in their classroom, careful consideration must be given to which corpora (and software) are most suitable for their young learners and which will allow for the creation of effective and motivating corpus activities.

'Scaffolding' in the young learner classroom

This section will examine DDL from a socio-cultural perspective and consider the use of L1 in classroom instruction as well as teacher and peer mediation.

L1 use in the classroom. In the post-study interview, participants were asked about the use of Japanese by the teacher during the class, with all claiming that they found it helpful to some extent. Student D explicitly stated that she would have had difficulties if the lesson had been exclusively in English.

There is some research on the use of L1 in the EFL classroom, although to date little has focused on young learners of elementary school age (Copland & Yonetsugi, 2016). For the teacher, communicating in the students' L1 can be a useful time-saving strategy (Harbord, 1992) or used as a tool for controlling classroom behavior (Macaro, 2001). Also, in the young learner EFL classroom, there are learning benefits to allowing students to use their first language during instruction (Shintani, 2014; Tognini & Oliver, 2012).

The efficacy of teachers and students speaking in the L1 during the lesson can also be applied more specifically to young learners undertaking DDL instruction. When introducing DDL activities to elementary school children, it is possible that a significant amount of L1 will be required in the initial stages to ensure that tasks can be completed successfully. In Kim's (2019) study, the researcher wrote the materials in both English and Hangul (the participants' L1) for ease of understanding. Caution is advised, though, as it may encourage students to primarily focus on their written L1, thus potentially making the English text redundant.

Teacher mediation. The findings from the present study, while limited, do

suggest that the teacher's role and the classroom dynamics play an integral part in the perceptions of DDL instruction by students in this age group. As a reminder, the researcher and participants in this study did not previously know each other. This would naturally have some impact on the relationship between the teacher and the students as they needed to build a rapport in a short space of time and without being in the same physical space. The high enjoyment ratings given by the study participants suggest that this was not much of an impediment to learning or motivation to study. In the post-class questionnaires, students were asked directly about the teacher's support and generally showed a high level of satisfaction.

The teacher clearly plays a role in how the students perceive their classroom instruction and this is likely to have an impact on young learner attitudes toward the study of foreign languages. In Takahashi and Fujiwara's (2016) study on Japanese elementary school sixth-graders using DDL activities, they discovered that there were students who appreciated the hints provided by the teacher when they were having difficulties with comprehension. Guidance from the teacher can reduce students' difficulties with the activities and Kim (2019) found that those students who referred to guidance in the post-study interviews often also had positive opinions on DDL grammar study.

Peer Mediation. The present study sought to gain a deeper understanding of how the participants viewed the social element of undertaking DDL within the classroom context and whether this had any impact on their attitudes toward the activities. Several responses hint at the value that young learners place on peer–peer interaction in the EFL classroom. A friendly, co-operative atmosphere that fosters support between students in a young learner foreign language classroom can be a key factor in the successful integration of DDL activities.

Takahashi and Fujiwara (2016) found that some participants liked learning as a group, with one student commenting on checking answers with their peers. The twelve-year-old participants in Kim's (2019) study also spoke of the pleasure they derived from learning together and sharing ideas with other members of the class. One student observed that they found they were able to memorize the language rule as a result of peer discussions, while two other students talked about peers assisting each other with answering the questions. In MacGregor's (2014) study of learners aged 9–15, some participants claimed to enjoy doing the various DDL activities because of the opportunity to co-operate with peers and engage in teamwork. There is some evidence that young learners value working with their peers in the classroom and, while more research is required to look at the extent of motivational effects on learners, it appears that regular peer co-operation and collaboration does have a positive impact on the perception of DDL activities by young learners.

Study limitations

While this study produced insightful findings, some limitations need to be addressed. Firstly, the small sample size of five participants means that the data cannot be extrapolated to a wider population and generalizations should be avoided. Secondly, the course of one-hour lessons would have been more informative had it been extended to multiple sessions over a period of several months. Additionally, in seeking to gain insight into the perceptions and attitudes of participants of upper elementary school age, we must accept that there can be difficulties in eliciting in-depth and thoughtful responses as children may struggle to articulate their true thoughts. There is also the possibility that the students seek to please the teacher/researcher by providing inflated favorable opinions, although there is no way of proving this. Finally, a pilot study conducted prior to the main study would have aided in developing suitable materials and provided a chance to test the corpus software (as well as the questionnaire and interview questions), but time limitations made this unfeasible.

Future research possibilities

To provide richer data, the present study could be expanded into a longitudinal project conducted with a larger group of students over a longer period of time. The setting could be in a physical classroom and possibly with a regular Japanese elementary school class. Data triangulation could involve analysis of teacher diary entries or teacher observations of each lesson. This could allow the researcher to understand more clearly the reasons behind activity ratings or why a participant made a particular claim in the post-class questionnaires or post-study interviews.

Additionally, while there is some evidence that young learners generally have positive attitudes toward corpus-based classroom activities, more research is required into the actual learning effects gained from DDL instruction. Large-scale, longitudinal studies that investigate learning effects would be beneficial for advancing our understanding of the impact DDL can have on learners and their language development, as well as research that focuses on SLA theories within the approach (Flowerdew, 2015; O'Keeffe, 2021b). This particularly applies to young learners of EFL, who have arguably long been neglected as research subjects in foreign language learning research.

Conclusion

This study investigated the perceptions and attitudes of Japanese sixth-grade elementary school students toward learning English verb–noun collocations using DDL activities in an EFL class environment. It was found that the participants mostly enjoyed the lessons and showed eagerness toward studying English through such activities in the future. There was a preference for the paper-based activities over the computer-based activities from all five students; however, positive feedback on both forms of DDL do suggest that more

computer (and corpus) training could improve students' perceptions of activities involving direct corpus use. Questionnaire and interview responses regarding classroom interaction and relationships highlight the importance of teacher and peer mediation in young-learner EFL instruction. Finally, the evidence presented here, while tentative, suggests that elementary school learners are more likely to willingly participate in corpus-based study that sits on the right side of O'Keefe's (2021a) DDL framework. In a socio-culturally focused form of DDL, the students can complete a variety of both paper-based and computer-based activities, working together with the teacher and their peers to study familiar language included in the curriculum. A sufficient amount of 'scaffolding' would hopefully build learner confidence in the direct use of corpus software, and this could result in students engaging in a more constructivist-focused, autonomous type of DDL study in future years.

References

Allen-Tamai, M., Ano, K., & Hamanaka, N. (Eds.). (2020a). *New horizon elementary 5.* Tokyo: Tokyo Shoseki.

Allen-Tamai, M., Ano, K., & Hamanaka, N. (Eds.). (2020b). *New horizon elementary 6.* Tokyo: Tokyo Shoseki.

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261–292). John Benjamins. https://doi.org/10.1075/scl.52.11bou

Boulton, A. (2021). Research in data-driven learning. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 9–34). John Benjamins. https://doi.org/10.1075/scl.102

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*(3), 307–328. https://doi.org/10.1017/S0958344007000535

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

Copland, F., & Yonetsugi, E. (2016). Teaching English to young learners: Supporting the case for the bilingual native English speaker teacher. *Classroom Discourse, 7*(3), 221–238. https://doi.org/10.1080/19463014.2016.1192050

Cote, T. J., & Milliner, B. (2017). Preparing Japanese students' digital literacy for study abroad: Is more training needed? *The JALT CALL Journal, 13*(3), 187–197. https://doi.org/10.1111/lang.12224

Crosthwaite, P. (2020). Data-driven learning and younger learners: Introduction to the volume. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 1–10). Routledge.

Crosthwaite, P., & Stell, A. (2020). "It helps me get ideas on how to use my words": Elementary school students' initial reactions to corpus use in a private tutoring setting. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 150–170). Routledge.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Oxford University Press.

Eder, D., & Fingerson, L. (2003). Interviewing children and adolescents. In J. Holstein & J. F. Gubrium (Eds.), *Inside interviewing: New lenses, new concerns* (pp. 33–53). SAGE.

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.69

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL, 26*(2), 128–146. https://doi.org/10.1017/S0958344014000093

Gougeon, L., & Cross, J. S. (2021). Japanese elementary schools' playful programming curriculum considerations: Readiness, limitations and teacher training. *2021 IEEE International Conference on Engineering, Technology & Education (TALE),* 23–28. https://doi.org/10.1109/TALE52509.2021.9678771

Hadley, G., & Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Language Learning & Technology, 21*(3), 131–152. http://hdl.handle.net/10125/44624

Harbord, J. (1992). The use of the mother tongue in the classroom. *ELT Journal, 46*(4), 350–355. https://doi.org/10.1093/elt/46.4.350

Heary, C., & Hennessy, E. (2006). Focus groups versus individual interviews with children: A comparison of data. *The Irish Journal of Psychology, 27*(1), 58–68. https://doi.org/10.1080/03033910.2006.10446228

Kim, H. (2019). The perception of teachers and learners towards an exploratory corpus-based grammar instruction in a Korean EFL elementary school context. *Elementary English Education, 25*(1), 123–152. http://dx.doi.org/10.25231/pee.2019.25.1.123

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

Loewen, S., & Reinders, H. (2011). *Key concepts in second language acquisition.* Palgrave Macmillan.

Macaro, E. (2001). Analysing student teachers' codeswitching in foreign language classrooms: Theories and decision-making. *The Modern Language Journal, 85*(4), 531–548. https://doi.org/10.1111/0026-7902.00124

MacGregor, A. (2014). Young learners and lexical awareness: Children's engagement with wordlists and concordances. *TESOL Journal, 5*(1), 120–149. https://doi.org/10.1002/tesj.102

MEXT. (2021). *GIGA sukuru koso-no jitsugen-ni tsuite* [The realization of the GIGA school concept]. Retrieved September 12, 2023, from https://www.mext.go.jp/a_menu/other/index_00001.htm.

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies, 22,* 1–18.

Nishigaki, C. (2022). *Eigo no ru-ru wo hakken shiyo!* DDL Study. Retrieved September 12, 2023, from https://e.ddl-study.org/

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

O'Keeffe, A. (2021b). Data-driven learning, theories of learning and second language acquisition: In search of intersections. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: corpora in language education* (pp. 35–55). John Benjamins.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pinter, A. (2015). Researching young learners. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp. 439–455). Bloomsbury.

Rose, H., McKinley, J., & Briggs Baffoe-Djan, J. (2019). *Data collection research methods in applied linguistics.* Bloomsbury.

Sealey, A., & Thompson, P. (2004). 'What do you call the dull words?' Elementary school children using corpus-based approaches to learn about language. *English in Education, 38*(1), 80–91. https://doi.org/10.1111/j.1754-8845.2004.tb00774.x

Sealey, A., & Thompson, P. (2006). "Nice things get said": corpus evidence and the national literacy strategy. *Literacy, 40*(1), 22–28. https://doi.org/10.1111/j.1467-9345.2006.415.x

Sealey, A., & Thompson, P. (2007). Corpus, concordance, classification: Young learners in the L1 classroom. *Language Awareness, 16*(3), 208–223. https://doi.2org/10.2167/la453.0

Shintani, N. (2014). Using tasks with young beginner learners: The role of the teacher. *Innovation in Language Learning and Teaching, 8*(3), 279–294. https://doi.org/10.1080/17501229.2013.861466

Sketch Engine. (2023). Retrieved September 12, 2023, from https://auth.sketchengine.eu/#login

Sonix, Inc. (2023). Sonix. Retrieved September 12, 2023, from https://my.sonix.ai/accounts/sign_in

Szudarski, P. (2020). Effects of data-driven learning on enhancing the phraseological knowledge of secondary school learners of L2 English. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 133–149). Routledge.

Takahashi, S., & Fujiwara, Y. (2016). Effects of inductive learning based on data-driven learning at elementary schools in Japan. *JES Journal, 16*(1), 84–99. https://doi.org/10.20597/jesjournal.16.01_84

Tognini, R., & Oliver, R. (2012). L1 use in elementary and secondary foreign language classrooms and its contribution to learning. In E. A. Soler & M-P. Safont-Jordà (Eds.), *Discourse and language learning across L2 instructional settings* (pp. 53–77). Brill. https://doi.org/10.1163/9789401208598_005

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Weisser, M. (2016). *Practical corpus linguistics: An introduction to corpus-based language analysis.* Wiley Blackwell.

Appendix A

Pre-study survey

Q1   What's your name?
Q2   How old are you?
Q3   What is your gender?
Q4   What age did you start learning English?
Q5   Where have you studied English?
Q6   How long do you study English for every week?
Q7   Have you ever taken the EIKEN test?
Q8   Which technological devices have you used before?
Q9   Have you ever used a technological device to learn English?
Q10   How do you like to study English?
Q11   How do you rate your elementary school English classes?
Q12   What do you think about your elementary school English classes? What do you like about your classes? What do you dislike? Please write down your thoughts.
Q13   How have you studied English grammar before?

Appendix B

Post-class questionnaire

Q1   How was today's class? (Likert scale of 1–7)
Q2   How difficult was today's class? (1–7)
Q3   Please say what you enjoyed about today's class. (open-ended)

Q4   Please say if there is anything you found boring about today's class. (open-ended)
Q5   What was good about today's activity? (open-ended)
Q6   What was difficult about today's activity? (open-ended)
Q7   Would you like to do today's activity again? (multiple-choice)
Q8   Please give your reasons for your answer to Q7. (open-ended)
Q9   How would you like to do this activity again? (multiple-choice)
Q10 How would you rate the teacher's support today? (1–7)
Q11 Please state what was good about the teacher's support or what you think needs changing. (open-ended)
Q12 Were there any times when you worked with your classmates? If so, please write it down. (open-ended)

Appendix C

Post-study interview

Q1   What was the most memorable aspect of the course?
Q2   During the lessons the teacher sometimes used Japanese. Did you find this helpful? Do you think it would have been better to use no Japanese?
Q3   What difficulties did you have in the lessons and how were these solved?
Q4   How would you rank the four different activities?
     (1= most favorite, 4= least favorite)
Q5   Why did you choose Activity __/__ as your most/least favorite activity?
Q6   Which did you prefer, the computer-based or paper-based activities? Why?
Q7   Which website did you find easier to use, Sketch Engine or ddl-study? Why?
Q8   Do you think that you will use corpus software after this course to study English? Why? Why not?
Q9   If you would like to use corpus software again, would you prefer to work by yourself or with others? Why?