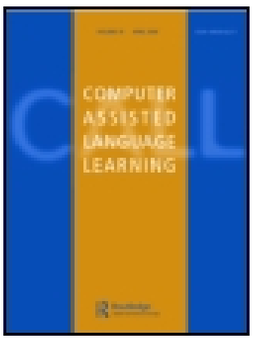

Computer Assisted Language Learning

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ncal20

A mixed methods case study on the use and impact of web-based lexicographic tools on L2 writing

Gabriela Adela Gánem-Gutiérrez & Alexander Gilmore

To cite this article: Gabriela Adela Gánem-Gutiérrez & Alexander Gilmore (2021): A mixed methods case study on the use and impact of web-based lexicographic tools on L2 writing, Computer Assisted Language Learning, DOI: 10.1080/09588221.2021.1987273

To link to this article: https://doi.org/10.1080/09588221.2021.1987273

COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2021.1987273

A mixed methods case study on the use and impact of web-based lexicographic tools on L2 writing

Gabriela Adela Gánem-Gutiérrez[a] and Alexander Gilmore[b]

[a]Centre for Research in Language Development throughout the Lifespan (LaDeLi), Department of Language and Linguistics, University of Essex, Colchester, UK; [b]Department of Civil Engineering, University of Tokyo, Tokyo, Japan

ABSTRACT

This study used a mixed methods research design to investigate use of Web-based lexicographic tools during real-time second language (L2) writing activity in an English as a foreign language university context and the impact of their use on writing quality. Participants included 22 adult Japanese EFL learners, elementary to advanced language proficiency, and their tutor. Data were collected using digital screen capture and eye-tracking technologies while participants wrote a 35-minute essay. Supplementary stimulated retrospective recalls were also conducted to document participants' reflections on the use of these tools during the full writing event. Results revealed that participants spent, on average, 15% of composing time consulting online tools, with bilingual resources the most used. Importantly, mediation analysis found that L2 proficiency was the main, and direct, predictor of essay score. These findings as well as specific patterns of use discerned are discussed with reference to the qualitative data before proposals for further research and pedagogical implications are considered.

KEYWORDS

Web-based lexicographic tools;
L2 writing;
writing performance;
digital screen-capture;
eye-tracking;
mixed-methods

Introduction

Web-based lexicographic tools such as online monolingual and bilingual dictionaries, thesauruses, concordances,[1] and web search engines, e.g. Google, have become widely used by second language (L2) learners, particularly for academic writing (Lew, 2016; Lew & Szarowska, 2017; C. Yoon, 2016a, 2016b). They have been shown to constitute an

CONTACT Gabriela Adela Gánem-Gutiérrez aganem@essex.ac.uk Centre for Research in Language Development throughout the Lifespan (LaDeLi), Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

Supplemental data for this article can be accessed here.

important mediational tool underlying L2 writing processes such as formulation and revision (Gánem-Gutiérrez & Gilmore, 2018). Their potential, as well as some of their limitations, have been widely acknowledged (Frankenberg-Garcia, 2020; C. Yoon, 2016b). However, it remains uncertain how these tools are being used by L2 writers who have received no specific training. Crucially, we still lack knowledge regarding the actual impact of these tools on writing quality and how their use interacts with other key variables such as L2 proficiency (Lai & Chen, 2015; Tono, Satake, & Miura, 2014).

This study addresses the above gaps in our knowledge through a mixed-methods design, underpinned by real-time digital screen-capture (DSC) and eye-tracking technologies. More specifically, our chosen methodology contributes to language teaching research in the following ways: First, we use statistical analyses, including mediation analysis, a technique rarely seen in language research articles to investigate profiles of use of Web-based lexicographic tools across a range of L2 proficiencies. Crucially, we investigate the potential impact that using those resources might have on the actual quality of the written product, taking language proficiency into account. Second, we use DSC video with gaze data replay to facilitate and enhance stimulated retrospective recalls and thus demonstrate how an eye-tracker can support research and pedagogy in innovative ways. Finally, we aim to contribute to strengthening pedagogical practice by providing insights into L2 writers' thinking and rationale for use of these tools. This is explored in the discussion by including detailed examples from stimulated retrospective recalls.

Background

There has been considerable research on L2 learners' use of web-based resources for correcting errors (Gilmore, 2009; Chon, 2009; Conroy, 2010; Fujii, 2007; Mraček, 2020; Tono et al., 2014; H. Yoon, 2008), look-up strategies and reasons for accessing the resources (Kennedy & Miceli, 2010, 2017; H. Yoon, 2008), as well as assessment of search success and students' attitudes towards the tools (Conroy, 2010; Quinn, 2015; Varley, 2009; H. Yoon, 2008; H. Yoon & Hirvela, 2004). Online dictionaries and thesauruses have been praised for facilitating language searches and for offering multimodal information through hyperlinks (Tono, 2011), despite the potential for distraction due to advertisements in free online versions (Dziemianko, 2019). Their use to support L2 writing while composing is, however, seriously under-studied. Furthermore, there is a tendency to focus on a single specific type of tool, with concordancers dominating, while online dictionaries and web search engines have received much less attention.

The study of single resources has contributed to our understanding of both the patterns of use and potential value of such tools; nonetheless, study of a range of tools used simultaneously is necessary. Combined tool affordances have the potential to improve naturalness in writing by showing authentic strings of collocations and contextualised language within specific genres and registers (Conroy, 2010; Fujii, 2007; Gilmore, 2009; Kennedy & Miceli, 2010). It is thus now necessary to study contexts where L2 writers have access to multiple resources and use them spontaneously throughout a writing task – an approach which is more authentic so carries greater ecological validity (Levy & Steel, 2015; Steel & Levy, 2013). To date, such studies are rare with three standing out: C. Yoon (2016a, 2016b); Lai and Chen (2015); and Frankenberg-Garcia (2005). These together provide an informative account of the extent to which lexicographic tools were used by participants with access to multiple resources, and the level of success achieved from the searches.

C. Yoon (2016a, 2016b) investigated use of online resources by six advanced ESL Korean graduate writers while completing an authentic writing assignment over several weeks. Students had access to eight different online reference resources including concordance tools, web search engines such as Google, and online bi/monolingual dictionaries. Data comprised three hours of screen recordings, stimulated recall sessions where participants discussed intentions behind their search queries (N = 515), and an interview and survey where advantages and disadvantages of using the tools were assessed. Lai and Chen (2015) design also included stimulated recall based on sample video clips of the writing process. Their participants (N = 14) were intermediate proficiency Taiwanese college students with access to bilingual and monolingual reference tools, including corpora and dictionaries (1,039 look-ups). In contrast, Frankenberg-Garcia (2005) relied on participants' self-completed answer grids describing what they had accessed while completing a translation task. The sixteen advanced English level Portuguese undergraduate students reported 146 queries using a wide range of lexicographic resources, including paper-based dictionaries and specialised online tools such as Eurodicautom, the online multilingual term bank of the European Union.

The results of these studies suggest that online bilingual resources were the most popular for consultation: 57% of queries in Frankenberg-Garcia's research (although this included paper-based dictionaries). Lai and Chen reported 61.6% bilingual dictionary queries and 32.4% using a bilingual concordancer; while C. Yoon's (2016a) participants used bilingual resources in 25% of queries. There were considerable differences in the use of corpus-based tools. The Korean and Taiwanese participants consulted concordancers for a similar

percentage of queries, 36% and 38.4% respectively, while the Portuguese students used this for only 11% of searches. Findings also differed regarding search engines, with Yoon reporting around 18% use but Frankenberg-Garcia only 6%. In terms of usefulness of the resources, Yoon reports that 69.4% of consultations resulted in a correct text formulation or revision, 12.7% were incorrect and 18% were abandoned because of the difficulty in 'finding relevant items from the query results' (C. Yoon, 2016a, p. 219). Reasonable levels of success were observed in Frankenberg-Garcia's consultations: 58% when using Eurodicautom, 69% when using corpora, and 67% in the case of search engines. Lai and Chen (2015) did not consider usefulness of the resources.

In all the studies individual preferences are prevalent, but some group patterns were nonetheless observed. For instance, students in Frankenberg-Garcia (2005) and Lai and Chen (2015) favoured use of a single resource to solve a particular problem, e.g. dictionaries to find the meaning of a word, and corpora for collocation, rather than consulting different resources in combination as with Yoon's participants.

Overall, these studies conclude that the resources have the potential to effectively support L2 writers and one thing is certain – these tools are increasingly pervasive in the L2 context so we need to better understand their potential. This study is therefore designed to contribute to that understanding and address a number of gaps in the literature. First, in contrast to most existing research, this study investigated access to a range of Web-based resources since we know that increasing Internet connectivity has resulted in L2 learners' simultaneous use of multiple tools. Second, research in this area has almost exclusively targeted advanced learners (Mueller & Jacobsen, 2016) thus we included a wide range of L2 proficiencies, from elementary to advanced. Third, recent cognitive models of writing (Leijten, Van Waes, Schriver, & Hayes, 2014) include *searching* as a fundamental part of the process, recognising that contemporary writers typically make use of online resources to access information while composing. Nonetheless, in L2 writing, we know relatively little about the extent or proportion of composing time spent accessing such resources, but see Gánem-Gutiérrez and Gilmore (2018); Stapleton (2010); and Hvelplund (2017). Crucially fourth, to the best of our knowledge, no publication to date reports on how use of online lexicographic tools relates to L2 proficiency and writing performance for a composition as a whole; this is echoed by Tono et al. (2014, p. 152) who state that 'there is no empirical evidence that unsupervised access to corpus data by students with varying degrees of proficiency ensures accuracy in revising their L2 writing.' It, therefore, remains difficult to ascertain what the actual impact of using these resources might be on writing quality (considering time, length and accuracy).

Our study uniquely investigates this by studying both the role of online resources in the real-time writing process and the overall quality of the writing products, in relation to writer proficiency.

Finally, and as highlighted in the introduction, we believe that a mixed-methods approach, underpinned by technologies such as digital screen-capture and eye-tracking, has the power to render more robust and accurate accounts of lexicographic tool use in L2 writing than the previous predominantly qualitative accounts. This, we argue, is a necessary methodological stepping-stone that will enable meaningful comparisons with future studies and, ultimately, much needed generalizable conclusions. The present study therefore addressed the following research questions:

RQ1: What distinct profiles of online lexicographic resource use can be observed in the participants as they write?

RQ2: Which factors, L2 language proficiency and/or type of resource used, independently predict essay length, quality, and time spent on the task?

RQ3: Do L2 proficiency and use of lexicographic resources affect essay writing directly, or are they mediated through other variables?

Method

Participants

This case study included 22 EFL students (6 males; 16 females) from two Japanese universities and an EFL tutor/researcher. An English native-speaker participant (P01) was also included to provide some benchmark figures, although his data was excluded from the statistical analysis. Participants' L1 was Japanese, except for three, for whom Japanese was an L2 and whose L1s were Mandarin Chinese, Korean, and Spanish. Ages ranged from 18 to 40 years (mean = 21.4; median = 20; $SD = 5.58$), with length of time learning English from 6 to 14 years and self-reported writing expertise assessed as elementary to advanced. Participants' L2 English proficiency levels ranged from elementary to advanced (see below and Table 1 in Appendix, Supplementary material for details).

Data gathering tools and procedures for data collection

All data (obtained with full written consent) was collected by the tutor/researcher on an individual basis during three stages:

Stage 1: precomposition stage

Participants completed a 116 item C-test in order to estimate their relative English proficiency (Gilmore, 2011; Grotjahn, 2010).

Stage 2: composition stage

Participants were first familiarized with the hardware and software, which included an eye-tracker, word processor and the online lexicographic resources available to them while writing. These resources were selected based on their reported popularity and use by university students in Japan (Gilmore's unpublished survey data):

a. Eijiro ALC, a bilingual concordancing tool which allows users to search for keywords or phrases in English or Japanese. It is free and well used (with a Google PageRank of 6/10) although more accurate in the English > Japanese direction; the J > E version is simply reverse engineered without any human editing and includes some strange or misleading translations. In addition, Japanese words which would not be used as translations for English words, particularly those associated with Japanese culture (e.g. 'shuukatsu': to make arrangements for one's own death), are omitted completely.
b. The J-E dictionary at Goo, the free online version of the Progressive Japanese-English Dictionary, published by Shogakukan. Lexicographers such as Tom Gally (personal communication) consider it a good, mid-sized dictionary written specifically with the needs of Japanese-L1 users in mind, and the English translations are, overall, accurate and natural.
c. The Oxford Advanced Learner's Dictionary (OALD), a free online version of the world's best-selling monolingual dictionary aimed at a non-native audience (Google PageRank of 6/10). It provides definitions, grammatical information, example sentences, collocations, synonyms and idioms associated with a search word.
d. The Merriam Webster online dictionary and thesaurus also a well-known American English monolingual resource for native speakers or learners (Google PageRank of 8/10). Similarly to the OALD, the dictionary provides definitions, grammatical information, example sentences, collocations for search words, while the thesaurus function lists (near) synonyms and (near) antonyms.

Participants were also free to use any other online resources of their choosing via the web browser, Firefox, where all the above resources were opened for convenience. It is important to note that, in the interest of realism, students were not specifically trained in the use of these tools nor were they told they had to use them. Furthermore, we were interested in documenting the full composing process (Gánem-Gutiérrez & Gilmore, 2018) where 'the interaction with multiple sources—intentionally and unintentionally—has become an inherent part' of contemporary L2 writing practices (Leijten et al., 2014; Leijten & van Waes, 2013, p. 383).

Students had 10 minutes planning time (not included for analysis in this study) before writing an IELTS style argumentative essay on the topic: 'Education should be free for everyone. To what extent do you agree or disagree with this statement?' An argumentative essay was chosen given its potential for knowledge-transforming and problem-solving demands (Roca de Larios, Manchón, Murphy, & Marín, 2008); the essay topic was chosen for its familiarity and engagement potential. To enable comparison between participants and in line with IELTS practice tasks, each participant was given 35 minutes to write the essay.

Real-time digital screen capture (DSC) and eye-tracking data. The eye-tracker suite, Tobii T60/Studio 2.2,[2] was calibrated and used to gather data comprising of visual records (DSC) of the whole onscreen L2 writing event (13 hrs in total) with eye gaze data overlaid. This technology provides an excellent way of unobtrusively tracking moment-to-moment activity including the use of online resources (Latif, 2008, 2019) and renders a powerful visualisation of L2 writing processes by capturing on-screen activity, recorded sound, keystrokes and mouse clicks and integrating eye gaze data. The latter was exclusively used to support the stimulated retrospective recall (SRR) and thus represents an innovative use of this technology.

Digital video recordings of the participants' interaction with the computer and their paper notes were also collected in order to capture all possible activity during composing time (see Figure 1 for an illustration of all the elements captured for subsequent analyses).

Stage 3: stimulated retrospective recall (SRR) stage

In order to support interpretation of students' use of lexicographic resources and L2 writing behaviour, as well as to provide data for complementary qualitative analysis, we conducted SRRs. Following general guidelines on stimulated recall methodology (Gass & Mackey, 2000), the SRR protocol (based on the DSC video with gaze data replay, see Figure 1) was initiated after a 10-minute break – while the writing event was still fresh in participants' minds – using the following instructions:

> We will now watch your composition video, and I would like you to talk me through what was going on in your mind as you were writing your essay. You can press the pause button whenever you want to make a comment, and if I pause, I would also like you to tell me what you were thinking at the time.

The SRR was also recorded using Tobii Studio 2.2 and produced a dataset of 27 hrs 6 mins, which was transcribed in full.

In sum, the research design responds to calls to make use of more accurate and robust methodologies in L2 writing studies (Gilmore &

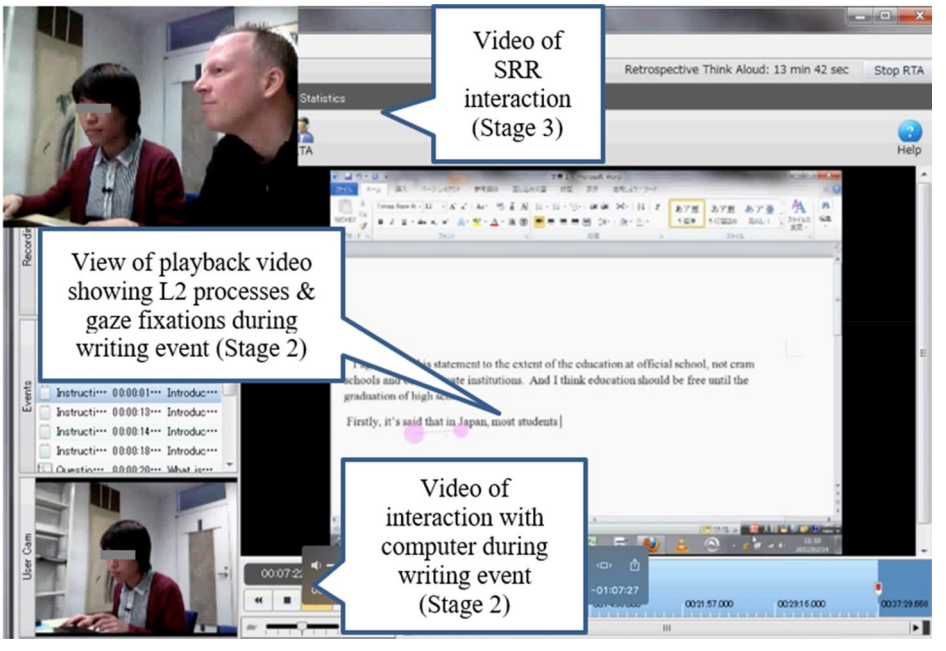

Figure 1. Overview of visual data available for analysis.

Gánem-Gutiérrez, 2020; Hamel & Séror, 2016; Séror, 2013) given the fact that such designs are still scarce (but see Gánem-Gutiérrez & Gilmore, 2018; Park & Kinginger, 2010; Révész, Michel, & Lee, 2019).

Data analysis procedures

C-test and essay

The C-test was scored based on the exact word scoring method (Weir, 1990) according one point to each correct answer (maximum score = 116 points). The essays were blind-rated by three native speaker teachers with training and experience in language testing. Using the IELTS Task 2 writing band descriptors and scoring procedures,[3] each essay was rated on four dimensions (task achievement, coherence and cohesion, lexical resources, grammatical range and accuracy), which were then averaged to render a global score for each composition. The result for each participant was the mean of the three raters' global scores (see Appendix, Supplementary material). Interrater reliability was excellent (Cronbach's α = .971). C-test scores and essay scores correlated strongly, $r = .75$, $p < .001$.

Digital screen capture data

In preparation for analysis, the DSC videos of real-time L2 writing behaviour were segmented into episodes and coded for L2 processes (text construction, revising, rereading, use of external resources, pausing

or other) using Elan tools v.4.8.1. (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). In this report, our focus is exclusively on episodes involving the use of lexicographic resources during the writing event – for definitions and an analysis of the other processes, refer to Gánem-Gutiérrez & Gilmore, 2018. A use of external resources episode was defined as a 'Period when students left the word processor to access external resources, for example, monolingual dictionary, bilingual dictionary, thesaurus, Web browser, or paper notes, as evident through screen capture data, video-recorded data, and retrospective stimulated recall protocols' (Gánem-Gutiérrez & Gilmore, 2018, p. 482). Once all files are segmented and coded, Elan produces descriptive statistics, e.g. frequency (number) and temporal duration (in seconds) of episodes. These data formed the basis for subsequent statistical analyses which were conducted using IBM SPSS v.23. The simultaneous procedure of DSC video segmentation into episodes and their coding followed a recursive process between the authors (Bernard & Ryan, 2010). After establishing inter-coder reliability based on 10% of the data (κ = .83) all video files were segmented and coded by one of the authors. Intra-coder reliability was subsequently checked on a random 5% of the data (κ = .93). Finally, and once all data had been coded, a second cycle of inter-coder reliability based on a random 10% of the data was conducted (κ = .86).

Stimulated retrospective recall (SRR) data

SRR data provided supporting information on the participants' thought processes during the essay writing task. Relevant episodes, where use of external resources was discussed, were identified from the written transcriptions and collated for analysis.

Statistical analysis

Researchers widely believe that L2 writing performance is influenced by a wide range of factors (Cumming, 1990; C. Yoon, 2016b). These fall broadly into two categories: abilities of the writer, and the nature of the task. In the present study, we controlled variables relating to the writing task: topic, maximum time allowed, and availability of external resources. The focus was on two key ways in which the participants varied: L2 language proficiency, and pattern of processing with respect to use of external resources. The design is, therefore, basically correlational. The analysis includes measures of writing performance (product length and quality, and time spent on the task) as dependent variables. The independent variables are L2 language proficiency (C-test), and measures of use of three types of online lexicographic tools (bilingual resources, monolingual dictionaries, and thesauruses).

The basic measures of each external resource were the number of times it was consulted (frequency) and the length of time it was used for (duration). We further generated derived scores for: (i) all resources considered together; (ii) range/repertoire of resources used by each person (out of three); (iii) mean length of time per consultation for each resource, and (iv) duration of use of each resource as a percent of total time spent on the task.

Only seven of the 21 variables passed the Kolmogorov-Smirnov test (with Lilliefors correction) of distribution normality. For pairwise correlation we therefore used non-parametric Spearman correlation. For multiple regression analyses we used optimal scaling, a non-parametric counterpart of normal multiple regression. For mediation analysis, a technique rarely seen in language research articles, we used the suggestions of Shrout and Bolger (2002), which include a (non-parametric) bootstrapping procedure to compute a confidence interval around the indirect effect (i.e. the path from the independent variable to the dependent through the mediator). If zero falls outside of this interval, significant mediation is said to be present. Specifically, we used the SPSS macro designed by Preacher and Hayes (2008) to perform all the required mediation analyses, both the parts using customary (parametric) multiple regression and those using bootstrapping.

Results

Upon scrutinising the descriptive statistics for each variable (Table 3 in Appendix, Supplementary material), it became immediately apparent that, regardless of which measure is used, use of bilingual resources (mean around 14 uses per essay) far exceeds that of the monolingual dictionaries (mean around 1) or thesauruses (mean .5). Hence, the figures shown in Table 3 Appendix, Supplementary material for overall resource use are heavily influenced by the use of bilingual resources. The various measures of use of each external resource for the most part correlate positively and highly with each other for the same resource ($p<.001$) so are not all pursued separately below. It is also clear that, overall, a minority of time spent on the task, just 15% on average, is devoted to use of lexicographic resources as a whole.

RQ1: What distinct profiles of online lexicographic resource use can be observed in the participants as they write?

Figures 2 and 3 summarise participants' use of mono/bi-lingual resources and thesaurus in terms of both total number of episodes involving a particular resource and duration, with participants numbered from high to low proficiency.

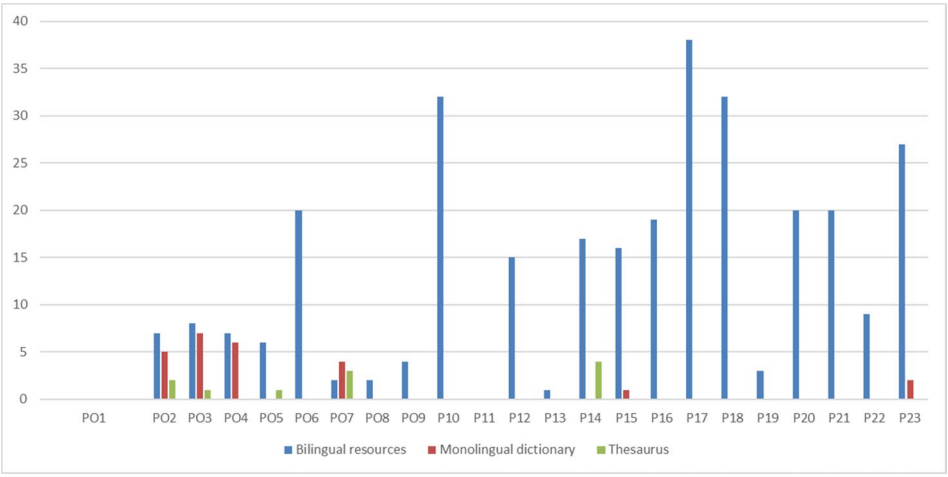

Figure 2. Total number of episodes involving bi/monolingual resources & thesaurus.

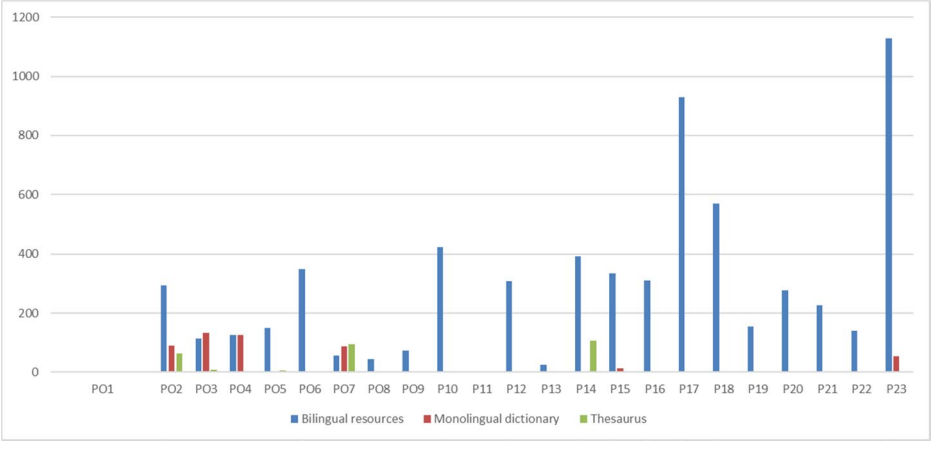

Figure 3. Total duration of episodes involving in bi/monolingual resources & thesaurus (seconds).

Patterns are similar whether the data is considered according to number of resource use episodes or duration, with five groups emerging: writers using no resources at all (N = 1, and the native speaker control, P01); writers using all three types (N = 3); writers using only bilingual resources (N = 13); writers using bilingual resources + monolingual dictionaries (N = 3); writers using bilingual resources + thesaurus (N = 2). Overall, as might be expected, use of monolingual lexicographical resources tends to be concentrated amongst the higher proficiency students (P02 highest proficiency) and bilingual resource use tends to increase amongst the lower proficiency students (P23 lowest proficiency).

RQ2: Which factors, L2 language proficiency and/or type of resource used, independently predict essay length, quality, and time spent on the task?

It is important to examine relationships of the independent variables (L2 proficiency and duration of use of external resources) with dependent variables (essay length, quality, and time spent on the task), where intercorrelations among the former are taken into account (which does not happen in simple correlation analyses, see Table 4 Appendix, Supplementary material). Therefore Optimal scaling was used (see Method section) with results in Table 1.

Use of bilingual resources emerges as highly negatively related both to percent of time spent on writing (of which it consumes on average 13.2%) and essay length, and negatively but non-significantly to quality (essay score). In contrast, essay quality is predicted significantly positively by L2 proficiency, and borderline significantly by thesaurus use (despite that use being low). Use of monolingual dictionary has no significant relationships.

RQ3: Do L2 proficiency and use of lexicographic resources affect essay writing directly, or are they mediated through other variables?

Multiple independent variables may relate to a dependent variable not directly but mediated through each other. Two such matters prompted by the above are as follows: 1. Does bilingual resource use negatively affect essay length, and to an extent score, only through the percent of task time it takes up, leaving less time for actual composing? 2. Does L2 proficiency affect essay writing score directly, or only through the use of external resources, especially thesaurus use and (negatively) bilingual resource use?

Such questions can be assessed statistically, using a technique rarely seen in language research articles. We present the results for the first analysis in detail to demonstrate the type of analysis and its output. Table 2 shows the results for the mediation analysis of number of use of bilingual resources episodes (the focal IV), percent time not using resources (potential mediator, MV), and essay length (DV). The main focus is on whether the IV affects the DV directly or through the mediator. We also include the IVs L2 proficiency (C-test scores) and numbers of monolingual dictionaries and thesauruses episodes so that the effects of these IVs are controlled for. The difference from conventional regression or optimal scaling analyses is that not only are the effects of multiple IVs on a DV assessed together, but also the effect of an IV on a potential mediator variable is considered, and the two-step effect of the IV on DV via the MV is calculated.

Table 2 shows that in this case the focal IV, bilingual resource use, strongly but negatively predicts the MV, percent time spent on writing ($p<.001$) (conventionally referred to as path a). The MV, in turn, positively predicts essay length ($p=.008$) on path b. The combination of

Table 1. Optimal scaling of the relationships between independent variables and essay performance variables.

Dependent:	Essay length				Essay score				Percent time on writing			
Predictor	Beta	df	F	sig	Beta	df	F	sig	Beta	df	F	sig
C-test	.154	1	0.540	.472	.492	1	5.550	.034	−.011	2	0.015	.906
Bilingual resource	−.854	1	6.498	.021	−.198	2	1.273	.311	−1.00	2	153.1	<.001
Monolingual dictionary	−.167	1	0.176	.680	−.267	3	1.659	.221	−.005	1	0.000	.987
Thesaurus	.027	1	0.007	.937	.533	1	4.618	.050	-.017	6	0.015	.985

Table 2. Analysis of percent time composing as a potential mediator between number of bilingual resource episodes (IV) and essay length (DV).

Effect and Path	Variables	B Coefficient	SE	T	p
IV to MV (Path a)	Number of bilingual resource episodes → percent time composing	−0.84	0.16	−5.39	<.001
Direct Effect of MV on DV (Path b)	Percent time composing → essay length	7.03	2.33	3.02	.008
Total Effect of IV on DV (Path c)	Number of bilingual resource episodes → essay length (not considering mediator)	−3.03	1.81	−1.67	.113
Direct Effect of IV on DV (Path c')	Number of bilingual resource episodes → essay length (taking mediator into account)	2.85	2.45	1.16	.262
Partial Effect of Control Variables on DV	C-test → essay length	0.22	1.33	0.17	.869
	Number of monolingual dictionary episodes → essay length	16.64	8.95	1.86	.082
	Number of thesaurus episodes → essay length	18.05	15.55	1.16	.263
				Bias corrected 95% confidence interval*	
Indirect Effect of IV on DV through MV (Path ab)	Number of bilingual resource episodes → percent time composing → essay length	−5.87, −6.93*	3.19	−12.03 to −0.25	

*Bootstrapped estimates (from 1000 resamples).

those two relationships, path ab, is assessed at the end of the table, where the confidence interval −12.03 to −0.25 does not contain the value zero. This shows that there is a significant non-zero negative relationship between bilingual resource use and essay length via the mediator. In contrast, neither estimate for the direct link between bilingual resource use and essay length emerges as significant (paths c and c'). Hence, there is no evidence for a direct effect of bilingual resource use on essay length: it only has an effect via the mediating variable of percent time spent on writing. The equivalent analysis for essay score as the DV shows that, in the presence of the control variables, bilingual resource use does not predict it either directly (path c', p=.440) or indirectly (path ab, confidence interval contains zero: −0.065 to 0.049). Thus the implication is that considerable use of bilingual resources has a negative effect by its consumption of time, at least in a time limited writing task such as ours, and is not necessarily bad due to other reasons such as providing faulty information that would directly affect score.

Addressing our second question, it was shown unequivocally that L2 proficiency has a direct effect on essay score (path c', p=.003), and indeed, independently, thesaurus use also has a significant effect on essay score (path b: p=.005). However, the indirect effect of L2 proficiency via thesaurus use (path ab) is not significant (confidence interval

contains zero: −0.002 to 0.017), although we know that thesaurus use presupposes some level of L2 ability. A parallel analysis exploring whether L2 proficiency predicts essay score directly or mediated (negatively) through bilingual resource use found that both were the case (path c', p=.003; path ab confidence interval does not (by a very small margin) contain zero: 0.001 to 0.027), though bilingual resource use, as we have already seen, did not have an independent effect (path b, p=.067).

Further qualitative exploration, and discussion

Given the complexity of the relationships between L2 proficiency, use of lexicographic tools, and L2 writing measures already revealed by the quantitative findings above, our discussion will focus exclusively on the main findings. Furthermore, our methodological commitment dictates that the discussion of statistical results is contextualised and, we argue, enriched, through the qualitative perspective afforded by the SRR data. As detailed in the Method section, the SRR was, in turn, facilitated by the use of the screen-capture of real-time writing activity as well as gaze data produced by the eye-tracker (Figure 1). The SRR thus represents an important source for understanding some aspects underlying the use of the online resources.

First, the results showed a significant positive relationship between L2 proficiency and essay score. Importantly, the mediation analysis indicated that this association was overwhelmingly direct and only marginally had effect due to more proficient participants occupying less of their available time on bilingual resource use or using thesauruses more. This finding represents an important first step in addressing a key knowledge gap as identified in our literature review and highlighted by Tono et al. (2014) by showing that greater unsupervised use of lexicographic resources does not necessarily affect overall product quality of L2 writing. The native speaker benchmark figures (see Appendix, Supplementary material), although not included in the statistical analysis, contribute to the validity of these findings. Compared to the student writers, and as might be predicted, the native speaker produced a longer essay within the allotted time without any consultation of external resources and obtained the highest score available (9.0).

We also found that, on average, use of lexicographic tools only amounted to 15% of the total time on task. Comparisons with other work are extremely difficult because this aspect has been seriously neglected in studies of lexicographic tool use. This is unfortunate given the importance of understanding the temporal dimension in L2 writing processes (Gánem-Gutiérrez & Gilmore, 2018; Manchón, Roca de Larios, & Murphy, 2009). Two researchers that have turned their attention to

this matter are Hvelplund (2017) and Stapleton (2010) although their studies differ contextually and methodologically from ours. Hvelplund (2017) found that his participants (18 professional translators) spent 19% of their composing time consulting non-specialised lexicographic tools for a short translation task; in other words, similar resources, and not a dissimilar proportion of time, to those consulted by our participants. Despite a lack of specific figures reported in Stapleton (2010), his study serves as an interesting point of contrast because it focuses on a non-time-limited task, a master's student writing a 4000-word essay over a period of three weeks and reports 'frequent application of electronic tools' (Stapleton, 2010, p. 304).

Nevertheless, participants who spent less time consulting online resources in our study tended to write more (rho(21)=-.466, p =.010), and, since longer essays were generally scored higher by raters (rho(21)=.640, p =.001), this trade-off of length over accuracy might reap rewards under timed-conditions such as exams, for example. Observations from the retrospective recall data confirm that the time factor affects willingness to use resources for some participants; for instance, 'so eh so so I should have searched *give* in (.) dictionary but I maybe eh I run out of time and I I wrote other words' (P03). Similar views from L2 students have been reported in other studies, e.g. Lai and Chen (2015).

The detailed patterns of lexicographic tool use reported in the Results section can be summarised as follows: One L2 writer, and our native speaker control, made no use of resources; the majority of participants (N = 13) used bilingual resources exclusively while eight participants showed more sophisticated behaviour by using a combination of tools (3 used all three types whereas 5 used a bilingual resource plus either a monolingual one or a thesaurus).

In relation to such patterns, our data corroborates that individual differences and preferences play an important role, as suggested in previous studies (e.g. Hvelplund 2017; Lai & Chen, 2015; Mraček, 2020; C. Yoon, 2016b). Amongst the 22 participants, the range of overall external resource consultations varied considerably, from 0 to 38. Although we might have expected weaker students to rely more on such resources when writing, that was not seen straightforwardly in Figures 2 and 3 (where bars do not rise systematically left to right), and L2 proficiency was not significantly related to the percentage of time spent on writing versus using lexicographic tools (Table 1). The SRR transcripts for the four no/low-frequency users (P11, P13, P08, P09, see Appendix, Supplementary material) show that they opted to fall back on their own L2 knowledge during the writing event, despite encountering lexical and grammatical problems in a similar way to other participants who

consulted the external tools. For example, P11 explains a word change from *thus* to *and* in his essay by commenting, 'I don't know the meaning of thus completely [laughs] so I put another word'. He therefore appeared to choose a quick, although not necessarily ideal, solution of a well-known 'all purpose' word, even though a brief consultation of an online dictionary could have clarified the difference in meaning between the two target words. As well as potentially adding sophistication to the essay, consulting the dictionary might have also contributed an opportunity for developing his interlanguage (Laufer & Hill, 2000; Lew & Adamska-Sałaciak, 2015).

As mentioned above, only one student (P11) used no available resources while writing. Although he does not explain his reasons during the SRR, it is clear that, whatever his motivation, it is not because he has no need of support – numerous episodes arise during the composition process where he could have benefitted from dictionary/thesaurus consultations. For example, he mistakenly uses the word *worship* rather than *welfare* in his essay, writing *It has been thought that people who have higher knowledge get higher jobs, which have good-paying, worship, and so on*. During the video play-back of the writing event, P11 pauses to reread this sentence and his eye gaze focuses on *worship*, suggesting some attention to that term is occurring (Hvelplund, 2017; Roberts & Siyanova-Chanturia, 2013), and this, along with the odd choice of word given the context, prompts the tutor to query it:

In line 3, after reflection, P11 realizes his mistake when his intended word *welfare* springs to mind and asks the tutor to clarify the meaning of *worship*. In line 11 he then defines *welfare* as something provided by companies in *higher jobs* to employees who have *higher knowledge* and this elicits the suggestion of *benefits* from the tutor as a more appropriate choice. Had the student checked one of the lexicographic resources available during the writing event, he may well have been able to resolve these issues himself, but see discussion regarding training below.

Overall, and as might be expected, use of monolingual lexicographical resources tended to characterize the higher proficiency students (dictionary rho(21)=.421, p =.051; thesaurus rho(21)=.500, p =.018), which presumably reflected a greater confidence or familiarity with these English-only sources. Bilingual resource use tended to increase amongst the lower proficiency students (rho(21)=-.464, p =.030) although most participants relied predominantly on bilingual resources for lexicographic support, 89% of episodes, consistent with previous reports on this issue (e.g. Lai & Chen, 2015; Lew & Adamska-Sałaciak, 2015; Lew & Szarowska, 2017).

Both L2 proficiency and thesaurus use were found to have a direct and positive effect on essay score, although an indirect effect of L2

proficiency via thesaurus use was not significant. In other words, more proficient students wrote better essays, as did those who referred to thesauruses more frequently. However, despite the apparent benefits of thesaurus consultations, they accounted for only 0.57% of total task time (Table 3 Appendix, Supplementary material), corresponding to just 11 searches amongst all 22 participants. Of these searches 5 were carried out by a single student, P14, who also stood out as being a more sophisticated dictionary user. P14 uses the thesaurus strategically to avoid repetition in her essay and confirm her understanding of specific words. For example, her first thesaurus search is triggered after she begins a new paragraph: *At first, in public education including elementary school, middle school and high school, the education should be free for everyone. The reason is that in those public educations, we can learn the basic knowledge for living*. She immediately deletes *for living* and changes the adjective *basic* to *fundamental*, commenting in the SRR 'It's more academic word […] sounds more academic'. This demonstrates a level of understanding of some subtle differences between the two potential choices here: although in the COCA corpus both words are seen predominantly in academic texts (*basic* = 41.45%; *fundamental* = 50.23%), *basic* is over twice as frequent as *fundamental* (*basic* = 46,200 hits; *fundamental* = 20,509 hits) and in Nation's BNC/COCA headword list[4] also appears in the first 1000 most frequent words compared to the third 1000 most frequent words for *fundamental*. This writer seems to have had sufficient exposure to English to become sensitive to this distinction, although in this case both words could be considered appropriate.

P14 then copy-pastes the word *fundamental* from her essay into the online thesaurus, explaining in the SRR, 'checked again fundamental or any other appropriate word'. The results page in the Merriam Webster thesaurus (see Figure 4) is cluttered with display advertising around the periphery, including banners, Flash animations, and video (see Dziemianko, 2019 for adverse effect of ads), but the student is able to focus exclusively on the search results in the center of the page, suggesting some familiarity with this online resource.

P14's eye gaze data shows that she tracks straight down from the search word *fundamental* at the top of the results box to the synonyms section which lists 12 suggestions given in alphabetical order (see Figure 4). Unfortunately, no guidance is provided as to the frequency of these words in the English language (e.g. *abecedarian* has only 19 hits in COCA, mostly related to the Carolina Abecedarian Project in 1972), nor their register (e.g. *meat-and-potatoes* is informal American English), or common collocates, an area of notorious difficulty for L2 learners to master (Boers & Webb, 2018; Frankenberg-Garcia, 2020). Thus, writers who did not already know the words would need to perform follow-up

searches in mono/bi-lingual dictionaries for this information to be of any real value. In this case, P14's main goal is to confirm her understanding of *fundamental* and after 16 seconds, having satisfied herself that it does indeed carry her intended meaning, she returns to her essay to continue composing. In her next sentence she writes *In addition, we can also learn essential social skills as a member of society*, picking up another synonym (*essential*) from the thesaurus results for *fundamental* to avoid repetition in her academic writing.

Our final reflections in this section relate to the important aspect of training which, while not a component of our research design for the reasons given in the Method section, is clearly of relevance to the broader discussion on lexicographic tool use. The studies reviewed in the Background section included training as part of their designs, albeit of varying lengths and types. Despite this, the impact of the lexicographic tool was, as put by Lai and Chen (2015, p. 352), 'not what one would have expected'… 'even with considerable training.' While students might perceive resources as being beneficial, the actual evidence is less straightforward. The core argument we would like to pose here, however, is not one that questions the necessity of training in use of online lexicographic resources; on the contrary, we consider this essential, at least for our participants as we will further illustrate below, but very possibly

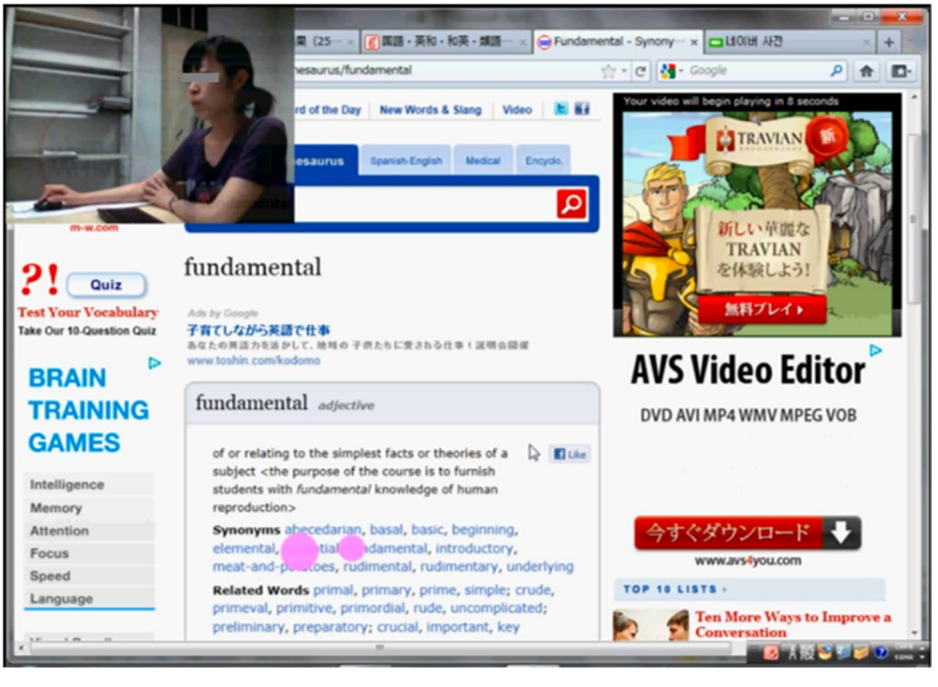

Figure 4. Eye gaze data from writing event for P14 showing her attention on synonyms of fundamental.

in the wider L2 context. We therefore concur with Chon's (2009) thoughts:

> Research on dictionary users and use does not give enough attention to the cognitive process of using dictionaries in L2 writing, regardless of dictionary type or medium, though successful language production depends considerably on the ability to make appropriate lexical choices in dictionary entries, and to incorporate chosen items into continuous text following accepted collocational and syntactic norms (Chon, 2009, p. 24).

These issues can be further discussed in light of findings from the mediation analysis which indicated that bilingual resource use has an effect on essay length via the mediating variable of percent time spent on writing but does not predict essay score either directly or indirectly. Hence, despite the time, and cognitive cost (Hvelplund, 2017) involved in consulting one of the bilingual resources, it did not appear to consistently improve the overall quality of the text. Learners therefore clearly need to weigh up the benefits of pausing the text construction process in order to consult a dictionary rather than relying on their own lexical or grammatical resources. This is particularly true if dictionaries are not used effectively as seems to be the case with many of the students in this study.

The majority of bilingual searches (96.6%) were from Japanese to English which is to be expected in a writing task where the predominant need is to find suitable English translations for L1 words and phrases. Participants did not seem to distinguish between the two bilingual resources available during the writing event (Eijiro ALC and Goo), generally choosing either one or the other for the entire task (see also Lai & Chen, 2015). Only one student (P14) systematically used both, explaining her rationale in the SRR: 'For checking authentic just one to one words I usually check the Goo. Mm and some words wide range of words check ALC that time. ALC and Eijiro mm suggest many word […] Lots of choices […] Or just quick reference is convenient using Goo.' The fact that Eijiro ALC was primarily used in the J > E direction is worrying since, as mentioned in the Method section, this version is simply reverse engineered without any human editing and contains numerous errors and misleading information (see Lew & Szarowska, 2017 for a discussion on quality issues in lexicographic resources). For example, a query in Eijiro ALC by P14 for a translation of the Japanese word 欠点 (*ketten* = weak point) produced the search results shown in Figure 5:

This search was triggered by a need to find an appropriate adjective to complete the clause *The _____ of this system is*…. As can be seen above, of the 25 suggested translations, only around nine (demerit, drawback, failing, flaw, limitation, shortcoming, trouble, weak point, weakness) seem possible in this context and many of the others (bad

patch, blemish, claw-back, hole, vitium, wart) are positively unhelpful. Students familiar with typical dictionary lay-outs might expect the search results appearing at the top of the list to be the most common or appropriate (Tono, 2011) but this is not the case here and probably the most obvious choice, *disadvantage*, is not included at all. P14 spends 18 seconds scanning through the search results and the eye-tracking data, Figure 6, shows that she fixates on 7 of the suggested translations (claw-back > demerit > claw-back > inadequacy > objection > deficit > bad patch > shortcoming). She moves the cursor arrow next to the word *deficit*, suggesting a particular interest in this option which was subsequently confirmed in the SRR where she comments: 'I thought deficit is not appropriate in this paper so just eh eh gave up then (laughs) using the proper another word […] first eh top of my mind disadvantage is better'. Ultimately then, she rejects all of the suggested translations and falls back on her own L2 knowledge to write *However, the disadvantage is* in her essay.


欠点

- bad patch
- blemish
- claw-back
- defect
- deficit
- demerit（長所と比べた）〔通例、demerits〕
- drawback
- failing〔fault の婉曲的表現〕
- fault
- flaw
- frailty（人間の意志の弱さから来る）〔通例、frailties〕
- hole
- imperfection
- inadequacy
- kink（機器・計画などの）
- limitation（弱みとなる）〔通例、limitations〕
- negative〈話〉
- shortcoming（人・組織・物・制度などの）〔【用法】通例、複数形〕
- trouble
- vitium〈ラテン語〉
- want（性格の）
- wart
- weak point
- weakness〔可算〕
- wrinkle（問題を起こす）

Figure 5. Search results for (ketten = weak point) in Eijiro ALC.

Figure 6. Eye gaze data from writing event for P14 showing her first fixation on claw-back.

In sum, we argue that our examples underscore the need for training in the use of online lexicographic resources for L2 writing, but that training needs to be informed by a deeper understanding of how L2 writers actually use those tools on the one hand, and of the complex cognitive processes involved, on the other. This kind of research requires robust, mixed-methods approaches, that also take advantage of current technologies (Hamel & Séror, 2016; Hvelplund, 2017; Lew, Grzelak, & Leszkowicz, 2013). While the affordances offered by some of these resources (Frankenberg-Garcia, 2020; C. Yoon, 2016b) render them potentially powerful tools, we know that 'consulting electronic dictionaries or thesauruses may not be a guaranteed means of solving lexical problems in L2 writing' (Chon, 2009, p. 24).

Conclusion and pedagogical implications

The overarching aim of this study was to contribute to our understanding of complex cognitive processes in L2 writing as currently practised. We specifically refer to the increasing use of computer-based applications, online lexicographic resources being a case in point. In particular, the paper aimed to fill important gaps in the literature through its objectives: to investigate profiles of use of these tools across a range of L2 proficiencies and, crucially, to ascertain the potential impact that using those resources might have on the actual quality of the written product, taking language proficiency into account. To achieve these goals we used a

mixed-methods research design, aided by digital-screen capture and eye-tracking technology.

Our findings on relationships between L2 proficiency, writing performance, and use of lexicographic resources offer interesting paths for reflection. While the fact that L2 proficiency appears to be the main predictor for writing quality is not entirely surprising (Cumming, 1989), our study is unique in showing that, at least in the context of our investigation, that result was not to a great extent mediated by use of lexicographic tools. There are at least two possibly interrelated implications of this, one is that those learners who could, potentially, be main beneficiaries from their use are not necessarily exploiting them to improve their essay scores. The second one is that even when the resources are being used, they might not be having the positive impact that we, educationalists, would wish for.

These points can be further explored in relation to the impact of length of time spent consulting them and the range of tools used. Regarding time, we found that some of our participants felt under time pressure to finish the task. However, this feeling is not necessarily unique to the type of task, timed essay, used; students completing regular academic assignments, for example, have mentioned time pressure as a reason for under-using similar tools (C. Yoon, 2016b). Furthermore, it was evident from our data that a limited range of resources characterised the patterns of their use. This clearly has direct implications regarding training. It could be argued that our results reflect the fact that training was not part of the research design; however, previous studies show that even when training is given, the potential of these resources is far from being maximised (Kennedy & Miceli, 2017; Lai & Chen, 2015).

It is therefore evident that, as educators and researchers, we need to continue our efforts to ensure that L2 writers become really empowered to maximise the potential of these tools. This requires a multifaceted approach which includes helping them see the specific affordances of the various resources, the importance of adopting a strategic approach to linguistic problem-solving, and also ensuring that research ultimately feeds into better pedagogy on the one hand, and improvement of the resources, on the other. This approach is illustrated in recent work by Frankenberg-Garcia (2020), for example; and should be part of a general strengthening of what is referred to as digital literacy for academic purposes, see Conroy (2010); Frankenberg-Garcia (2005); and C. Yoon (2016b) among others. Furthermore, we would like to argue that research into the use of these tools for L2 writing needs to adopt designs that capture *complete* L2 composing events to achieve a more nuanced understanding of how learners use these tools and for what purposes in real-time. We hope that the detailed examples we provided might go some way into helping teachers understand some of the issues students face when

consulting similar tools and, in turn, aid the development of training resources based on students' actual behaviour and needs. Further investigations into the various matters raised in this paper should help to contextualise our findings and continue efforts to strengthen pedagogies for L2 writing.

Notes

1. Concordancers are powerful electronic tools which enable users to search corpora such as the British National Corpus (BNC) or COBUILD; these comprise large collections of texts produced by native speakers of English. Users type in a query word or phrase and the concordancer generates 'concordance lines' extracted from the corpus and thus providing examples in context of naturally occurring data (Gilmore, 2009; Hyland, 2003).
2. Tobii T60 collects raw eye-movement data points every 16.7 milliseconds, with each data point given a time stamp and x/y coordinates that are subsequently used to establish the location of the fixation.
3. IELTS Task 2 writing band descriptors (public version) available at: https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf.
4. For more information, see: https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf.

Acknowledgments

We would like to thank the editor of *Computer Assisted Language Learning* and the anonymous reviewers for their constructive comments on earlier versions of this paper. We are deeply grateful to Phil Scholfield for his invaluable help with the statistical analysis. A special thank-you to our participants as well as to Michiyo Yamazaki for her help at earlier stages of the project.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributors

Gabriela Adela Gánem-Gutiérrez is a Senior Lecturer in Applied Linguistics in the Department of Language & Linguistics at the University of Essex. Her main research interests include L2 pedagogy, explicit L2 learning, Sociocultural theory, applied Cognitive Linguistics for L2 learning and teaching, and CALL.

Alexander Gilmore is Associate Professor (English for Academic Purposes) in the Department of Civil Engineering at the University of Tokyo, Japan. He has worked as a language teacher, teacher trainer, and applied linguistics researcher in Spain, Mexico, Saudi Arabia, England and Japan. His research interests include authenticity in language learning, discourse analysis, corpus analysis, materials design and language pedagogy.

ORCID

Alexander Gilmore http://orcid.org/0000-0001-6380-0601

References

Bernard, H. R., & Ryan, G. W. (2010). *Analyzing qualitative data: Systematic approaches*. London: Sage.

Boers, F., & Webb, S. (2018). Teaching and learning collocation in adult second and foreign language learning. *Language Teaching*, *51*(1), 77–89. doi:10.1017/S0261444817000301

Chon, Y. V. (2009). The electronic dictionary for writing: A solution or a problem? *International Journal of Lexicography*, *22*(1), 23–54. doi:10.1093/ijl/ecn034

Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology*, *26*(6), 861–882. doi:10.14742/ajet.1047

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, *39*(1), 81–141. doi:10.1111/j.1467-1770.1989.tb00592.x

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, *7*(4), 482–511. doi:10.1177/0741088390007004003

Dziemianko, A. (2019). The role of online dictionary advertisements in language reception, production, and retention. *ReCALL*, *31*(01), 5–22. doi:10.1017/S0958344018000149

Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography, 18*(3), 335–355. doi:10.1093/ijl/eci015

Frankenberg-Garcia, A. (2020). Combining user needs, lexicographic data and digital writing environments. *Language Teaching*, *53*(1), 29–43. doi:10.1017/S0261444818000277

Fujii, Y. (2007). Making the most of search engines for Japanese to English translation: Benefits and challenges. *The Asian EFL Journal Professional Teaching Articles*, *23*, 1–36.

Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, *68*(2), 469–506. https://. doi:10.1111/lang.12280

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.

Gilmore, A. (2009). Using online Corpora to develop students' writing skills. *ELT Journal*, *63*(4), 363–372. doi:10.1093/elt/ccn056

Gilmore, A. (2011). "I prefer not text": Developing Japanese learners' communicative competence with authentic materials. *Language Learning*, *61*(3), 786–819. doi:10.1111/j.1467-9922.2011.00634.x

Gilmore, A., & Gánem-Gutiérrez, G. A. (2020). Investigating complexity in L2 writing with mixed methods approaches. In G. G. Fogal & M. H. Verspoor (Eds.), *Complex dynamic systems theory and L2 writing development* (pp. 183–206). Amsterdam: John Benjamins. doi:10.1075/lllt.54.08gil

Grotjahn, R. (2010). *The C-Test: Contributions from current research*. Bern: Peter Lang.

Hamel, M.-J., & Séror, J. (2016). Video screen capture to document and scaffold the L2 writing process. In C. Caws & M.-J. Hamel (Eds.), *Language-learner computer interactions* (pp. 137–162). Amsterdam: John Benjamins.

Hvelplund, K. T. (2017). Translators' use of digital resources during translation. *HERMES – Journal of Language and Communication in Business*, *56*(56), 71–87. doi:10.7146/hjlcb.v0i56.97205

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, *14*(1), 28–44.

Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, *30*(1–2), 91–114. doi:10.1080/09588221.2016.1264427

Lai, S.-L., & Chen, H.-J H. (2015). Dictionaries vs concordancers: Actual practice of the two different tools in EFL writing. *Computer Assisted Language Learning*, *28*(4), 341–363. doi:10.1080/09588221.2013.839567

Latif, M. M. A. (2008). A state-of-the-art review of the real-time computer-aided study of the writing process. *International Journal of English Studies*, *8*(1), 29–50.

Latif, M. M. A. (2019). Eye-tracking in recent L2 learner process research: A review of areas, issues, and methodological approaches. *System*, *83*, 25–35. doi:10.1016/j.system.2019.02.008

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*(2), 61–81.

Leijten, M., & van Waes, L. (2013). Keystroke logging in writing research using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358–392. doi:10.1177/0741088313491692

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, *5*(3), 285–337. doi:10.17239/jowr-2014.05.03.3

Levy, M., & Steel, C. (2015). Language learner perspectives on the functionality and use of electronic language dictionaries. *ReCALL*, *27*(2), 177–196. doi:10.1017/S095834401400038X

Lew, R. (2016). Dictionaries for learners of English. *Language Teaching*, *49*(2), 291–294. doi:10.1017/S026144481500049X

Lew, R., & Adamska-Sałaciak, A. (2015). A case for bilingual learners' dictionaries. *ELT Journal*, *69*(1), 47–57. doi:10.1093/elt/ccu038

Lew, R., Grzelak, M., & Leszkowicz, M. (2013). How dictionary users choose senses in bilingual dictionary entries: An Eye-Tracking Study. *Lexikos*, *23*(1), 228–254. doi:10.5788/23-1-1213

Lew, R., & Szarowska, A. (2017). Evaluating online bilingual dictionaries: The case of popular free English-Polish dictionaries. *ReCALL*, *29*(2), 138–159. doi:10.1017/S0958344016000252

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes: Implications for theory. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 102–129). Bristol: Multilingual Matters.

Mraček, D. (2020). Uncovering the Potential of the Google Search Engine for L2 Learning and L2 Translator Training. In E. Popescu, T. Hao, T.-C. Hsu, H. Xie, M. Temperini, & W. Chen (Eds.), *Emerging technologies for education* (Vol. 11984, pp. 109–121). Cham: Springer International Publishing. doi:10.1007/978-3-030-38778-5_14

Mueller, C. M., & Jacobsen, N. D. (2016). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, *28*(1), 3–21. doi:10.1017/S0958344015000142

Park, K., & Kinginger, C. (2010). Writing/thinking in real time: Digital video and corpus query analysis. *Language Learning & Technology*, *14*(3), 31–50.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, *69*(2), 165–177. doi:10.1093/elt/ccu062

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition*, *41*(3), 605–631. doi:10.1017/S027226311900024X

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*(2), 213–235. doi:10.1017/S0272263112000861

Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, *17*(1), 30–47. doi:10.1016/j.jslw.2007.08.005

Séror, J. (2013). Technologie de capture d'écran: une fenêtre numérique sur le processus d'écriture des étudiants [Screen capture technology: A digital window into students' writing processes]. *La Revue Canadienne de L'apprentissage et de La Technologie*, *39*(3), 1–16. http://cjlt.csj.ualberta.ca/index.php/cjlt/article/view/707.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445.

Stapleton, P. (2010). Writing in an electronic age: A case study of L2 composing processes. *Journal of English for Academic Purposes, 9*(4), 295–307. doi:10.1016/j.jeap.2010.10.002

Steel, C. H., & Levy, M. (2013). Language students and their technologies: Charting the evolution 2006–2011. *ReCALL*, *25*(3), 306–320. doi:10.1017/S0958344013000128

Tono, Y. (2011). Application of eye-tracking in EFL learners' dictionary look-up process research. *International Journal of Lexicography*, *24*(1), 124–153. doi:10.1093/ijl/ecq043

Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL*, *26*(2), 147–162. doi:10.1017/S095834401400007X

Varley, S. (2009). I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, *22*(2), 133–152. doi:10.1080/09588220902778294

Weir, C. (1990). *Communicative language testing*. Hoboken, NJ: Prentice Hall.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A Professional Framework for Multimodality Research*. Proceedings of LREC 2006. Fifth International Conference on Language Resources and Evaluation.

Yoon, C. (2016a). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, *20*(1), 209–229.

Yoon, C. (2016b). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers. *Journal of Second Language Writing*, *32*, 67–80. doi:10.1016/j.jslw.2016.04.002

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, *12*(2), 31–48.

Yoon, H., & Hirvela, A. (2004). ESL student attitude toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257–283. doi:10.1016/j.jslw.2004.06.002