

State of the Scholarship

METHODOLOGICAL ADVANCES IN INVESTIGATING L2 WRITING PROCESSES CHALLENGES AND PERSPECTIVES

David Galbraith

University of Southampton

*Ineke Vedder**

University of Amsterdam

This special issue is devoted to describing and evaluating new methods for analyzing online writing processes in a second language (L2). As all the contributors have stated, the aim is to identify the cognitive processes underlying writing. The ultimate aim is to employ observations and experimental manipulations of these processes to build theories that can be used to help writers write more effectively in their second language and, as López-Serrano, Roca de Larios, and Manchón (2019) argue, to facilitate learning of the second language.

This project has two sides: (i) developing methods for recording and analyzing behavior as it unfolds during writing and (ii) relating this behavior to underlying cognitive processes. Furthermore, these two sides to the question are interdependent: what behavior is recorded and how it is analyzed depends on the researcher's theory of the cognitive processes involved, and, to a certain extent, theories of these cognitive processes depend on what it is possible to observe. In this commentary on the articles in the special issue, we will first consider the theoretical models within which the research has been framed. We will then consider the contribution of the articles to developing methods for observing the writing process, and how these have been related to underlying cognitive processes. Finally, we will consider directions for future research and some of the possible implications for second language writing instruction.

*Correspondence concerning this article should be addressed to Ineke Vedder, Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam. E-mail: i.vedder@uva.nl

MODELS OF THE WRITING PROCESS

The articles in this issue have taken as their basic framework the classic cognitive models of writing developed by Hayes and Flower (1980, 1986; Hayes, 2012a), and subsequently modified and developed by Bereiter and Scardamalia (1987) and Kellogg (1996), among others. A key feature of empirical research informed by these models was the use of verbal protocols to provide a rich picture of the thinking behind the text. As well as providing a characterization of the different kinds of basic processes involved—planning, translation, and reviewing, which occurred recursively throughout writing—this body of research had two themes, which have dominated research in the field since.

First, writing is goal directed; it is an example of what Bereiter and Scardamalia (1989) called *intentional cognition*, and the written product depends fundamentally on the goals toward which it was directed rather than on just, for example, the linguistic proficiency of the writer. This led to a characterization of novice writing as being largely a matter of knowledge telling, in which the writer's goal is essentially to transcribe content as it is retrieved from long-term memory into an appropriate written form. By contrast, expert writing was characterized as a knowledge-transforming process, in which the writer's goal was to design a text to have a rhetorical effect on the reader, and in which content was created and transformed to achieve this goal, rather than simply being retrieved from long-term memory.

Second, a fundamental constraint on the writing process is cognitive overload: conflicting demands on limited cognitive resources may prevent writers from carrying out the component processes effectively, even if these are directed at appropriate goals (Flower & Hayes, 1980). Thus, a further distinction between novice and expert writers is that novices are more likely to employ a “single draft” strategy, in which they try to write a complete text straight away, carrying out all the component processes at the same time, whereas expert writers are more likely to focus on different components at different points in writing, developing more elaborate plans for text, and revising more extensively after completing a draft (Bereiter & Scardamalia, 1987; Hayes & Flower, 1986; Kellogg, 1988).

This focus on cognitive load was a relatively general feature of early models of writing but was made more precise by Kellogg's (1996) analysis of how different components of the writing process mapped on to a componential model of working memory (see also Révész & Michel, 2019). This more precise description of cognitive capacity enables more specific analysis of the potential conflicts for resources during writing, focusing not just on those between the global writing processes of planning, translation, and reviewing but also on more specific ones between processes (e.g., lexical retrieval and sentence production) assumed to draw on the same components of working memory (see Kellogg, Whiteford, Turner, Cahill, & Mertens, 2013; Olive, 2014, for reviews). This more detailed view of working memory as the limited capacity system within which writing processes are carried out is now a standard component of cognitive models of writing, and has been incorporated explicitly in Hayes's (2012a) most recent versions of the original Hayes and Flower model of writing.

As Révész and Michel remark in the introduction to the special issue, early research on writing in the first language (L1) tended to focus on the higher-level thinking processes

involved in writing, rather than on the detailed processes by which thoughts were formulated in written language. It is only relatively recently that research has allocated more detailed attention to the processes involved in text production (Chenoweth & Hayes, 2001; Hayes, 2012b; Hayes & Chenoweth, 2006; Kaufer, Hayes, & Flower, 1986). An important finding here has been that writers produce shorter P-bursts when writing in L2 than in L1, and that the length of the bursts in L2 varies depending on the writer's L2 experience (Chenoweth & Hayes, 2001). (Chenoweth and Hayes define P-bursts as bursts of language produced between pauses of two seconds or more.) Furthermore, L1 writers and writers with more L2 experience produced a higher proportion of P-bursts relative to R-bursts (bursts terminated by a revision rather than a pause) (Chenoweth & Hayes, 2001; Hayes, 2012b; Hayes & Chenoweth, 2006). This raises the question of whether L2 translation is a potential impediment to higher level thinking. It is reflected in the studies in this issue by Barkaoui, and Chukarev-Hudilainen, Feng, Saricaoglu, and Torrance, which explore whether difficulties with the translation process impede higher-level idea generation and content-planning processes.

The dual-process model of writing (see Galbraith & Baaijen, 2018, for a recent review), by contrast, questions this basic assumption that the translation process is a passive process, responsible for translating the output of ideational planning into words, and suggests instead that translation can be an active knowledge-constituting process in its own right. This leads to an alternative view in which a central conflict in writing, over and above that represented by cognitive overload from having too many things to think about, is between the emergent content produced by the knowledge-constituting process and preestablished content generated by explicit planning processes. One consequence of this is that, although it is possible that less efficient translation will impede higher-level planning, it is also possible that higher-level planning processes will reduce the knowledge-constituting effects of translation. Although there is some support for this in recent L1 research (Baaijen & Galbraith, 2018), no research has been carried out in an L2 context. There are important questions about how any such knowledge-constituting effect is affected when writing, or indeed speaking, in a second language. To what extent is an L2 capable of serving as a vehicle of the writer's thinking, and how is this affected by L2 proficiency (cf. Roca de Larios, Nicolás-Conesa, & Coyle, 2018)?

APPROACHES TO ANALYZING ONLINE PROCESSES

Having sketched some of the broader theoretical context within which these studies have been carried out, we turn now to discussing the individual articles. To do so, we have divided them up into three groups: those that use think-aloud protocols (TAP), and hence provide information about the content of writer's thoughts during writing (López-Serrano, Roca de Larios, & Manchón, 2019); those that use either keystroke logging and/or eye movements, and hence provide moment-by-moment information about the distribution of different types of processes (Barkaoui, 2019; Chukarev-Hudilainen et al., 2019; Leijten, Van Waes, Schrijver, Bernolet, & Vangehuchten, 2019); and those that combine keystroke logging with eye-tracking and retrospective verbal protocols, and hence provide information about the distribution of different types of processes and (retrospectively) about the content of the writer's thoughts (Révész, Michel, & Lee, 2019).

DIRECT OBSERVATION OF THE THINKING BEHIND THE TEXT

A contrast that is often made between TAP-informed research and eye-tracking studies is that, while TAP studies are assumed to give access to higher-level thinking processes, they do not provide much information about the specific processes involved in translation. Although there are general issues with TAP methods arising from reactivity and the theory-laden nature of self-reports, the article by López-Serrano et al. demonstrates that, in fact, TAP methods can provide a window into some of the writer's concerns as they grope their way to formulating a proposition in words. Consider the example given by López-Serrano et al. of a language-related episode (LRE) (Table 2, López-Serrano et al., 2019). This shows a writer's protocol, expressed in their L2 (English in this case) as they formulated the final clause of a sentence, "as I *have been* an eager learner since my early childhood" (our italics). In the final written product, this would appear straightforward as a reasonably formulated clause. By taking us behind the scenes, López-Serrano et al. reveal that the italicized words were produced only after considerable reflection. Furthermore, although keystroke logging combined with eye-tracking would represent the clause as: "as I ~~started~~ (PAUSE) (RE-READING within sentence) was (PAUSE) ~~was~~ have been an eager learner since my early childhood," and hence would capture the reflective event, it would not capture the fact that the concept of "learnt" is the topic of thought during the pauses and is introduced as soon as "started" is deleted. TAPs, therefore, can provide additional information about specific features of the process by which translation is carried out.

As López-Serrano et al. suggest, it would be valuable to examine how other online methods such as keystroke logging could be used to examine the units of analysis they have developed (see their Figures 2 and 3). Of the dimensions that López-Serrano et al. identified, linguistic focus seems to be comparable to the distinction made by keystroke-logging studies between different levels of text production (e.g., word, sentence, and paragraph), particularly when transition times between keystrokes rather than pauses above a certain threshold are analyzed. Similarly, the classification of language-related episodes in terms of the extent to which problems have been resolved should be equally applicable to a keystroke log. By contrast, there is a larger disparity for inferences about the strategies that writers are using. Of those that López-Serrano et al. identify in Figure 3 of their article, only rereading, and generating and assessment of alternatives seem to be readily available in a keystroke log. A central question here would be whether similar units could be identified through retrospective verbal protocols or stimulated recall.

Perhaps the most important additional information provided by TAPs is information about the writers' goals. For the classical models of writing that we discussed earlier, differences in writers' goals are a key factor in writing expertise. It is noteworthy that, of the articles in this issue, López-Serrano et al.'s contribution is the only one that discusses these in much detail. This is an example of how the observational method that one chooses can influence the theoretical account that can be constructed. López-Serrano et al. conceptualize LREs as problem-solving strategy clusters, in which episodes are integrated by the goals toward which they are directed. This led them to classify LREs in terms of their orientation: (i) a compensatory orientation (in which the writer's goal was to compensate for deficiencies in their L2 knowledge), and (ii) an upgrading orientation (in which the writer's goal was to revise and improve existing expressions). They suggest

that the upgrading orientation may “trigger the expression of more complex or precise ideas.” This aspect of López-Serrano et al.’s analytic scheme is particularly interesting because it does not seem to be directly inferable from a keystroke log. Information about orientation would be valuable in disambiguating keystroke observations.

Furthermore, the distinction is similar to that made in recent research in L1 on individual differences in writing beliefs (White & Bruning, 2005). For example, Baaijen, Galbraith, and De Glopper (2014) suggested that writers with high transactional beliefs—writers who view writing as a process of developing ideas in the course of writing—orientate revision toward the development of their understanding. By contrast, writers with low transactional beliefs—writers who believe that writing is a matter of translating preconceived ideas into text—orientate revision toward compensating for errors in the text. However, in the absence of direct information about the writer’s goals during writing, this suggestion remained an inference based on the individual difference measure of writing beliefs. Retrospective protocols analyzed using the orientation category developed by López-Serrano et al. could in principle enable this hypothesis to be tested. In general, if the categories identified by López-Serrano et al. can be reliably identified retrospectively, they would be a valuable tool for assessing the effects of a variety of variables on writers’ goals. For example, it would be interesting to test whether the relative balance of compensatory and upgrading goals varied depending on writers’ L2 proficiency.

INDIRECT OBSERVATION OF ONLINE WRITING PROCESSES

The next group of articles to be discussed all used keystroke logging and/or eye-tracking without the additional information about the content of the writers’ thoughts available from concurrent or retrospective verbal protocols. These methods provide a wealth of data for analysis. However, this research is still in its infancy, and there is no consensus on how to analyze such data, or on how to assess their relationship with other variables. This collection of articles, therefore, offers a valuable opportunity to consider the range of approaches that can be taken to using these data, and to assess the strengths and weaknesses of these approaches. For a more general discussion of the problems that arise in relating keystroke measures to underlying cognitive processes, and recommendations designed to increase this alignment, see Baaijen, Galbraith, and De Glopper (2012), Galbraith and Baaijen (2019), and the collection of articles in Lindgren and Sullivan (2019).

The article by Barkaoui includes a thorough review of previous research into pauses during writing, and then reports the results of a study assessing the effect of task type (either an integrated or an independent writing task), language proficiency, and keyboarding skill on the characteristics of pauses (frequency and duration) at different locations within the text and at different points in time during writing. Pauses are defined using a threshold of two seconds, and hence reflect episodes during writing where the writer breaks off from text production to engage in some form of conscious reflection. As discussed in Barkaoui’s review, this is a common practice within the field (and is used in both the article by Barkaoui, 2019, and the article by Révész et al., 2019), but has advantages and disadvantages depending on the aim of the research. As a threshold, two seconds does help isolate moments during writing where the normal flow of writing is

interrupted, and hence where the writer has stepped back to engage in conscious reflection. However, in doing so, it focuses on these episodes alone and ignores the majority of transitions between units in the text. Baaijen et al. (2012), in their study of L1 (Dutch) writers, used mixture modeling to analyze the distribution of transition times between words, and found that well over 90% (varying for different writers) of the transitions between words took less than this amount of time. Furthermore, there was evidence for two different distributions below this threshold, one that Baaijen et al. suggested reflected word-retrieval processes, and another that reflected higher-level subsentence structuring operations. In selecting a threshold for the definition of a pause, therefore, the researcher is choosing to concentrate on relatively rare episodes of reflection, and to ignore data that might be informative about some of the more automatic processes involved in language production (see also Chenu, Pellegrino, Jisa, & Fayol, 2014). The incidence and location of these reflective episodes, nevertheless, remains valuable as Barkaoui's and Révész et al.'s (both 2019) findings indicate.

The most clear-cut finding is that lower language proficiency and keyboarding skill increase the number of reflective episodes and reduce the fluency of writing. This suggests that text production is generally more problematic for these groups, and that this is detectable in patterns of pause production during writing. That said, there was relatively little evidence that these two factors affected pause duration. This is probably because the analysis was restricted to pauses above the two-second threshold, thus reducing the sensitivity of the analysis. This may also account for the relative lack of difference in pausing behavior between linguistic locations. Typically, research has found clear-cut differences between the pause durations within words, between words, between sentences, and between paragraphs, with durations increasing with the size of the linguistic unit (e.g., Chanquoy, Foulin, & Fayol, 1996; Matsushashi, 1981; Medimorec & Risko, 2017; Phinney & Khouri, 1993; Schilperoord, 1996; Spelman Miller, 2000). Barkaoui also reports that pauses tended to be more frequent at word boundaries. It is important to recognize here that this is a consequence of the greater frequency of word boundaries compared to the boundaries between higher-level units such as sentences and paragraphs. Thus, 40 pauses at word boundaries for a text of 300 words corresponds to 13% of words being associated with pauses. By contrast, 2.5 pauses at paragraph boundaries for a five-paragraph text would correspond to 50% of paragraphs being associated with pauses. Looked at like this, Barkaoui's findings are consistent with higher-level units being associated with a higher proportion of reflective thought during writing. At lower levels, such episodes are more likely to reflect problems with the language production process; at higher levels, they are more likely to reflect global planning and revision operations (writers often use paragraph boundaries to review and edit previously produced text) (Baaijen et al., 2012; Chukharev-Hudilainen et al., 2019; Révész et al., 2019).

Overall, the article by Barkaoui demonstrates that task type, language proficiency, and keyboarding skill have systematic effects on the distribution of pauses above a two-second threshold during writing. Some of the difficulties with interpreting these effects illustrate the potential value of using transition times rather than threshold-defined pauses for analysis, and of complementing keystroke analysis with eye-tracking data. An interesting suggestion, in accordance with the findings of Révész, Kourtali, and Mazgutova

(2017), to be further explored in future research, is the observation that providing content support to L2 writers, as was the case in Barkaoui's study, can reduce processing load, particularly for cognitively more challenging tasks, allowing writers to devote more attention to linguistic encoding processes.

The article by Leijten, Van Waes, Schrijver, Bernolet, and Vangehuchten (2019) is similar to the article by Barkaoui in relying exclusively on keystroke logging for information about writing processes. In their case, however, they focus on writers' use of sources and assess how this varies as function of language of writing (L1 Dutch and L2 English) and at two different points in the academic year, and how it relates to the quality of the texts they write. In the course of this article they demonstrate a number of features of keystroke-logging methodology. First, they show how keystroke logging can provide at least some information about reading behavior. In particular, so long as the sources are electronically available, one can identify various features of how a writer interacts with sources during writing. This kind of measure would have helped confirm Barkaoui's plausible suggestion that the more extended pauses associated with the initial stages of the integrated task could reflect time spent reading and interacting with the source texts. Second, they present a case study showing how an individual writer moved back and forth, both within their text and between the developing text and the sources. In doing so, they show that keystroke logs can provide information, not just about the location and duration of episodes of reflective thought (pauses) but also about the linearity with which text is produced (cf. Baaijen & Galbraith, 2018).

Leijten et al. focus on the writers' interaction with sources, and the most innovative feature of their study is the use of confirmatory factor analysis to describe three components of this interaction: (i) how long the writer spends reading the sources during the initial phase of writing; (ii) how frequently the writer switches between different sources during writing; and (iii) the balance between reading of sources and writing text, with writers at one extreme focusing initially on reading sources and then on writing text, and writers at the other extreme referring back to sources more frequently throughout the course of writing text. They also confirm through measurement invariance analysis that the same factor structure can be used to describe writing in L1 and L2. That said, it is not entirely clear how the factor structure should be interpreted: there are correlations between the factors (in some cases larger than the factor loadings for individual items), and some of the measures look as if they might be alternative ways of measuring the same feature rather than separate indicators of an underlying construct. It is important to establish the reliability of the factor structure in future research, and to further explore alternative possible interpretations. Nevertheless, Leijten et al.'s results do strongly suggest that the extent to which writers switch between sources may be an indicator of variations in source use: writers do this more in their L2 than in L1, which may reflect something about reading in their second language, and writers who do this more in their L1 produce better quality text. There is a similar positive relationship with text quality in L2 but this is not statistically significant. It is important to note here that the lack of significance for L2 is most probably because of the smaller sample size for this group rather than because of any significant difference in effect size, so it remains an open question whether the relationship between switching across sources and text quality is general across L1 and L2.

Stepping back from the results of the study, the article as a whole indicates the potential that keystroke logging has for recording data from large samples of participants, and suggests ways in which it can provide valuable information even in the absence of direct observation of reading behavior through eye-tracking or information about the content of processes from retrospective verbal protocols. Other aspects that could be investigated include the impact on source use of linguistic proficiency in different L2s, the role of individual learner characteristics in relation to the choice of specific source-use strategies, and the linguistic comparison of source text and target text (in terms of lexical diversity and “text borrowing”).

The study, interestingly, also makes a number of suggestions for instructional practice. In the academic context, source-based writing forms a key element of academic writing proficiency and may be an index of successful academic achievement: a text that makes no use of any external sources would be fundamentally nonacademic. Appropriate source use, however, turns out to be difficult, especially for low-proficiency L2 writers, as writers have to balance between the demands of intertextuality and the risk of being accused of plagiarism. Leijten et al. thus propose that more attention needs to be paid to source-based writing in L2 (and L1) writing instruction, and suggest that this should include techniques for summarizing, paraphrasing, and text editing during larger writing assignments, such as essays, for which the combination of independent and integrated writing is necessary (see also Davis, 2013; McGinley, 1992).

The final article in this group, by Chukharev-Hudilainen et al., provides the first example in the special issue of a study that goes beyond collecting keystrokes to collect information about eye movements. This is a key step in that it offers the possibility of disambiguating the nature of pauses that occur during writing, and therefore of resolving some of the difficulties of interpretation that we have noted about pure keystroke-based measures. Note also that, unlike the other articles in this volume, pauses are defined simply in terms of the transition time between key presses (or interkeystroke intervals [IKIs]), rather than in terms of a particular threshold.

The first major contribution of this article is the presentation of CyWrite. To our knowledge, this is the first system that enables automated analysis of keystroke and eye-movement data collected under naturalistic conditions, and is likely to be an invaluable, low-cost tool for researchers in the field. Furthermore, the ready availability of playback and interactive visualizations of writing sessions provides many pedagogical possibilities. Chukharev-Hudilainen et al. then use CyWrite to assess differences between L1 and L2 writing processes by comparing the logs of 24 adult Turkish speakers writing in Turkish (L1) and English (L2). A second important innovation was the use of more linguistically sophisticated text boundaries, including clause boundaries, rather than just the overt orthographic boundaries indicated by punctuation that have been used in much previous research.

This study has two important findings. First, although pauses were generally longer at each text boundary (within-word, word-initial, clause-initial, sentence-initial) for writing in L2 compared to L1, this difference was not present for clause boundaries in L2, where the pauses were similar in duration to those preceding individual words. In other words, within sentences, words appeared to be produced more linearly in L2, without evidence of a hierarchical substructure. As the authors stress, this effect was not anticipated, and needs replication. Furthermore, the precise reasons why this might occur are open to

interpretation. Chukharev-Hudilainen et al. suggest that it may be because writers in L2 fail to plan subclauses in advance. We would suggest that it may be that the resources required for retrieving individual lexical items may reduce the resources available for syntactic planning: sentence production in L2 becomes a less hierarchically structured process and a more word-by-word production process.

Second, eye-movement analysis showed that pauses at higher-level units were not just longer in duration but more likely to involve looking back in the text, and over longer distances. When writing in L2, writers were, as predicted, more likely to look back than when writing in L1. However, there was an interesting interaction with language of writing: writers looked back within a sentence to a similar extent for both languages, but looked back to a different sentence more frequently when writing in L2. This pattern of results is consistent with the idea that the greater difficulty that writers have in producing language in L2 has a distinctive effect on their need to reconstitute the overall theme of their text (by reading over previous sentences). This suggests that problems with producing text have a negative effect on the higher-level processes involved in generating content. It would be interesting to assess directly whether the extent to which writers' pause times for lower-level units in L2 are elevated correlates with the frequency with which they read back over previous sentences. Chukharev-Hudilainen et al. suggest that the greater tendency in L2 to look back at previous sentences may be problematic because of its disruptive effects on fluency. An alternative possibility is that this is an important compensatory activity: writers in L2 need to do this to ensure that the current text is cohesive and coherent with previously produced text. Regardless of interpretation, however, these are potentially important empirical findings, arising from the study's incorporation of eye-tracking measures, and its coding of text in terms of linguistic units.

Finally, Chukharev-Hudilainen et al. note the possibility that the differences they observed could, in theory, be because of differences between the two languages (Turkish and English) rather than because one is L1 and the other is L2. This is an important methodological point: to be sure that a difference is an L1/L2 difference, one needs to demonstrate that the effect occurs when the mapping of the languages is reversed. If English was L1 and Turkish L2, would the same differences as Chukharev-Hudilainen et al. observed still be present, or would the differences reflect the language being used?

COMBINING DIRECT AND INDIRECT OBSERVATION

The final article that we discuss, by Révész et al., extends the methods used by including stimulated recall to provide further insight into the content of processes carried out during pauses measured by keystroke logging and eye movements. They extend the analyses of these to consider not just the processes occurring during pauses but also the revisions that writers carry out during writing in an L2. It is worth noting that, although this enabled them to capture a wider range of writing behavior, and they were able to access more information about the content through stimulated recall, the fact that a pause was defined in terms of a two-second threshold means that the analysis is relatively more coarse-grained than in Chukharev-Hudilainen et al.'s study. Furthermore, as in the Barkaoui article, little mention was made of the baseline frequency of occurrence of different text locations (individual keystrokes within words occur many times more frequently than sentence or paragraph boundaries do, making it difficult to compare raw frequencies of

events at different boundaries). Nevertheless, Révész et al.'s results are very similar to those of Chukharev-Hudilainen et al.'s study—we will not therefore reiterate them here—and support the hypothesis that different kinds of processing are associated with different text boundaries. Crucially, they also provide complementary evidence, through the use of stimulated recall, about the content of processing at different boundaries, and extend this, in their analysis of revisions, to the revision process as well as the planning process.

First, they suggest that pauses at different locations, though defined equivalently in each case as episodes of reflective thought taking longer than two seconds, appear to be employed for different reasons depending on the locations: pauses at the more local word level appear to be overwhelmingly concerned with the language processes involved in translating thought into words; pauses at the sentence level are more likely to be concerned with planning and, in some cases, revising content. The hypothetical explanation proposed by studies that use keystroke logging and/or eye movements alone is that the longer duration of pauses at sentence boundaries reflects greater amounts of content planning at these boundaries than at lower-level text boundaries. This receives direct support in Révész et al.'s study from the additional information provided by the stimulated recall protocols that they collected. Furthermore, the stimulated recall protocols also enable them to clarify the nature of the revisions made at different text boundaries. Thus, they find that within-sentence revisions tend to be more focused on revisions of language, whereas between-sentence revisions tend to be more focused on revision of content. Overall, however, at least in the L2 context of this study, the participants focused more on language-related revisions than on content revisions.

Finally, this study, along with the article by López-Serrano et al. (2019), demonstrates the value of verbal protocols in elucidating the content of writers' thoughts about even the relatively fine-grained processes involved in the formulation (or translation) process. Although such data have to be treated with caution because of the risk that they reflect post-hoc rationalizations, when combined with complementary information from more objective measures, they provide a valuable insight into the goals that drive the writing process at different text locations.

CONCLUSIONS AND FUTURE DIRECTIONS

Research into the online processes taking place during L2 writing and their cognitive interpretation is still very much in its infancy. As the articles in this special issue illustrate, it has tended to focus on observational studies of specific elements of the process—the duration of pauses at different locations and/or the nature and extent of revisions at different locations—and there is variation in how these elements are defined and analyzed. However, these articles also illustrate how the range of behaviors under analysis can be extended (Leijten et al.; López-Serrano et al.), how observations can be triangulated by combining complementary methods (Chukharev-Hudilainen et al.; Révész et al.), and how relevant independent variables (e.g., task type) can be manipulated (Barkaoui; Chukharev-Hudilainen et al.; Leijten et al.). Furthermore, the technical innovations described by Chukharev-Hudilainen et al., and the innovative schemes for analyzing protocols described by López-Serrano et al. and Révész et al.,

indicate that research in this area is moving to a stage where it can test specific hypotheses about L2 writing processes.

Perhaps the most general theme to emerge from these studies is the distinction between within- and between-sentence processes. Sentence boundaries appear to be an important hinge in the writing process, broadly separating global-planning processes from the more local processes involved in sentence production. A second noticeable feature is that, with the exception of López-Serrano et al., they have focused on isolated measures—pauses and revisions or uses of sources—rather than on how the different processes that they reflect operate in combination. Furthermore, with the exception of Leijten et al., they have not examined how writing process measures are associated with outcomes such as text quality or the development of the writer's understanding. The study by Baaijen and Galbraith (2018), which used keystroke logging to assess these relationships for writers writing under different planning conditions in L1, found systematic relationships between writing processes and outcome variables such as text quality and the development of the writer's understanding. However, the results also suggested that these may not be directly related to single indicators of writing processes, and that they may be moderated by variations in overall drafting strategy. An important question for future research is about the nature of such relationships when writing in an L2. The methods described in this special issue, and particularly the greater insights provided when keystroke logging is combined with eye-tracking and verbal protocols, promise to provide tools with which to explore these complexities.

Furthermore, because many of the articles observed large differences between individuals, future studies should take into account as moderating variables individual cognitive and motivational variables that have been found to influence writing (sub) processes and text quality. These variables include phonological short-term memory ability, age, writing engagement, writing attitude and writing beliefs, self-efficacy, and goal achievement. They manifest themselves in the tendency to develop complex goals for writing and, as a result, to develop language learning opportunities (Ortega, 2012). The necessity to examine in more detail the role of these individual factors has been emphasized in many studies, with the ultimate goal to help learners become autonomous writers, capable of self-monitoring for language errors and editing their own work. Reflexivity of L2 (and L1) writers toward their own texts, learner agency, and self-regulation, in a psychologically safe learning environment where learners *seek* for feedback instead of *receiving* it, should thus be encouraged (Segers, 2013).

In the present state of development, the online measures described in this issue show most promise for increasing our understanding of the cognitive processes and dynamic interaction between them in the course of writing, rather than as providing measures of direct utility for writing instruction. An important first step in developing their pedagogical applicability is to assess how the range of different measures identified in these studies relate to the quality of text in L2, both with respect to linguistic complexity, accuracy, and fluency (CAF; cf. Housen, Kuiken, & Vedder, 2012), and in terms of adequacy of content, register, and specific task requirements (Kuiken & Vedder, 2017, 2018). Second, it is unlikely that a single measure of writing processes will show a straightforward relationship with text quality. Rather, the relationship is likely to be with more general features of the writing process, like the relative balance between sentence production processes and higher-level planning processes, and how these

processes are coordinated. Given the importance of goals in determining the moment-by-moment actions that a writer takes, interventions that target the writer's goals, and their coordination in the face of conflicting demands are likely to be beneficial.

Finally, future research should also focus on the inclusion of L2 writers from different linguistic backgrounds and writing traditions to investigate (cross-)linguistic development of L2 writing ability over time. A related issue for further research concerns the effects of writing instruction on L2 learners' (meta)cognitive writing processes, particularly with respect to the extent to which these processes can be influenced, and adjusted and writing problems can be remediated by pedagogical intervention.

REFERENCES

- Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction, 36*, 199–223. <https://doi.org/10.1080/07370008.2018.1456431>.
- Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication, 29*, 246–277. doi: 10.1177/0741088312451108.
- Baaijen, V. M., Galbraith, D., & de Glopper, K. (2014). Effects of writing beliefs and planning on writing performance. *Learning and Instruction, 33*, 81–91. <http://dx.doi.org/10.1016/j.learninstruc.2014.04.001>.
- Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes? *Studies in Second Language Acquisition, 41*, 529–554.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.
- Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Eds.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361–392). Hillsdale, NJ: Erlbaum.
- Chanquoy, L., Foulon, J. N., & Fayol, M. (1996). Writing in adults: A real time approach. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 36–43). Amsterdam, The Netherlands: Amsterdam University Press.
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80–98. doi:10.1177/0741088301018001004.
- Chenu, F., Pellegrino, F., Jisa, H., & Fayol, M. (2014). Interword and intraword pause threshold in writing. *Frontiers in Psychology, 5*, 1–7. doi: 10.3389/fpsyg.2014.00182.
- Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., & Feng, H.-H. (2019). Combined deployable keystroke logging and eye-tracking for investigating L2 writing fluency. *Studies in Second Language Acquisition, 41*, 583–604.
- Davis, M. (2013). The development of source use by international postgraduate students. *Journal of English for Academic Purposes, 12*, 125–135. <http://dx.doi.org/10.1016/j.jeap.2012.11.008>.
- Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 31–50). Hillsdale, NJ: Erlbaum.
- Galbraith, D., & Baaijen, V. M. (2018). The work of writing: Raiding the inarticulate. *Educational Psychologist, 53*, 238–257. <https://doi.org/10.1080/00461520.2018.1505515>.
- Galbraith, D., & Baaijen, V. M. (2019). Aligning keystroke with cognitive processes in writing. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 306–325). Leiden, The Netherlands: Brill Publishing. https://doi.org/10.1163/9789004392526_015.
- Hayes, J. R. (2012a). Modeling and remodeling writing. *Written Communication, 29*, 369–388. <https://doi.org/10.1177/0741088312451260>.
- Hayes, J. R. (2012b). From idea to text. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 65–79). London, UK: Sage Publications Ltd.
- Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication, 23*, 135–149. doi:10.1177/0741088306286283.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3–30). Hillsdale, NJ: Erlbaum.
- Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist, 41*, 1106–1113. <http://dx.doi.org/10.1037/0003-066X.41.10.1106>.

- Housen, A., Kuiken, F., & Vedder, I. (Eds.) (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam, The Netherlands: John Benjamins.
- Kaufner, D. S., Hayes, J. R., & Flower, L. S. (1986). Composing written sentences. *Research in the Teaching of English*, 20, 121–140.
- Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 355–365. doi: 10.1037/0278-7393.14.2.355.
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing* (pp. 57–71). Mahwah, NJ: Erlbaum.
- Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research*, 5, 159–190. doi:10.17239/jowr-2013.05.02.1
- Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing*, 34, 321–336. doi: 10.1177/0265532216663991.
- Kuiken, F., & Vedder, I. (2018). Assessing functional adequacy of L2 performance in a task-based approach. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics* (pp. 266–285). Amsterdam, The Netherlands: John Benjamins.
- Leijten, M., Van Waes, L., Schrijver, I., Berolet, S., & Vangehuchten, L. (2019). Mapping master's students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition*, 41, 555–582.
- Lindgren, E., & Sullivan, K. P. H. (2019). *Observing writing: Insights from keystroke logging and handwriting*. Leiden, Netherlands: Brill Publishing. doi: <https://doi.org/10.1163/9789004392526>.
- López-Serrano, S., Roca de Larios, J., & Manchón, R. M. (2019). Language reflection fostered by individual L2 writing tasks. *Studies in Second Language Acquisition*, 41, 503–527.
- Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English*, 15, 113–134.
- McGinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, 27, 227–248. doi: 10.2307/747793.
- Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing*, 30, 1267–1285. doi: 10.1007/s11145-017-9723-7.
- Olive, T. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research*, 6, 141–171. doi: 10.17239/jowr-2014.06.02.4.
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). Berlin, Germany: Mouton de Gruyter.
- Phinney, M., & Khouri, S. (1993). Computers, revision and ESL writers: The role of experience. *Journal of Second Language Writing*, 2, 257–277. doi: 10.1016/1060-3743(93)90021-t.
- Révész, A., & Michel, M. (2019). Introduction. *Studies in Second Language Acquisition*, 41, 491–501.
- Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67, 208–241. doi:10.1111/lang.12205.
- 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, 605–631.
- Roca de Larios, J., Nicolás-Conesa, F., & Coyle, Y. (2018). Focus on writers: Process and strategies. In R. M. Manchón, & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 267–286). Boston, MA: Walter de Gruyter.
- Schilperoord, J. (1996). The distribution of pause time in written text production. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 21–35). Amsterdam, The Netherlands: Amsterdam University Press.
- Segers, M. S. R. (2013). Understanding informal feedback in the workplace: The impact of the position in the organizational hierarchy. *European Journal of Training and Development* 37, 72–85. doi.org/10.1108/03090591311293293.
- Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, 4, 123–148. doi: 10.1111/lang.12205.
- White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, 30, 166–189. <http://dx.doi.org/10.1016/j.cedpsych.2004.07.002>.