

# Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing

Jim Ranalli  
Hui-Hsien Feng  
Evgeny Chukharev-Hudilainen

## ABSTRACT

Assessment for learning (AfL) seeks to support instruction by providing information about students' current state of learning, the desired end state of learning, and ways to close the gap. AfL of second-language (L2) writing faces challenges insofar as feedback from instructors tends to focus on written products while neglecting most of the processes that gave rise to them, such as planning, formulation, and evaluation. Meanwhile, researchers studying writing processes have been using keystroke logging (KL) and eye-tracking (ET) to analyze and visualize process engagement. This study explores whether such technologies can support more meaningful AfL of L2 writing. Two Chinese L1 students studying at a U.S. university who served as case studies completed a series of argumentative writing tasks while a KL-ET system traced their processes and then produced visualizations that were used for individualized tutoring. Data sources included the visualizations, tutoring-session transcripts, the participants' assessed final essays, and written reflections. Findings showed the technologies, in combination with the assessment dialogues they facilitated, made it possible to (1) position the participants in relation to developmental models of writing; (2) identify and address problems with planning, formulation, and revision; and (3) reveal deep-seated motivational issues that constrained the participants' learning.

*Keywords:* assessment for learning; writing processes; L2 writing; keystroke logging; eye tracking; process tracing

## 1. Introduction

Assessment for learning (AfL) is assessment meant to support learning and teaching as opposed to other types that support sorting, certifying, or accountability functions (William, 2011). While AfL has been a central focus of curriculum reform in several countries for more than a decade,

AfL research in L2 writing remains scarce (Lee, 2017). Part of the problem may have to do with the capacity for AfL, and L2 writing assessment more generally, to take account of writing processes, such as planning, formulating, and revising.

Research has shown major differences between the ways skilled and unskilled writers engage in these processes (Roca de Larios, Murphy, & Marín, 2002) as well as connections between patterns of process engagement and writing quality. In some studies, nearly 80% of the variance in writing quality was explained by the type and timing of the processes writers engaged in (Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Rijlaarsdam & Van den Bergh, 2006). And yet, most L2 writing instructors probably know very little about the way their students go about producing the assignments they submit as part of classroom assessment: how much time they spent, for example, and whether and how they engaged in specific processes. Even if instructors adopt a “process” approach and assign multiple drafts, these drafts are still written products that bear little information about the specific processes that went into their creation.

Meanwhile, the tools that researchers have used to study writing processes have been increasing in power and sophistication. They now include technologies that are largely unobtrusive and increasingly accessible and affordable such that they could be scaled up for use in instructional settings. The present study investigated whether and how such process-tracing technologies might support more meaningful AfL of L2 writing. The study contributes to the existing research on L2 writing assessment by investigating whether new technologies can expand the focus and potential benefits of AfL while also addressing calls for assessments that capture more of the writing construct (Cumming, 2002; Deane, 2013). In line with the theme of this special issue of *Assessing Writing*, it also demonstrates unique and significant advantages of computer-based over paper-based writing.

## **2. Literature review**

### *2.1 Assessment for learning and feedback in L2 writing*

Assessment for learning has been defined as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where learners are in their learning, where they

need to go, and how best to get there” (Broadfoot et al., 2002, pp. 2-3). While AfL has become an important component of curriculum reform in Australia, Hong Kong, the U.K., and the U.S., it has so far inspired little research in the area of L2 writing (Lee, 2017). Such research as exists has been based in secondary and tertiary settings in EFL contexts in Hong Kong and Taiwan. Many of these studies have focused on teachers’ motivations and strategies for implementing AfL and factors that facilitated or constrained AfL initiatives (Lee, 2011; Lee & Coniam, 2013; Lee & Falvey, 2014; Mak & Lee, 2014). Another line of research has examined the influence of AfL on student writing outcomes and student attitudes about AfL innovations (Huang, 2012, 2016; Lee, 2011) such as providing indirect feedback to help learners develop more independence in error correction.

Feedback is central to AfL because it is the means to convey information about where students’ abilities currently lie in relation to their goals and about the ways they can progress toward those goals (Black & Wiliam, 1998). Hattie and Timperley’s (2007) influential work on feedback, often cited in the general AfL literature, posits four types: feedback about the student him or herself (FS), such as praise; feedback about the task (FT), such as correctness or alignment with a rubric; feedback about the processing of the task (FP), such as the type of behavior needed to make improvements; and feedback about self-regulation (FR), such as information to support self-evaluation. Research shows that, whereas FS is the least effective, FP and FR can contribute powerfully to deep processing and mastery of tasks. FT, which is the most common type of feedback, is powerful when used in conjunction with FP and FR, although in practice this rarely happens (Hattie & Timperley, 2007).

Much of the feedback provided in conventional L2 writing classrooms can be described in terms of FT. Written corrective feedback (WCF) addressing linguistic issues may support focus on form (Doughty, 2001) and thus promote learning beyond the current task (although the WCF research on this question is mixed; see Bitchener & Ferris, 2012). The other common types of feedback on writing—namely, instructors’ written or oral comments about content, organization, and other higher-level concerns—may only serve to improve the current text; that is, it may be too task-specific to generalize to future writing and thus less supportive of learning. FP and FR, meanwhile, are difficult for writing instructors to address because of the lack of information

about either of these dimensions of students' task engagement. However, information about students' task processing could be gathered by using some of the same methods employed by writing process researchers (described below).

In AfL, it is vital that assessment information be communicated to students in ways that help them clarify goals and understand evaluative criteria (Chong, 2017). AfL is informed by motivation theory that emphasizes learning (i.e., mastery) goals as opposed to performance goals, and when conducted appropriately, AfL is seen to both support and be supported by learner motivation (Lee, 2017). One way to ensure that feedback is conveyed in ways that students can understand and make use of is for instructors to engage in individual “assessment dialogues” with students (Carless, 2006), which can be mutually beneficial insofar as they may also help instructors synthesize and interpret assessment information (Chong, 2017).

## *2.2 Processes in writing theory and research*

Processes are foundational to the major cognitive models of writing (Bereiter & Scardamalia, 1987; Flower & Hayes, 1981; Hayes, 2012; Kellogg, 2008; Leijten, Van Waes, Schriver, & Hayes, 2014), including the pair of models proposed by Bereiter and Scardamalia (1987) to account for differences between novice and skilled writers, which, although based in L1 research, has informed theorizing about L2 writing (Weigle, 2002). In this distinction, the *knowledge-telling* model describes the novice approach, wherein writing is a simple act of retrieving information from memory and “telling” what you know. By contrast, skilled writers exhibit *knowledge-transforming*, which involves an interaction between the author's mental representation of her ideas and a separate mental representation of the text, with discrepancies between the two occasioning problem-solving and rethinking of the original ideas. (Kellogg [2008] expanded this dichotomous conceptualization by adding a third, top-level model, *knowledge-crafting*, which characterizes the work of professional writers.) According to Bereiter and Scardamalia (1987), it is not possible to identify the underlying approach simply by studying a given text because knowledge of topic, genre, and language affects writing outcomes. Rather, it is cognitive processes and the different ways they are enacted that make such differentiation possible. These processes include:

### 2.2.1 Planning

Planning typically encompasses generating ideas and organizing them into a structure. Unskilled writers tend to forego planning whereas skilled writers use advanced (i.e., before writing) or emergent (i.e., while writing) forms of planning or both (Cumming, 1989; Victori, 1999).

Advanced planning has been found to ease the attentional burden on unskilled writers (Kellogg, 1988), and in a study involving timed writing tests, it corresponded with higher writing quality (Worden, 2009).

### 2.2.2 Formulation

Also referred to as *translation*, this process involves the conversion of ideas into language. In L2 writing and among novice writers, this is typically the process to which the most time is devoted, but as writing skills develop, other skills such as planning and revision increase in proportion (Roca de Larios et al., 2008). Writing-skills development is associated with an increase in the interruption of formulation by other processes (e.g., planning, evaluating, revising), which can be interpreted as recursiveness indicative of problem-solving (Roca de Larios, Marín, & Murphy, 2001). This should be distinguished from the simultaneous enactment of formulation and planning (or other processes), which in novices can easily lead to attentional overload (Kellogg, 1988). Some research shows writers process both form and concepts more when formulating in an L2 than when writing in L1 because of the increased cognitive effort from working in a language that is not yet automatized (Lindgren & Sullivan, 2006).

### 2.2.3 Evaluation

Evaluation involves rereading the text produced so far to determine whether it needs modification. The process is understudied compared to its companion process, revision. However, eye-tracking can facilitate greater research focus on this process (Wengelin et al., 2009). Evaluation is worth considering in its own right since positive evaluations that do not result in revisions may still be of interest, and since revisions may proceed from different sources of evaluation (oneself, instructors, peers, or automated analyses). In addition, evaluations that determine a need for major changes may give rise to redrafting (i.e., formulating parts of the text anew) as opposed to revision, which in such cases is more cognitively demanding (Flower et al., 1986). In a study comparing distributions of writing processes across test and non-test

conditions, time allocated to evaluation in the latter stages of writing was found to constitute the only significant difference, with twice as much evaluation observed in the non-test condition (Khuder & Harwood, 2015).

#### 2.2.4 *Revision*

This well-researched process involves modifying previously written text. Research shows unskilled writers in general focus more on local revision (i.e., changes at sentence level) whereas skilled writers engage in both local and global revision (i.e., changes involving larger units of discourse). Among L2 writers, aversion to global revision has been attributed to fear of confronting linguistic issues they cannot remediate (Uzawa, 1996). While unskilled L2 writers may engage in local revision randomly, skilled L2 writers tend to leave local revision until later (Zamel, 1983). Writers working in an L2 have been found to make more pre-contextual revisions (i.e., revisions at the point of inscription) involving both form and concept than when writing in their L1 (Lindgren & Sullivan, 2006).

#### 2.2.5 *Task definition*

Task definition, also known as *task representation*, is the process whereby a writer forms and refines her understanding of a task's requirements. This process can be conceptualized at multiple levels, from a particular assignment to a genre to a mental model of academic writing (Nicolás-Conesa, 2012). A study of EFL student writers at the beginning and end of a nine-month writing program found that they maintained conceptualizations that were largely product-based, although some by the end showed evidence of a process orientation as well (Nicolás-Conesa, Roca de Larios, & Coyle, 2014).

One important underlying theme in L2 writing process research is the greater complexity that writing tasks take on because attentional capacity must be devoted to language processing that could otherwise be devoted to higher-level concerns, such as content and rhetoric (Roca de Larios et al., 2002; Spelman Miller, Lindgren, & Sullivan, 2008). As the above research summary indicates, student writers have been observed to engage in the same processes differently depending on whether they are writing in a first (i.e., automatized) or second (i.e., not yet automatized) language. Yet L2 writers have also been shown to exercise agency, reusing and

reshaping their L1 knowledge and experience differently depending on the task and audience, albeit with these choices being influenced by L2 proficiency and amount of L1/L2 writing experience (Rinnert, Kobayashi, & Katayama, 2015). This suggests efforts to develop interventions aimed at L2 writing processes must (1) prioritize helping L2 student writers manage cognitive load; and (2) adopt an individualized approach based on assessing the needs of particular students, which is consistent with AfL theory.

Finally, we note that the process research summarized above has been accomplished over time using a variety of methods. Older methods included direct observation, observation via videotaping, and thinkaloud protocols (which are still frequently used) while newer methods include keystroke logging and eye-tracking. Whereas the former are clearly impractical for classroom applications, the same cannot be said for the latter, hence the motivation for the present study.

### *2.3 Process tracing in writing research*

As the preceding section suggests, relatively little research in the field of writing assessment has investigated cognitive processes. This can be attributed in part to the historically predominant form in the field: the timed, impromptu writing test, in which recursiveness, global revision, and other important aspects of writing-as-process can play little part (Wolcott, 1987). Nevertheless, there have been repeated calls for assessments that can measure processes (Cho, 2003; Deane, 2013; Graham, Harris, & Hebert, 2011; O'Brien, 1992; Wolcott, 1987) and thus expand the coverage of the construct of writing ability (Deane, 2013). To achieve this, recent, large-scale standardized assessments have experimented with keystroke logging (Almond et al., 2012; National Assessment of Educational Progress, 2011).

At the same time, writing process (WP) researchers have been using different types of visualization to analyze keystroke logging (and in some cases eye-tracking) data and report their findings, which is necessary because the huge data sets that these technologies produce make it impossible for unaided human faculties to identify patterns and regularities (Bécotte et al., 2015). Despite their varied forms, WP visualizations typically convey information about the timing and location of writing behaviors of interest. They include written text marked up with special

symbols, known as S-notation (Severinson Eklundh & Kollberg, 1996); timelines combining keystroke and eye-tracking information (Wengelin et al., 2009), progression diagrams showing number of revisions as a function of both timing and location (Perrin, 2003); dynamic, GIS-based representations (Lindgren et al., 2007); and graph visualizations resembling network diagrams (Caporossi & Leblay, 2011). One issue in the design of WP visualizations is how to achieve sufficient amounts of macro- and micro-level detail so as to support analysis of larger trends (e.g., to see how much time was allocated to formulation versus revision) while also allowing text-level views (e.g., to see what type of revision was made at a particular point in time). The solution used in the current study was to dynamically link a linear graph inspired by those in Lindgren and Sullivan (2002) and Leijten and Van Waes (2013) with a replay visualization of character-by-character reconstruction of texts like those described in Severinson Eklundh and Kollberg (1996) and Leijten and Van Waes (2013). (See description in Methods.)

To date, it appears only a single line of research has used process-tracing technologies to provide formative feedback on L2 writing. Lindgren and Sullivan used keystroke logs to generate a visual playback of the entire production of student texts, which served as stimuli for self- and peer assessment of composing behavior. In two small-scale studies (Lindgren, Stevenson, & Sullivan, 2008; Sullivan & Lindgren, 2002), participants who watched replays of their own and a peer's composing reported gaining unique insights into their own writing habits, such as a tendency to repeatedly revise a certain passage and, as a result, lose the thread of the larger argument. No follow-up was conducted, however, to see if these insights resulted in changes to the behaviors identified by the participants. Also, 30-minute writing tasks were used, which made it feasible to derive insights from a linear viewing of the replay; this might not be the case with longer replays of untimed writing.

#### *2.4 The current study*

The following research question guided the study: Do process-tracing technologies provide insights about L2 student writing that can enhance instructors' efforts to create more meaningful assessment for learning? We focus on the role of instructors because student involvement in AfL, sometimes referred to as Assessment as Learning (AaL), requires attention to metacognitive and self-regulatory abilities that went beyond the scope of this exploratory study (although AaL



would be a next logical step). In addressing this research question, an underlying aim was to produce information that could inform future attempts to scale up this approach were it to prove promising.

### **3. Methods**

#### *3.1 Context*

The study took place at a large research university in the Midwestern U.S. Because of constraints imposed by the institutional review board, the research could not be integrated into ongoing writing classes, so we recruited students from these classes and had them complete writing tasks outside of their normal coursework. Specifically, we recruited from two sequenced, developmental writing courses to which international students are assigned on the basis of an English placement test. The lower-level course focuses on academic writing at the sentence and paragraph levels while the higher-level course assigns essay-length tasks.

#### *3.2 Participants*

Of 14 participants recruited for the study, six completed all four writing tasks. From those six, we selected two using purposive sampling; specifically, sequential sampling to identify confirming and disconfirming cases (Teddlie & Tashakkori, 2008). Both participants were Mandarin L1 speakers from mainland China. Other biodata is summarized in Table 1. The participants were paid \$15 an hour.

[INSERT TABLE 1 ABOUT HERE]

#### *3.3 Writing and process-tracing tools*

Writing and process tracing were conducted in a web-based tool called CyWrite, which was developed by the authors. The CyWrite system features a text editor that provides a familiar word-processing experience while also permitting capturing of the process of composition with combined keystroke logging and eye tracking. As the user composes text, the CyWrite editor unobtrusively records time-aligned logs of keystrokes, text changes, and eye fixations. Eye

tracking is performed by a low-cost device mounted under the computer screen. The editor interfaces with the eye tracker via a protocol that provides a real-time feed of eye-fixation coordinates.

The three logs are streamed live to a server where they are analyzed and persistently stored. The logged events are then rendered in a post-session viewer (Figure 1), in which user activity is reconstructed in a visualization called *playback* that resembles high-fidelity screen-capture recordings with an overlaid gaze-point marker. Above the playback area, another visualization called the *process graph* is generated, containing variables measured in characters on the Y-scale and time measured in minutes on the X-scale. The plotted variables are: total number of typed characters, including deleted ones (“process” in blue); total length of text (“product” in green); offset in text of the character rendered in the top-left corner of the viewport (“scrolling” in pink); offset in text of the fixated character (“fixation” in yellow); and offset in text of the cursor position (“cursor” in red). Gaps in the plotted lines indicate periods when the writer has switched focus to another window in the operating system, such as an online dictionary page, or when the eye tracker was being recalibrated. The process graph and playback features are dynamically linked such that moving the playhead to a point in the graph will show what was happening in playback at that moment.

[INSERT FIGURE 1 ABOUT HERE]

Figure 1: Post-session viewer showing process graph above and playback below

### *3.4 Materials and measures*

Four writing tasks, adapted from prompts created by the Educational Testing Service, were assigned. We chose argumentative tasks since these elicit more complex interplays of processes (Roca de Larios et al., 2002) compared to narrative prompts, for example. The topics, which addressed generation-gap disputes over traditions, controversial trends and fads, computers and privilege, and the value of gap years, are provided in Appendix A. Each task required participants to write an essay of at least 400 words for a specified audience (faculty and students at the site university) and purpose (to persuade readers to “see the issue the same way as you”).

The participants were encouraged to take as much time and as many writing sessions as needed to complete it to their satisfaction. So that we could capture as much information about their writing processes as possible, the participants were told to do all of their planning, formulating, and revising in the CyWrite files prepared for them. In addition, the prompt for each task was displayed in copyable but uneditable text at the top of the CyWrite file so that participants' interactions with the prompt could also be captured.

Along with the essays, the participants also wrote four reflection tasks (Appendix B), two at the beginning of the project and two at the end, which were adapted from Nicolás-Conesa et al. (2014). One pair of these reflections probed students' mental models of academic writing while a second pair probed a variable unrelated to the current study, hence its exclusion here.

### *3.5 Analyses*

In addition to the automated analyses that produced the process visualizations, several other analyses were performed automatically by the CyWrite system. Summary statistics were calculated for time spent writing per session, number of characters typed and deleted, and word counts.

The completed essays were exported from CyWrite into MS Word format and then evaluated using the Comment and Track Changes functions by the first author, who also acted as writing instructor. The evaluated essays, transcripts of the follow-up sessions, the pre- and post-project reflections, and the sheets on which the participants' process goals were assigned, were imported into the qualitative software nVivo for coding and analysis. The process-trace data was screen-recorded, segmented, and coded using TechSmith Morae, a usability research tool.

Four separate coding schemes were used for the analysis of these different forms of data (Appendix C). The reflection-data codes were adapted from Nicolás-Conesa et al. (2014). The coding schemes for the follow-up session transcripts and process-viewer data were developed using a combined top-down and bottom-up approach, based on categories in Roca de Larios et al. (2008) as well as new categories identified during repeated cycles of reviewing the data with reference to the research question and previous literature. The coding scheme for the evaluated

essays was based on assessment features from the International English Language Testing System (IELTS) rubrics<sup>1</sup> and the first and second authors' combined experience in ESL writing instruction.

For all but one of the data types, some data was reserved for calibration and final refinements of the coding scheme; the remaining data were then annotated by a second coder. The total number of annotations, the percentage of annotations performed by the second coder, and reliability coefficients, are provided in Table 2. Reliability was excellent across the data sources based on the scale interpretations in Strijbos and Stahl (2007).

[INSERT TABLE 2 ABOUT HERE]

### *3.6 Procedures*

The study began in the second half of the Fall 2016 semester and continued through Spring 2017. Participants started by completing the questionnaire and initial reflection tasks. They then wrote their first writing task in a computer lab on machines equipped with the CyWrite software and eye trackers. With the help of project staff, they scheduled their writing sessions on days and times that were convenient for them. Once they completed a task, a tutoring session with the first author—referred to as a “follow-up session” to reflect its formative-assessment focus—would be scheduled.

In the follow-up sessions, which lasted 60-90 minutes, the participant and instructor would first discuss the finished essay, as would happen in a typical writing conference. The discussion would then turn to writing processes, with the participant and instructor reviewing the process

---

<sup>1</sup> Each essay was also rated using the public version of the IELTS Task 2 writing band descriptors ([http://takeielts.britishcouncil.org/sites/default/files/IELTS\\_task\\_2\\_Writing\\_band\\_descriptors.pdf](http://takeielts.britishcouncil.org/sites/default/files/IELTS_task_2_Writing_band_descriptors.pdf)). This analytic rating scale was used in place of the TOEFL holistic scale selected because its four categories (*task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy*) provide more formative information to learners and because IELTS Task 2 prompts resemble TOEFL prompts in their argumentative focus and in the types of response they elicit. We attempted to validate the writing quality ratings assigned by the first author as part of feedback, but variation in writing quality among the essays was below the sensitivity threshold of the instrument. On a scale of 0-9, the average rating was 5.43, with a standard deviation of only 0.71. Further, when the first and the second authors independently rated the essays, they only achieved reliability of  $\alpha_{\text{interval}} = 0.27$  despite appropriate calibration. Thus, although the quality ratings were useful for providing feedback to participants, they were not reliable enough to incorporate into the analyses reported here.

data. (The first follow-up session was spent familiarizing the participant with the process graph and playback capabilities, the writing processes, and practicing graph interpretation). Depending on the outcome of the product and process reviews, the session might then include (1) observation of process data for experienced “model” writers who had completed the same writing task, (2) instruction in strategies for approaching a particular process or decoupling processes, or (3) both. The sessions would then end with the instructor assigning process goals for the next writing task that addressed the major problem areas pointed out during the session. A sheet summarizing these goals would be given to the participant to have in front of her whenever she was working on the next writing task.

Upon completing the final writing task and follow-up session, participants wrote their final reflections and completed an exit survey.

## **4. Results**

A summary of the participants’ process and product information across the four writing tasks is provided in Table 3. This data is also referenced in the individual narratives that follow.

[INSERT TABLE 3 ABOUT HERE]

In these narratives, attributions are made to specific data sources where relevant by means of codes (e.g., *FSI* = the first follow-up session, *RT2* = the second reflection task, *WT3* = the third writing task).

### *4.1 Zedong: Development constrained by limited motivation*

Zedong came to the project with little academic writing experience, claiming never to have written an essay of more than 300 words in English, and this only for exam-preparation purposes. He described his primary motivation for participating as financial although he said he also wished to improve his writing. A freshman studying computer science and assigned to the lower-level writing course, Zedong was living and studying overseas for the first time. He said he hoped to work and study further in the U.S., but he believed writing in English would not be essential for his future career. The only form of writing he enjoyed was computer coding; other

types were “not that interesting” (FS4). He had scored a 6.5 overall on the IELTS test, which translates to a TOEFL score of between 79-93, and which was one step above the minimum score needed to enter the university.

His essay in response to WT1 consisted of a single-paragraph of almost the exact word length specified in the assignment. In it, Zedong had strayed from the prompt. Rather than arguing whether older people were justified in getting upset when younger people flouted traditions, the text focused on negative consequences of an annual spring festival in his hometown (e.g., pollution, traffic). The composition was also critiqued for non-sequiturs, uneven support for main ideas, and a statement at the end that seemed to contradict the thesis he had laid out at the beginning.

Analysis of the process data showed Zedong had completed the task in a single session lasting about 90 minutes (Figure 2), despite encouragement to take as long as he needed and to spread the work over at least two sessions. He had begun formulating after less than four minutes of reading the prompt and in the absence of any external plan. He said he had come up with a few ideas with which to get started and had expected more to occur to him as he wrote (FS1).

His need to generate ideas while formulating contributed to a fractured and slow overall process. Figure 2 shows a shallow upward slope of the green product line, as well as numerous gaps indicating Google and dictionary searches to find words to express his ideas or to check spelling. Several instances of “flatlining” (i.e., stretches with no slope and no apparent keyboarding or re-reading) are also evident, during which Zedong said he was “just thinking about what to write next” (FS1).

[INSERT FIGURE 2 ABOUT HERE]

Figure 2. Process-graph from Zedong’s first and only writing session in response to WT1.

In terms of revision, the process data showed only a brief (less than one minute) stage near the end, during which he made two changes to the second sentence of the text: adding a second

headword to create a compound noun phrase and substituting the word “disappointed” for “sad” to describe the feeling of losing one’s traditions. No rereading of the entire text was evident. Asked to reconcile this with his project-initial reflection, in which he stated the importance of revision and re-reading “to see is the logic make sense to the readers” (RT1), he laughed and admitted: “I didn’t revision so much. Just I don’t have the habit to. I don’t like to revise” (FS1).

Our goals for Zedong thus focused on helping him transition from a linear, knowledge-telling approach to a more recursive, knowledge-transforming approach including a fuller complement of processes. Zedong was told to include an external planning stage at the beginning, rereading of the prompt at regular intervals to ensure he stayed on track, and at least 20 minutes of evaluation and revision at the end, with a global focus preceding a local focus. We also suggested a “placeholder strategy” in which he was to write an L1 word, or his best approximation of the spelling of an L2 word he was unsure about so that he could continue formulating without losing his train of thought. These items, annotated with a symbol such as “##” to make finding them again easier, could be looked up after formulation during a dedicated resourcing or revision stage. One rationale for these goals was to allow Zedong to compartmentalize processes so that he was doing only one thing at a time and thus avoiding attentional overload.

In terms of planning, he progressed from simple plans consisting of a few main points (WT2) to a longer outline that broke down his argument into "benefits" and "deficits" (WT4). His planning technique became more interactive, but he also had a habit of dismantling his plan as he wrote, cutting and pasting portions into the text proper to convert to parts of the essay so that the complete plan was no longer available for purposes of further refinement or comparison during evaluation. Despite this, he also showed increasing propensity to compartmentalize formulation, planning, and resourcing (with the help of the placeholder strategy, which he used a great deal), facilitating a more productive overall process with less flatlining.

Evaluation and revision were more challenging for Zedong. For WT2, he had to be encouraged by project staff to return for a second session in which he could evaluate and review his work.

This session (Figure 3) lasted 19 minutes. The cursor line indicates two passes through the text, which playback showed to consist of local changes.

[INSERT FIGURE 3 ABOUT HERE]

Figure 3. Process graph from the second of Zedong's two writing sessions for WT2.

To address this, we showed him process data of a model writer engaging in global evaluation and discussed how the changes benefited the final product. We also provided specific questions to focus his global evaluation on issues of task response, content, organization, and audience awareness (Appendix D). Finally, we emphasized that writing is time-consuming even for skilled writers working in their mother tongue and that spending more time would be a gainful strategy. However, Zedong's process data for WT3 and WT4 showed similarly abbreviated evaluation and revision stages with predominantly local changes in evidence. Asked about this in the follow-up sessions, Zedong said he was avoiding global evaluation of a complete draft because he feared the amount of work required if he were to find major problems. He preferred instead to focus on smaller chunks of text because these were "flexible [enough] to make changes" (FS2).

The lack of global-level evaluation and revision was particularly reflected in the critiques of coherence problems on Zedong's submission for WT4. Although he was better addressing the prompt and generating better ideas, the way he organized and connected them in the essay still fell short, and dealing with these problems required scrutinizing his work at a global level, which he expressed unwillingness to do.

In short, while he made some strides, Zedong's development seemed constrained by limited motivation. Despite being compensated on an hourly basis, he never spent more than two sessions, or much more than two hours, on any writing task. He seemed willing to take on board strategies that created efficiencies and made writing easier (e.g., planning), but those addressing writing quality that might entail substantial effort (i.e., global evaluation and revision) lost in his cost-benefit analysis. These seemed likely to be a significant challenge for him going forward. That being said, by the end of the project, he was using the process terms (e.g., *formulation*,



*global evaluation*) appropriately to discuss his writing and showed some capacity to interpret his graphs and critique his process engagement.

#### 4.2 Mingyu: Overplanning, under-evaluating, and tending to go off topic

Mingyu was also a freshman on her first international study experience who was majoring in Computer Engineering. She was more proficient in English than Zedong, having scored a 98 on the TOEFL exam—well above the 71 needed for entrance to the university—and she had placed into the higher-level writing course. She expressed uncertainty about her future plans beyond “coding every day in front of the computer” (FS4). She believed training in academic writing was important insofar as it developed “logical thinking” (FS4), but beyond this, she did not anticipate a need for such writing in her future. She said she found it difficult to write according to prescribed topics, but “when I just write whatever I want, that seems fine to me” (FS4).

Mingyu’s submission for WT1 was a five-paragraph essay consisting of 642 words. While it was well organized, the essay was critiqued for having strayed from the prompt. Rather than arguing whether or not adults were justified in feeling upset when young people flouted tradition, she wrote a descriptive piece focusing on two psychological concepts that could explain people’s attachment to traditions: “Consistency theory” and the “‘Knew it all along’ effect” (WT1). Most of the essay was devoted to illustrating these ideas.

Analysis of the process data for this task showed Mingyu had completed it in a single session totaling 88 minutes (two recordings with a small break in between), nearly the same amount of time spent by Zedong. Unlike Zedong, however, Mingyu had begun by creating an external plan consisting of ideas for major points indexed to numbered paragraphs in the essay. She started formulating at the 13-minute mark and showed some ability to compartmentalize processes by shifting between the evolving text below and her plan above, which she would update as new ideas occurred to her and which she used to replenish her ideas at junctures in the formulation of paragraphs.

There were, however, only two instances where she returned to the prompt at the top of the file; one of these (Figure 4 at the 45-minute mark) shows her reviewing it before an *ex-post facto*

attempt to connect her description of “Consistency theory” back to the topic of the assignment. In addition, the process data depicted a final evaluation and revision stage of only nine minutes, during which she made mostly local-level changes (substitutions for word choice and phrases elaborated for clarity), although she did add a sentence to aid the transition between paragraphs. These two process issues seemed clearly connected to the problems she had had keeping her essay aligned with the prompt, and review of her project-initial reflection suggested this was not an unusual occurrence for her. In discussing the importance of planning, she had written: “Making an outline for your article can help you better control the direction of your article, so that you will not be off topic” (RT1).

[INSERT FIGURE 4 ABOUT HERE]

Figure 4. Process graph from the first recording of Mingyu’s two writing sessions for WT1

Because she already seemed aware of the benefits of planning, our goals for her initially focused on avoiding topic drift by rereading the prompt at regular intervals during planning and formulating as well as before conducting global evaluation and revision. We supplied her with guiding questions to use during global evaluation (Appendix D) and told her to make sure this stage preceded a change of focus to local issues of word choice and grammar.

In her subsequent tasks, she continued to display a penchant for planning, employing different approaches (brainstorming, outlining, freewriting), with the size of her plans usually far exceeding the text proper in the number of characters and in the time spent developing them. Much of this material would not end up in the final product. At one point, she was working from two distinct sets of plans, one labeled “outline” consisting of bullet-point items, and another labeled “outline draft” containing more detailed ideas written in lengthy phrases or complete sentences. Her composing was also highly recursive and nonlinear, indicative of a knowledge-transformational approach. She wrote conclusions before body paragraphs and annotated sections of plan or text with bracketed comments such as “Doesn't explain why” and “this is not the topic!!!”

Despite these latter strategies, she continued to experience problems with drifting off topic. In WT2, instead of writing about a popular fad that she disliked, she chose to write about procrastination, again concentrating on psychological reasons for why people engage in it and the harm it can do. In WT3, she was more on track, providing different perspectives on the question of whether computers benefited a privileged few, but her thesis in the introduction was contradicted by a statement in the conclusion. In the final task, she went off topic yet again, spending most of the essay discussing the importance of having clear goals in college and whether it was even necessary to attend college, while the notion of gap years, the topic of the prompt, was mentioned only twice and never explained.

By the end of her participation in the project, we had identified additional causes for Mingyu's problems staying on task. First, she tended to get carried away by her own ideas, the logic of her thought process, and new information she discovered while resourcing. Her lengthy, complicated plans further facilitated topic drift by making it difficult for her to evaluate her plan in relation to the prompt. An additional cause was her approach to evaluation and revision once she had completed a full draft of her essay. The time she devoted to these processes was less than her case-study counterpart (Table 3), and playback showed her focus during these stages to be almost exclusively local. Once all the components of the text were in place, she would begin local-level evaluation and revision and then finish writing.

Asked about her reluctance to engage in global evaluation despite repeated encouragement and the provision of questions to guide this process, Mingyu expressed doubt about its value in light of the amount of planning she had done.

Mingyu: I didn't ask [these questions] ... after I wrote everything. I read these before I start expanding my ideas ...

Instructor: But you didn't do this after you made a complete draft?

Mingyu: No.

Instructor: How come?

Mingyu: Because I think I already spent enough time on this. (FS3)

Instead of using the questions to evaluate a completed draft of her essay, she was using them to develop and evaluate her plans. She seemed to assume that, as long as she had material in her plan addressing the guiding questions, the final product would also manifest these components and qualities, so there was no need to revisit the questions. “When I finished writing it, I read it again basically to just check grammar, and I think I just assumed that I did well on the task. And then, based on my ideas [in the plan], I read it again. Like, it seems good to me, so I didn't analyze from another aspect” (FS4).

Not surprisingly, Mingyu’s work did not evidence much progress across the four tasks. She seemed frustrated in the final follow-up session but also indicated that she was coming to the realization that she could not rely on planning alone to guarantee good results. “When I was writing this, I had a very clear outline, but when I look at it right now, I can’t remember the outline. I think when I translated my outline to my article, I didn’t write so clear that I can understand,” she said. “When I look at this now, I can’t see the logic in it” (FS4). Like Zedong, however, she finished the project having acquired more terms and concepts for describing the way she went about writing as well as some capacity to interpret the visualizations and make observations about her process engagement.

## **5. Discussion**

### *5.1 Summary of findings*

This study investigated whether process data in the form of visualizations generated from keystroke and eye-tracking logs provided insights about L2 writing that could enhance instructors’ efforts to create more meaningful assessment for learning of L2 writing. The findings clearly provide an affirmative answer. We summarize the findings here with reference to the aforementioned definition of AfL as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where learners are in their learning, where they need to go, and how best to get there” (Broadfoot et al., 2002, pp. 2-3).

### *5.1.1 Understanding where learners are in their learning*

The process data and the collaborative dialogues they facilitated were seen to provide rich information about where students are in their current stage of development of L2 writing skills. This information can be understood from two perspectives. First, it provided macro-level diagnoses of Zedong's and Mingyu's skills in relation to the developmental models proposed by Bereiter and Scardamalia (1987), with Zedong representing the knowledge-telling end of the spectrum and Mingyu displaying a knowledge-transforming approach, albeit one with gaps. These more general characterizations were immediately useful in determining subsequent attempts at intervention; in Zedong's case, adding those processes that were not yet represented in his repertoire, and in Mingyu's case, honing in on a specific process (task definition) with which she seemed to be having trouble. This connects to the second perspective: formative assessment of how students were performing at implementing recommended changes to their process engagement. These subsequent, finer-grained assessments were rich enough in detail to allow the instruction to be individually tailored, with the result that less time was spent on those issues that were more easily addressed (e.g., Zedong's uptake of the placeholder strategy) and more on those problems that proved more intractable (e.g., both participants' reluctance to engage in global evaluation).

### *5.1.2 Understanding where learners need to go and how best to get there*

In these respects as well, the findings showed process data providing insights that conventional, product-focused assessment for learning do not afford. The process goals provided manageable sets of targets for participants to aim for, and because they were often working on the same process or strategy across multiple tasks, there was a sense of connectedness and progression to the instruction that contrasted with what can seem like a succession of detached, standalone tasks in some writing courses. This allowed the participants to develop clearer understandings of individual processes and of process in general as an important aspect of writing as the project progressed.

In addition, the model-writers' process data, when contrasted with that of the participants, showed in visual terms what desired performance looked like and how it could be achieved. The dynamically linked process graph and playback afforded macro- and micro-level views of model

approaches, and because these were recorded data they could be viewed multiple times across follow-up sessions as needed. This affordance is particularly notable in light of research showing the positive effects of observational learning on writing skills development (Braaksma et al. 2004; Couzijn, 1999; Lindgren et al., 2008).

Finally, the process data afforded the instructor the ability to draw from other aspects of the WP literature beyond developmental models in setting goals and recommending specific strategies for achieving them. For example, we knew of ways to address Zedong's disfluencies in formulating because of WP studies (1) showing connections between disfluencies and absence of external planning (Kellogg, 1988) and (2) describing skilled L2 student writers' use of a placeholder-like strategy (Zamel, 1983). To be sure, this knowledge base may not provide solutions for, or even documentation of, every issue instructors may encounter in assessing students' process engagement; for instance, we had trouble finding reference to anything resembling Mingyu's problems with topic drift. However, the potential for process tracing to make this copious body of knowledge about, among other things, differences in skilled and less skilled writers more immediately relevant to L2 assessment and instruction is an important finding.

### *5.1.3 Other insights from process-tracing*

Beyond alignment with these core definitional properties, the study also found process-tracing to present additional affordances for AfL, including as a way of understanding specific, problematic features of written products. A case in point was Mingyu's awkward, after-the-fact attempt to connect a body paragraph back to the topic of the prompt, evidenced by an isolated downward spike in the process graph when she scrolled up to reread the instructions. Had her work on WT1 been assessed only on the basis of the final product, the poorly integrated sentence in question would likely have been regarded simply as a lapse in organization and forgotten. Instead, it was the first indication of an ingrained problem of topic drift related to task definition and global evaluation. The potential for problematic features of written products to be rendered not only more understandable but also instructional by salient events in the process data corroborates what Hattie and Timperley (2007) assert regarding feedback about the task (FT) and feedback about processing of the task (FP): namely, that FT is most useful when it supports FP. Specific

product-process connections such as this are one of the more intriguing findings of this study, and one that merits more research.

Finally, we note how process data provided a useful means of triangulating or validating process information from other sources; namely, student self-reports as expressed in reflective writing. It is probably not uncommon for learners to profess one thing while actually doing something else, as was the case with Zedong's statement valorizing revision in his initial reflection, which contrasted with the almost complete absence of revision in his work on the first task. Process data can reveal such discrepancies, which we hasten to acknowledge do not necessarily result from dishonesty. The cognitive demands of complex writing tasks make it difficult for learners to accurately monitor their engagement in writing processes (Couzjin, 1999), so process data may be useful as input for reflection. Mingyu's reflection about getting off topic was rendered more meaningful by the process data showing her to have a recurring problem in this regard. This raises the question of how process tracing can support not only instructors but learners in AfL, which is discussed below.

### *5.2 Implications for future applications of process-tracing to AfL of L2 writing*

The findings summarized above show clearly that process tracing can support more meaningful assessment for learning of L2 writing. There are several implications of these findings for current practices in L2 writing assessment and instruction. First and most obvious is that the potential benefits of process tracing may be hard to ignore. The experience of assessing writing in terms of both product and process will make it difficult to go back to assessing writing on the basis of product alone. This is because of the depth and breadth of perspective that process data can provide on why written products are the way they are. By assessing products only, instructors may be addressing only symptoms of underlying issues, resulting in lost opportunities to support student learning. The appeal of process-tracing is strengthened by calls for writing assessments that take account of processes (e.g., Deane, 2013) and the availability of a significant body of knowledge at the ready to inform and support its application to AfL.

That being said, we must recognize that incorporating a process focus into AfL would require major changes to current practices. Time would need to be created for this work in syllabuses

and timetables. Teachers and students would require significant preparation and training. As with all assessments, applications for particular contexts would necessitate domain analyses and the development of grading criteria and scoring guides. Given the large-scale nature of the changes—including significant teacher preparation and students’ openness to allowing their work to be scrutinized in this way—validity arguments would need to be developed and validation research conducted to demonstrate that the interpretations and uses of process-based assessments for learning are warranted.

The scale of the changes to current practice that might be involved therefore recommends a gradual and incremental adaptation of current practices. The first author is already experimenting with this in an ongoing writing course in his own work context. Adaptations to the approach used in the study so far include (1) limiting individual consultations with students to two: one at the beginning of the term, after the first writing assignment, and a second at the end of the term while students are working on the final assignment, with a combination of written commentary and visualization feedback provided for those assignments in between; (2) forgoing the use of eye-tracking so that students can write at times and in locations that are convenient for them; and (3) limiting process feedback to a narrower range of problems that are common to at least four of five students so that the use of class time to address these issues can be justified. At the time of writing, the adaptations have been restricted to sections of the course taught by the first author, but in the near future we plan to include additional sections with the help of cooperating instructors who will also serve as research participants in a study about factors that enable and constrain the implementation of process-focused AfL. We expect that previous studies of implementation of AfL-related innovations in existing curricula (e.g., Huang, 2016; Lee & Falvey, 2014) may prove informative in these efforts.

### *5.3 Limitations and future research*

Some limitations must be borne in mind while interpreting these findings. We obviously make no claims about the representativeness of the participants or the generalizability of our findings to larger groups. While we tried to enhance ecological validity wherever possible, the requirement that the participants complete the writing tasks in a campus computer lab may very well have influenced the way they went about it. Another important limitation was that it was



impossible for them to plan or take notes in their L1, due to the CyWrite software's inability to accommodate the use of Chinese characters. Some research has shown benefits for students planning in their L1 (e.g., Jones & Tetroe, 1987).

In addition to those research ideas already mentioned, we see the need for further related studies as follows. First, as a priority, research should evaluate what effects, if any, process-focused AfL has on relevant outcomes such as writing quality, self-efficacy for writing, and mental models of writing. In addition, design-based research (Design-Based Research Collective, 2003) should be used investigate which types of visualizations, or other means of conveying information about processes, are more effective for different audiences and purposes. Finally, and relatedly, research should be conducted to see whether, and under what conditions, process-tracing can be used directly by students to enable self-assessment that informs their learning (i.e., assessment as learning, or AaL; e.g., Lee, 2017).

## **6. Conclusion**

The motivation for this study was to explore whether process-tracing technologies used in writing research could be applied to assessment for learning of L2 writing. These technologies have shown potential not only for helping students better manage the behaviors through which writing is necessarily achieved but for allowing instructors to capture copious amounts of information about students' actual writing abilities with which to inform instruction. They will, we predict, become commonplace in the next five to 10 years. In the meantime, there is still much work to do in building a pedagogy that incorporates them, including much more piloting, evaluation, refinement, and scholarly discussion on the part of practitioners and researchers, as well as professional development on the part of teachers. Based on our relatively brief experience so far, however, we feel process-tracing for AfL holds considerable promise and venture so far as to say there may come a time when writing teachers, and perhaps language testers as well, may wonder how writing was ever assessed without it.

## Acknowledgments

The authors would like to thank the anonymous reviewers and editors for their helpful comments on earlier versions of this paper. We also gratefully acknowledge the National Science Foundation, which supported the project under Grant No. 1550122.

## References

- Almond, R. G., Deane, P., Quinlan, T., Wagner, M., & Sydorenko, T. (2012). A preliminary analysis of keystroke log data from a timed writing task. *ETS Research Report Series, 2012(2)*, i-61. doi: 10.1002/j.2333-8504.2012.tb02305.x
- Bécotte, H.-S., Caporossi, G., Hertz, A., & Leblay, C. (2015). *Writing and rewriting: Keystroke logging's colored numerical visualization*. Unpublished manuscript.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5(1)*, 7-74. doi: 10.1080/0969595980050102
- Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22(1)*, 1-36.
- Breetvelt, I., van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction, 12(2)*, 103-123. doi: 10.1207/s1532690xci1202\_2
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge: University of Cambridge School of Education.
- Caporossi, G., & Leblay, C. (2011). Online writing data representation: A graph theory approach. In *Proceedings of the 10th International Conference on Advances in Intelligent*

*Data Analysis X* (Porto, Portugal, October 29-31), 80-89. DOI: 10.1007/978-3-642-24800-9\_10

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-233.

Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8(3), 165-191. doi: [http://dx.doi.org/10.1016/S1075-2935\(03\)00018-7](http://dx.doi.org/10.1016/S1075-2935(03)00018-7)

Chong, S. W. (2017). Three paradigms of classroom assessment: Implications for written feedback research. *Language Assessment Quarterly*, 1-18. doi: 10.1080/15434303.2017.1405423

Design-Based Research Collective. (2003). Design-Based Research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8. doi: 10.2307/3699927

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: effects on learning and transfer. *Learning and Instruction*, 9(2), 109-142. doi: [http://dx.doi.org/10.1016/S0959-4752\(98\)00040-1](http://dx.doi.org/10.1016/S0959-4752(98)00040-1)

Cumming, A. (1989). Writing expertise and second-language proficiency. *Language Learning*, 39(1), 81-135. doi: 10.1111/j.1467-1770.1989.tb00592.x

Cumming, A. (2002). Assessing L2 writing: Alternative constructs and ethical dilemmas. *Assessing Writing*, 8(2), 73-83. doi: [http://dx.doi.org/10.1016/S1075-2935\(02\)00047-8](http://dx.doi.org/10.1016/S1075-2935(02)00047-8)

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18(1), 7-24. doi: <http://dx.doi.org/10.1016/j.asw.2012.10.002>

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and Second Language Acquisition* (pp. 206-257). Cambridge: Cambridge University Press.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387. doi: 10.2307/356600

- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37(1), 16-55. doi: 10.2307/357381
- Graham, S., Harris, K., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment*. Washington, DC: Carnegie Corporation of New York.
- Hattie, J., & Timperley, H. S. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi: 10.3102/003465430298487
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29(3), 369-388. doi: 10.1177/0741088312451260
- Huang, S.-C. (2012). Like a bell responding to a striker: Instruction contingent on assessment. *English Teaching: Practice and Critique*, 11(4), 99-119.
- Huang, S.-C. (2016). No longer a teacher monologue: Involving EFL writing learners in teachers' assessment and feedback processes. *Taiwan Journal of TESOL*, 13(1), 1-31.
- Jones, S., & Tetroe, J. (1987). Composing in a second language. In A. Matsushashi (Ed.), *Writing in real time: Modelling production processes* (pp. 34-57). New York: Longman.
- 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(2), 355-365. doi: 10.1037/0278-7393.14.2.355
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1), 1-26.
- Khuder, B., & Harwood, N. (2015). L2 writing in test and non-test situations: Process and product. *Journal of Writing Research*, 6(3), 233-278.
- Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching*, 5(1), 19-33.
- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer.

- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50. doi: <http://dx.doi.org/10.1016/j.jslw.2012.11.003>
- Lee, I., & Falvey, P. (2014). Perspectives on assessment for learning in Hong Kong writing classrooms. In D. Coniam (Ed.), *English language education and assessment: Recent developments in Hong Kong and Chinese Mainland* (pp. 221-236). Singapore: Springer.
- 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.
- Lindgren, E., Stevenson, M., & Sullivan, K. P. H. (2008). Supporting the reflective language learner with computer keystroke logging. In F. Zhang & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 189-204). Hershey, NY: Information Science Reference.
- Lindgren, E., & Sullivan, K. P. H. (2002). The LS graph: A methodology for visualizing writing revision. *Language Learning*, 52(3), 565-595. doi: 10.1111/1467-9922.00195
- Lindgren, E., & Sullivan, K. P. H. (2006). Analysing online revision. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer Keystroke Logging: Methods and Applications: Studies in Writing, Vol. 18* (pp. 157-188). Oxford: Elsevier.
- Lindgren, E., Sullivan, K. P. H., Lindgren, U., & Spellman Miller, K. (2007). GIS for writing: Applying Geographical Information Systems techniques to data mine writings' cognitive processes. In M. Torrance, L. Van Waes & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 83-96). Amsterdam: Elsevier.

- Mak, P., & Lee, I. (2014). Implementing assessment for learning in L2 writing: An activity theory perspective. *System*, 47, 73-87. doi: <http://dx.doi.org/10.1016/j.system.2014.09.018>
- National Assessment of Educational Progress. (2011). Writing 2011: National assessment of educational progress at grade 8 and 12 *National Assessment of Educational Progress*. Washington, DC: Institute of Educational Sciences.
- Nicolás-Conesa, F. (2012). *Development of mental models of writing in a foreign language context: Dynamics of goals and beliefs*. PhD dissertation, University of Murcia.
- Nicolás-Conesa, F., Roca de Larios, J., & Coyle, Y. (2014). Development of EFL students' mental models of writing and their effects on performance. *Journal of Second Language Writing*, 24(0), 1-19. doi: <http://dx.doi.org/10.1016/j.jslw.2014.02.004>
- O'Brien, C. W. (1992). A Large-Scale Assessment to Support the Process Paradigm. *The English Journal*, 81(2), 28-33. doi: 10.2307/819969
- Perrin, D. (2003). Progression analysis (PA): investigating writing strategies at the workplace. *Journal of Pragmatics*, 35(6), 907-921. doi: [https://doi.org/10.1016/S0378-2166\(02\)00125-X](https://doi.org/10.1016/S0378-2166(02)00125-X)
- Rijlaarsdam, G., & van den Bergh, H. (2006). Writing process theory: A functional dynamic approach. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 41-53). New York: Guilford Publications.
- Rinnert, C., Kobayashi, H., & Katayama, A. (2015). Argumentation text construction by Japanese as a foreign language writers: A dynamic view of transfer. *The Modern Language Journal*, 99(2), 213-245. doi: 10.1111/modl.12210
- Roca de Larios, J., Manchón, R. M., 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: <http://dx.doi.org/10.1016/j.jslw.2007.08.005>

- Roca de Larios, J., Marín, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, 51(3), 497-538. doi: 10.1111/0023-8333.00163
- Roca de Larios, J., Murphy, L., & Marín, J. (2002). A critical examination of writing process research. In S. Ransdell & M. L. Barbier (Eds.), *Studies in writing, Volume 11: New Directions for research in L2 writing* (Vol. 11-47). Netherlands: Kluwer Academic Publishers.
- Severinson Eklundh, K., & Kollberg, P. (1996). A computer tool and framework for analyzing online revisions. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Spelman Miller, K., Lindgren, E. V. A., & Sullivan, K. P. H. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly*, 42(3), 433-454. doi: 10.1002/j.1545-7249.2008.tb00140.x
- Strijbos, J.-W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small-group chat communication. *Learning and Instruction*, 17(4), 394-404.
- Sullivan, K. P. H., & Lindgren, E. (2002). Self- assessment in autonomous computer- aided second language writing. *ELT Journal*, 56(3), 258-266. doi: 10.1093/elt/56.3.258
- Teddlie, C., & Tashakkori, A. (2008). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles: Sage Publications.
- Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing*, 5(3), 271-294. doi: [http://dx.doi.org/10.1016/S1060-3743\(96\)90005-3](http://dx.doi.org/10.1016/S1060-3743(96)90005-3)

- Victori, M. (1999). An analysis of writing knowledge in EFL composing: A case study of two effective and two less effective writers. *System*, 27(4), 537-555. doi: 10.1016/s0346-251x(99)00049-4
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wengelin, Å., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, 41(2), 337-351. doi: 10.3758/brm.41.2.337
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14. doi: <http://dx.doi.org/10.1016/j.stueduc.2011.03.001>
- Wolcott, W. (1987). Writing instruction and assessment: The need for interplay between process and product. *College Composition and Communication*, 38(1), 40-46. doi: 10.2307/357585
- Worden, D. L. (2009). Finding process in product: Prewriting and revision in timed essay responses. *Assessing Writing*, 14(3), 157-177. doi: <http://dx.doi.org/10.1016/j.asw.2009.09.003>
- Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, 17(2), 165-187.

## APPENDICES

### Appendix A. Writing tasks used in the project

Writing task	Topic and prompt
1	Breaking Traditions: Many adults become upset when young people break with traditions of the past. Do you think that these adults are justified in reacting this way? Why or why not? Support your position with evidence from your own experience or the experiences of people you know.



2	Fads and Trends: Briefly describe a fad or trend that you dislike. Explain why it has attracted so many followers and why you dislike it. Develop your point of view by giving reasons and/or examples from your own experiences, observations or reading.
3	Computers and Privilege: Some people say that computer technology gives an unfair advantage to a privileged few. To what extent do you agree or disagree with the statement above? Support your views with specific examples from your own experience, observations or reading.
4	Gap Year: At least one major United States university officially recommends that high school students take a year off — a so-called gap year — before starting college. The gap year idea is gaining popularity. Supporters say it helps students mature and focus on their goals. Detractors say taking a year off from school will get students off track and that many will never go to college if they don't go right away. Do you think taking a gap year is a good idea? Why or why not?

**Appendix B. Reflection tasks**

*Assigned at the beginning of the project.*

Thinking back on what you have learned in your previous schooling, write a reflection in which you try to explain to a prospective international student from your home country what good academic writing is and what it involves. Your reflection should consist of 250-300 words, or 2-3 paragraphs.

*Assigned at the end of the project.*

Now that your participation in this project has come to an end, write another reflection in which you explain to a prospective international student from your home country what good academic writing is and what it involves. Your reflection should consist of 250-300 words, or 2-3 paragraphs.

**Appendix C.**

**Coding scheme for reflection tasks focusing on participants' mental models of writing  
(adapted from Nicolás-Conesa, 2012)**

---

**Dimensions of writing**

Ideational	E.g., "a good academic writing will help the writer illustrate their idea effectively. The audience should get the writer's idea from the essay."
Linguistic	E.g., "The grammar and spelling mistakes will make the writing hard to understand. "
Textual	E.g., "a good academic writing usually include the thesis statements, supporting points, and the conclusion"

**Orientation**

Process	E.g., "Before submit an academic writing we should check several time about the mistake we usually make, like tense, capital letter and space in front of the paragraph."
Product	E.g., "A good academic writing must contain several parts. "

---

**Coding scheme for process-visualization data**

---

(Re)reading prompt	Initial reading and any re-rereading of all or part of the prompt (which was always positioned at the top of the CyWrite file); indicated by at least three, successive fixations in the same direction
Planning - Brainstorming	Expressing ideas in word or phrase form with no indication of order or hierarchy

Planning - Freewriting	Expressing ideas as extended discursive text with few pauses, or revisions, as part of planning (i.e., not part of the text proper).
Planning - Outlining	Expressing ideas with some indication of order, hierarchy, or organizational role (e.g., numbering, use of labels such as "hook" or "conclusion")
Planning - Replenishing	Interrupting formulation to return to plan to obtain more ideas for formulation; includes copying and pasting of parts of plan into text body to be converted to text proper
Planning - Updating	Interrupting formulation to return to plan and modify it based on new or altered ideas; includes deletion of parts of plan
Evaluation	Instances of rereading of the text written so far to determine if changes are needed; distinguished from briefer lookback events by presents of at least three successive fixations in the same direction
Formulation	Expressing ideas in text intended for the final product; includes "precontextual" revision (i.e., revisions made at the point of inscription)
Revision - Local	Returning to previously written text to make changes at or below sentence level; does not include changes made at the point of inscription (i.e., during formulation)
Revision - Global	Returning to previously written text before the point of inscription to make changes above the sentence level.
Pausing	Periods of no apparent keyboarding, mouse movements, or reading activity
Resourcing - Language	Switching task windows to consult online references for language support
Resourcing - Content	Switching task windows to consult online references for content support

Placeholder strategy	Typing the transliteration of an L1 word, possibly with a notation such as "##"
"Notes to self" strategy	Typing comments or using highlighting to remind oneself of aspects of writing that need attention

---

### **Coding scheme for evaluated essays**

---

Adherence to task and topic	Issues with meeting task requirements as outlined in the prompt
Audience and purpose	Rhetorical concerns, such as assuming cultural knowledge on the part of the reader or failing to state a clear position in an argument
Definition of key term or concept	Issues related to central concepts or ideas; typically, leaving these undefined or insufficiently elaborated
Language issue	Issues related to grammar or lexis
Logical connections	Problems such as gaps in reasoning or contradictions
Organization of information	Includes problems with unity in paragraphs, redundancy, ordering
Support for main ideas	Issues with the relevance and quality of supporting ideas, the diversity of types of support, relative amount of support across main ideas
Thesis statement or controlling idea	Lack of clear thesis statements, topic sentences, or connections between the two; positioning of controlling ideas
Transitions	Lack of transitions, overuse of certain transitions, inappropriate use of transitions

---

## Coding scheme for follow-up session transcripts

---

Views on English and writing	E.g., "I don't think I need to do very much writing, but I definitely going to do a lot of reading."
Combining/ compartmentalizing processes	E.g., "This is like planning and formulating together."
Evaluation/revision	E.g., "I think if something wrong, it's at the end, it's difficult to change the order or make changes."
Formulating	E.g., "Just this time at the end I didn't type a lot. Actually I just can't write anymore. I don't know how to continue when writing."
Language/ resourcing	E.g., "I was confused by <i>justified</i> . I was thinking it was like another form of <i>justice</i> , so I just Googled it."
Planning	E.g., "I think I didn't make my plan very well. The structure of my plan, I have some problem, like they didn't connect."
Prompts/ topics	E.g., "The topic asked me to describe a fad or trend you dislike. I actually don't have a trend or fad I really don't like."
Strategies (Placeholder, Notes to self)	E.g., "Some method, like use a placeholder, like, make me focus on the assignment and do the details later."
Task definition	E.g., "At the very beginning I was writing down my ideas and then I went too far and I start talking about fairness, and then I realized after I checked the topic then I get rid of the fairness thing."
Technology issue	E.g., "The devices was running well last time. It's better than the first time."

Time

E.g., "I think I spent too much time in one article."

---

Appendix D. Goals sheet provided to one of the case-study participants.

## Writing Process Goals for Next Time

For the last writing task, create a much **more detailed plan** than you've been using. Your plan should include a **thesis statement, topic sentences for supporting points** in each paragraph, and **notes for your conclusion**. The weakest parts of your last essay were those that you hadn't planned anything for.

Remember to keep **PLANNING** and **FORMULATING** separate. If you find yourself trying to come up with ideas at the same time that you're writing sentences – stop! Go back up to your plan and generate ideas there. Then, when you have a good enough plan for that section, move back to the text and formulate it.

