Development of L2 writing: fluency and proficiency

This article is about L2 writing and its development, and it examines fluency in relation to proficiency. Fluency in L2 writing can be understood in two ways: the fluency of the written texts (or the outcome) or the fluency of the writing process. Traditionally, fluency measures have been based on the former. The fluency of the writing process, in contrast, is still an under-researched issue. University students studying English or Swedish as their major subject were asked to write (or type) two texts (a narrative and an argumentative text), and a keystroke-logging programme was used to record the writing process, including pauses and all revisions. The texts produced were assessed for their proficiency on a standard six-point scale. The fluency of the writing process was measured by both traditional (offline) and new (on-line) measures. The findings suggest that fluency measures (of both types) correlate with L2 writing proficiency only to some extent, and thus the relationship is not necessarily a linear one.

Keywords: L2 writing, writing process, writing outcome, fluency, proficiency
1 Introduction

This article is about second language (L2) writing and its fluency. Importantly, a distinction can be made between the fluency of the writing outcome (or the text produced) and that of the writing process. The latter is of much more recent interest in research on L2 writing. Computer keystroke logging techniques have been developed by psycholinguists, and these make it possible to trace the writing process of an L2 writer more closely (online) by identifying features such as pauses and revisions made to the text being composed. In addition, these techniques have enabled fluency to be evaluated in new ways. However, research has only begun to address these issues. As part of a bigger research project, the study to be reported below is about L2 writing and its fluency (in the latter sense): fluency will be assessed in relation to the writing outcomes, that is, texts produced by L2 writers of English and Swedish, and their proficiency level.

First, some background to the study will be provided by reviewing key concepts and issues in research on fluency in L2 writing. After that the details of the study that we conducted will be explained (its aims, data collection and processing, and findings) and finally the implications of the study will be discussed.

2 Background

2.1 Fluency: the written product

Often when we talk about the fluency of a piece of writing we are in fact referring to the perceived fluency of the final written product, not of the process. Even when the writing process is mentioned, it is often seen more as a set or series of end products (written words, sentences or texts) (Uppstad & Solheim 2007: 80). When assessing the fluency of the product, the writer is ignored and the judgment is based on the reader’s ideas of what makes a text or performance fluent.

Writing proficiency can be regarded as an aspect of language proficiency; it includes the ability to produce both a variety of genres and rhetorical features, and a range of vocabulary and syntactic structures (Wolfe-Quintero, Inagaki & Kim 1998: 2). Writing proficiency has been viewed and measured in different ways in recent literature on L2 writing fluency. One way of measuring writing proficiency has been to apply developmental measures, such as age or number of years of language instruction. This is based on the idea that the more experience a person has of using a language, the greater their proficiency. Another way of measuring proficiency has been to holistically judge text quality on the basis of a set of pre-established criteria. In the present study,
The Common European Framework of Reference for Languages (2001, hereafter referred to as the CEFR) has been used to assess language proficiency. In general, the CEFR is functional and communicative in its approach to language and language skills. The CEFR assessment scales and indeed the whole framework are built around the principle of being able to communicate in different situations and carry out different language functions (CEFR 2001: 116).

In addition to the general rating scales, the CEFR also gives more detailed descriptions for different text types. In level descriptions for writing, fluency does not play a major role but it is mentioned at all six levels from A1 to C2. For example, at level B1 the writer should be able to “reasonably fluently relate a straightforward narrative or description as a linear sequence of points” (CEFR 2001: 125), and an L2 writer at level C1 can “produce clear, smoothly flowing, well-structured text” (CEFR 2001: 125). It should be noted that in this study, the written products were assessed overall, not only for their fluency.

In the CEFR, cohesion is seen as the most significant characteristic of a fluently written product. At Beginner level (A1–A2), a language user is expected to use simple connectors to link simple sentences, and at Intermediate level the use of cohesive links is expected to be more subtle and varied. Thus, at Intermediate level (B1–B2), a language user should be able to produce a more complex text such as a narrative or a descriptive text that is coherent and cohesive, although some clumsiness is allowed in longer texts. At Advanced level (C1–C2), a language user can be expected to be very fluent. In the CEFR scales, this is made clear by mentions of not only a range of cohesive links but also a variety of organizational patterns. A text at this level should flow smoothly and be logical and effective in its structure. Since the types of texts a language user should be able to produce are also more complex than at lower levels, the language and metalanguage used should take the genre into account. Features mentioned in the CEFR to enhance fluency in the written text (such as the use of connectors), however, do not necessarily tell us anything about the writing process, and, whether it has been fluent or not.

2.2 Fluency: the writing process

Hayes and Flower’s (1980) cognitive model of the writing process was pioneering in that it attempted to break down the writing process into inter-related cognitive sub-processes such as planning, translating ideas into text, and revising. These processes, in turn, consisted of further sub-processes. The model has since been revised, but the basic ideas remain the same (for an overview of cognitive models of writing, see Alamargot & Chanquoy 2001). Writing is claimed to be fluent when transitions between sub-processes
run as automatically and smoothly as possible (cf. Schmidt 1992: 358, who sees fluency as an “automatic procedural skill”). It has been argued that L2 writers’ writing process is less fluent that that of first language (L1) writers because they have less automatized lexical retrieval procedures (Schoonen, Snellings, Stevenson & van Gelderen 2009; Lindgren, Spelman Miller & Sullivan 2008). Besides, a lack of linguistic knowledge and a restricted ability to retrieve linguistic forms quickly place a greater burden on working memory. This in turn may lead to quantitatively as well as qualitatively different pausing behavior, and more revisions and less focus on aspects of textual quality than is found with more proficient writers (Schoonen et al. 2009). Less proficient writers might need to allocate cognitive resources to sub-processes such as revising at lower levels (e.g. spelling) at the expense of higher level processing such as content, style and audience (e.g. Schoonen, van Gelderen, DeGlopper, Hulstijn, Simis, Snellings & Stevenson 2003; Wengelin 2002).

Measures of fluency reported in the literature on writing processes are typically associated with measures of speed of production, length and time. Writing fluency has even been defined as “the rate of production of text” (Chenoweth & Hayes 2001: 94). The underlying idea is that the more proficient a writer is, the faster he or she can retrieve vocabulary and linguistic chunks, which in turn decides the speed of transforming ideas into text. The same point is also found in Wolf-Quintero et al. (1998: 4): “Second language learners write more fluently, or write more in the same amount of time, as they become more proficient.” Based on the idea of speed as a crucial parameter, a traditional offline measure of fluency is to divide the number of words occurring in the final text by the total time spent on the task (i.e. words per minute). The same principles of length and speed apply to studies where writing fluency has been strictly measured on the basis of text products only. The most reliable measures – according to a meta-analysis by Wolf-Quintero et al. (1998) of previous research on fluency measures – include T-unit length, error-free T-unit length and clause length. However, these measures have limitations as they “do not shed light on the production processes that enable a writer to compose text more fluently” (Chenoweth & Hayes 2001: 82). Instead, Burst Size – the average number of words produced between pauses or grammatical revisions – has been put forward as a better descriptor of fluency, reflecting the process through which ideas are transformed into text: “An increase in burst size reflects an increase in the capacity of the translator to handle complex language structures.” (Chenoweth & Hayes 2001: 94).

Spelman Miller, Lindgren & Sullivan (2008) and Lindgren et al. (2008) further developed

---

1 However, Torrance and Galbraith (2006: 73) challenge the commonly held view that proficient writers deliberately and strategically move through repeated plan-translate-revise cycles. Instead, they claim that for most writers smooth flow is repeatedly interrupted.
the ideas of fluency and burst and added the measure of *Fluency during Burst.*\(^\text{2}\) The measures as they appear in Lindgren et al. (2008) are also applied in the present study (the measures will be described in detail in Section 3.2.2 below).

The theoretical modeling of the writing process, as presented above, suggests that there is a relatively straightforward relationship between the fluency of the writing process and the fluency of the finished product: the smoother and more effortless the process, the better the readability and quality of the text produced. However, this does not necessarily seem to be the case. In a study by Schoonen et al. (2009), Dutch university students wrote one text in their L1 (Dutch) and three texts in their L2 (English), and the texts were holistically rated by a panel of raters. The writers indeed turned out to be more fluent in their L1 than their L2, at least in terms of burst and an offline measure of fluency (words occurring in the final text divided by the total time spent on the task). However, no significant relationship was found between fluency and the level of rhetorical development in L2 texts, nor was there any evidence of a relationship between fluency and text quality in either Dutch or English. In a study by Nordqvist Palviainen (2007), measures of the writing process were related to text quality, and similar patterns of pausing, revision and speed behaviors were found in L2 texts whether they had been rated to be of good or poor quality. Again, other studies (e.g. Lindgren et al. 2008; Spelman Miller et al. 2008) have shown writing fluency to be a strong predictor of text quality, and Lindgren et al. (2008) found that high school students writing in their L1 (Swedish) and L2 (English) improved their writing fluency significantly in both languages over time. Previous research thus shows evidence both for and against a relationship between writing fluency and text quality. Because of these conflicting findings further research is needed to explore how the fluency of the writing process relates to levels of language proficiency, and this is in fact what this study attempts to do.

\(^2\) In contrast to Chenoweth and Hayes (2001), where words were the unit of analysis, characters were used as the unit of measurement in these studies. A word as a unit is problematic when using computer-based automatic syntactic analysis (cf. also the difficulties of defining what a ‘word’ is: when composing a text it is common to delete or revise (parts of) words before they have been fully spelled out). Using characters (instead of words) as a unit of analysis makes it possible to compare fluency in languages with different morpho-syntax: the length of words may differ considerably from one language to another.
3 The present study

3.1 Aim of the study

The study to be reported on is part of a research project entitled *Paths in Second Language Acquisition* (or *Topling*). Some members of the team are experts in language assessment, while others are scholars interested in SLA (Second Language Acquisition). Overall, the project is concerned with the development of L2 writing in three languages (Finnish as L2, English and Swedish) in terms of proficiency levels, as measured by two standard rating scales, and within the proficiency levels in terms of complexity, accuracy and fluency, development traced for aspects of text (e.g. conjunctions), grammar (e.g. questions) and/or lexicon.

Within the project, the aim of the present study was to find out how the L2 writing process (in terms of fluency, measured by traditional and more recent online measures) relates to the outcome (in terms of L2 language proficiency as assessed on a standard rating scale). For this purpose, groups of L2 writers were asked to complete two communicative writing tasks, and the writing process was recorded online, using special software. Features of the writing process were analyzed and the quality of the writing outcome (i.e. the proficiency of the L2 texts produced) evaluated.

3.2 Data collection and processing

L2 writing data were collected from first-year university students with English or Swedish as their major subject. As a rule, the students of English already had nine years of formal teaching and learning of the language behind them (Grade 3 through Grade 12), and the students of Swedish six years (Grade 7 through Grade 12). The majority of the writers were young women.

The students were asked to tell a story based on their own experiences (a narrative text) and to express their opinions on a specific topic (an argumentative text) and give justifications. The instructions given to the students for the writing of the first text type were as follows:

---

3 The project is funded by the Academy of Finland over the years 2010–2013 and led by Maisa Martin. For details, see https://www.jyu.fi/hum/laitokset/kielet/topling/en
And to write the second text type the Swedish majors were instructed as follows:

**Mieliipide**


1. Lika stor andel kvinnor och män till högskoleprogram!
2. Avgiftsfri högskoleutbildning – en börda för skattebetalarna?

The instructions and the titles for the narrative writing tasks (“A memorable event in my life”) were thus given in Finnish (i.e. in the students’ L1), and the titles for the argumentative task in the language the students had been studying (i.e. in either English or Swedish). For the second writing task the students could choose from two topics, adapted to suit these specific groups of L2 writers: “Quotas for female and male students on university study programmes!” or “Tuition-free studies in Higher Education: A burden to tax payers?” No guidelines were given as to the length of either type of text.

The students wrote – or typed – the texts in a computer laboratory, one text after the other. Using the Scriptlog software (Strömqvist & Karlsson 2002), the writing process was recorded online, and so record was kept of each pressing of a key on the keyboard, and of all pauses and revisions of the text (including deletions and additions). While producing the texts the students were, however, unaware that their writing process was being recorded in such detail. (For an overview of different types of keystroke logging software, see Sullivan & Lindgren 2006.)

**3.2.1 Assessing proficiency**

The quality of the L2 writers’ writing outcomes (or finished texts) was assessed by assigning them to an L2 proficiency level as follows. The texts produced (N = 103) were assessed by trained raters, using the six-point CEFR scale of language proficiency levels, ranging from A1 (Breakthrough or Basic User) to C2 (Mastery or Proficient User). The scales used were a compilation of criteria given in the CEFR (2001) on writing

---

4 Of the texts, 62 were written in English and 41 in Swedish. One text had to be excluded due to technical problems.
various types of texts. Importantly, **communicativeness** was the main criterion used in assessing the texts. Thus, the raters were not specifically instructed to focus on accuracy or fluency (see Section 2.2).

To ensure reliability, at least three of the four raters of the English texts, and two of the three raters of the Swedish texts, had to agree on the CEFR proficiency level. In addition, the rater who deviated from the others could deviate by only one CEFR level. If all the raters had disagreed, the text would have been excluded from further analysis. This was, however, not the case with any text.

Out of the total of 103 texts, 13 were rated to be at proficiency level B1 on the CEFR scale, 31 at level B2, 44 at level C1, and 15 at level C2. In other words, the texts varied from levels B1 to C2, being mostly of level B2 or C1 (and none of level A2 or A1).

### 3.2.2 Measuring fluency

The fluency of the L2 writing process data collected was measured by the following **online fluency measures** (for details see Lindgren et al. 2008), which capture the underlying process in unprecedented detail:

- **Fluency (linear)** = number of characters produced per minute in linear text
- **Burst** = total number of typed characters / (total number of revisions + total number of pauses)
- **Fluency during Burst** = total number of typed characters / (total writing time – total pausing time)

In addition, a traditional **off-line fluency measure** based on the product was applied to the L2 writing process data collected. This measure, here referred to as Fluency (product), has often been used in previous studies (e.g. Schoonen et al. 2003), and a decision was made to use it for the sake of comparison. It is defined as follows:

- **Fluency (product)** = number of characters produced per minute in final text.

Fluency (linear) and Fluency (product) are thus both measures of the general speed of the writing process. However, whereas the former includes all characters typed during the process, the latter ignores characters that may have been deleted during the writing session. The concept and measurement of Burst were first put forward by Kaufer, Hayes and Flower (1986) and were further employed by Chenoweth and Hayes (2001, see 5 The themes covered in the compilation of the scales included Overall written production; Written interaction; correspondence, notes, messages & forms; Creative writing; Thematic development and Coherence & cohesion (CEFR 2001: 61–62, 83–84 and 125).
Section 2.2), who argued that ideas transformed into text are produced as chunks. The more proficient a writer is, the more complex and longer these chunks (i.e. Bursts) are. Burst is here operationalised as the average number of characters typed between pauses (2 seconds or longer) and/or revisions (insertions or deletions of characters). Finally, Fluency during Burst refers to how fast (in typed characters per minute) the writer writes when typing. Hence, in Fluency during Burst, the time spent pausing is subtracted from the total time spent on the task.

In addition to the fluency measures applied in this study, Scriptlog provides other types of data on the writing session, revealing other information about the writing process. These will be reported in Section 3.3 below.

### 3.3 Findings

Some process and outcome measures of the texts in relation to the CEFR proficiency levels were compiled and are presented in Table 1.

<table>
<thead>
<tr>
<th>Proficiency levels</th>
<th>Total writing time (min)</th>
<th>Total tokens in final text</th>
<th>Deleted tokens</th>
<th>Proportion of pause time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Deviation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 N=13</td>
<td>11</td>
<td>3</td>
<td>1036</td>
<td>223</td>
</tr>
<tr>
<td>B2 N=31</td>
<td>16</td>
<td>6</td>
<td>1099</td>
<td>296</td>
</tr>
<tr>
<td>C1 N=44</td>
<td>22</td>
<td>9</td>
<td>1375</td>
<td>643</td>
</tr>
<tr>
<td>C2 N=15</td>
<td>39</td>
<td>11</td>
<td>186</td>
<td>1364</td>
</tr>
</tbody>
</table>

On average, a C2 text took almost four times as long to produce (in terms of minutes) than a B1 text; it was twice as long (in terms of number of characters in the final text); it was revised many more times (in terms of deletions) and less time was spent pausing (in terms of inactivity of more than two seconds) in producing it.

Importantly, there is also some variation within the proficiency levels (based on the standard deviation measures). This is particularly true of levels C1 and C2: the amount of time spent on writing, text length and deleted tokens differ a lot within the
levels. It is, however, worth noting that on average the number of deleted tokens in the final text clearly increases from level B to level C. Thus, while writers at levels C1 and C2 spent more time on their texts, they also seemed to rewrite more than writers on lower levels. In contrast, writers at levels B1 and B2 revised less, but spent more time pausing.

The main findings concerning the fluency measures in relation to the CEFR proficiency levels are summarized in Table 2.

<table>
<thead>
<tr>
<th>Proficiency levels</th>
<th>Fluency (linear)</th>
<th>Burst</th>
<th>Fluency during Burst</th>
<th>Fluency (product)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B1</strong> N=13</td>
<td>Mean 52.3</td>
<td>7.0</td>
<td>147.2</td>
<td>38.9</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 13.7</td>
<td>1.3</td>
<td>20.0</td>
<td>11.3</td>
</tr>
<tr>
<td><strong>B2</strong> N=31</td>
<td>Mean 72.7</td>
<td>9.0</td>
<td>172.0</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 29.0</td>
<td>2.2</td>
<td>32.7</td>
<td>22.3</td>
</tr>
<tr>
<td><strong>C1</strong> N=44</td>
<td>Mean 97.2</td>
<td>11.4</td>
<td>201.9</td>
<td>70.9</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 33.6</td>
<td>4.5</td>
<td>39.9</td>
<td>33.2</td>
</tr>
<tr>
<td><strong>C2</strong> N=15</td>
<td>Mean 104.1</td>
<td>12.5</td>
<td>213.7</td>
<td>69.7</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation 21.3</td>
<td>6.1</td>
<td>35.6</td>
<td>18.3</td>
</tr>
</tbody>
</table>

As was the case with the measures in Table 1, there are again significant differences among the levels. Thus, the general trend is that the higher the proficiency level, the better fluency (Fluency (linear) and Fluency (product)), the longer uninterrupted stretches of production (Burst), and the higher the speed during bursts (Fluency during Burst). A statistical non-parametric test (Independent-Samples Kruskal-Wallis Test) showed these differences to be significant (p=0.000) for all measures.

When looking at the numbers in Table 2 in greater detail, we find only minimal differences between levels C1 and C2. Therefore, all proficiency levels were also statistically analyzed in relation to each other. The statistical non-parametric test (Independent-Samples Mann-Whitney U Test) showed significant differences between all measures at levels B1 and B2 in relation to the other levels. The statistical analysis showed, however, no significant differences for any of the four fluency measures, between the two highest levels, i.e. between levels C1 and C2. Hence, the texts rated on these two levels did not differ in any respects regarding fluency.
4 Conclusions

Previous research on the relationship between L2 proficiency (as measured by assessed text quality) on the one hand and writing fluency on the other hand have provided somewhat contradictory results (e.g. Schoonen et al. 2003; Lindgren et al. 2008). Our study confirms that the issue is a complex one. Whereas there was a clear tendency for fluency to be better in level B2 texts than in level B1 texts, and better at Advanced level than Intermediate level, no such effect was found between the two highest levels, C1 and C2. The findings thus suggest that writing fluency develops in close connection with L2 proficiency at intermediate levels, but its role at more advanced levels is not significant.

The development of fluency is not necessarily strictly linear. Further exploration is needed of the relationship and directionality between writing fluency and L2 proficiency. Is there a causal relationship such that high proficiency in L2 leads to better fluency, or does better fluency lead to higher L2 proficiency? Lindgren et al. (2008) argue for the latter and suggest that students should be helped to achieve automatization and good writing fluency, as highly proceduralized rules for language production enhance lexical retrieval, which in turn frees up cognitive capacity from working memory (see also Spelman Miller et al. 2008; Schoonen et al. 2003).

Although the fluency measures were much the same at levels C1 and C2, there are other types of data available on the process that reveal differences between the levels. As Table 1 shows, compared with C1 texts, C2 texts took longer to compose, the final texts were considerably longer and more revisions had been made in them (in terms of deleted tokens). Level C2 writers hence used more time writing texts, and possibly produced more polished texts with greater lexical, grammatical and pragmatic variation for raters to assess. This study did not take into account whether a rater considered a text to be fluent or not, but this may have influenced their rating. So in the future it would be worthwhile to investigate how particularly C1 and C2 texts differ from each other linguistically, and what raters focus on when deciding between these two levels.

In the future, texts produced at Beginner level (A1 and A2) should also be analyzed to gain a more comprehensive picture of the development of L2 writing fluency from one level to another (i.e. across the whole proficiency scale). Furthermore, since the participants in this study were students of two different languages, the effects of writing in Swedish as opposed to English as L2 should be explored, as well as the writers’ fluency in their L1 (i.e. Finnish). Finally, the participants wrote a narrative and an argumentative text; the possible effects of text type should be considered in future research to see if the results are the same regardless of genre.
Research on writing fluency is a relatively recent field of study and much remains to be explored. Composing a text in L2 is certainly a complex cognitive enterprise and operationalized measures of fluency should be used with some caution. Adding other types of process data (such as provided above) gives a fuller picture, but also interdisciplinary approaches and methods are needed to gain a better understanding of the phenomenon.

**Literature**


