

University of Groningen

Long-term effects of explicit versus implicit instruction on EFL writing

Ke, Huimin; Luo, Yu; Steinkrauss, Rasmus; Piggott, Leslie

Published in:
Dutch Journal of Applied Linguistics

DOI:
[10.51751/dujal9361](https://doi.org/10.51751/dujal9361)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Ke, H., Luo, Y., Steinkrauss, R., & Piggott, L. (2021). Long-term effects of explicit versus implicit instruction on EFL writing. *Dutch Journal of Applied Linguistics*, 10. <https://doi.org/10.51751/dujal9361>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Long-term effects of explicit versus implicit instruction on EFL writing

Huimin Ke¹, Yu Luo¹, Leslie Piggott² and Rasmus Steinkrauss¹

¹University of Groningen | ²Utrecht University

Abstract This study investigated the long-term effectiveness of explicit versus implicit instruction in a classroom setting. The participants were 114 Dutch secondary school students learning English as an L2; a control group received explicit and an intervention group received implicit instruction in a meaning-based context for two academic years. Instructional effects were measured via a timed writing task. The writing products of the two groups were compared in terms of holistic text quality, writing complexity, accuracy, and fluency. The results revealed that explicit and implicit instruction were equally effective in terms of promoting global writing proficiency, writing complexity and fluency. Regarding accuracy, interesting differences were found in the learners' command of tense/aspect related verb phrases. The findings suggest that the benefit of explicit instruction lies in the use of correct verb forms, whereas the benefit of implicit instruction lies in the correct choice of tense/aspect in the communicative context.

Keywords explicit/implicit instruction, EFL writing, longitudinal study, tense/aspect

Article history

Received: May 01, 2020

Accepted: April 13, 2021

Online: December 8, 2021

Corresponding author

Huimin Ke, h.ke@rug.nl

Author contributions

Huimin Ke, investigation, methodology, data analysis, writing – original draft, writing – review and editing, visualization; Yu Luo, investigation, methodology, data analysis, writing – original draft, visualization; Leslie Piggott, conceptualization, methodology, resources, investigation, review; Rasmus Steinkrauss, supervision, methodology, writing – review and editing

Copyright

© Author(s); licensed under Creative Commons

Attribution 4.0. This allows for unrestricted use, as long as the author(s) and source are credited.

Funding information

No funding is reported for this study.

Conflicting interests

The authors have declared that there were no conflicting interests.

Supporting information

Open datasets

1 Introduction

In the past two decades, the effectiveness of explicit/implicit instruction has been extensively researched in the field of instructed second language acquisition (ISLA). Alongside the accumulating findings, researchers have meta-analyzed the empirical evidence to seek conclusive answers to the efficacy debate. Following Norris and Ortega's (2000) seminal work, more researchers have continued with meta-syntheses and updated the knowledge of ISLA research. Spada and Tomita (2010) and Goo and colleagues (Goo et al., 2015) suggest that explicit instruction is more effective than implicit instruction in promoting L2 development. However, as acknowledged by the reviewers, the existing

evidence is inconclusive due to the prevalence of constrained tests and brief treatments. It is therefore not surprising that the field calls for classroom-based studies to compare the effects of explicit versus implicit L2 instruction in authentic educational contexts (DeKeyser & Botana, 2019).

The current study investigated the longitudinal effects of explicit and implicit L2 instruction in authentic mainstream classroom settings after a period of two years of instruction. The learning outcome was measured by a free-response writing task. The data was adopted from a larger project, namely Piggott (2019; this study will be reviewed in detail later). In the previous study, a complex Multilevel Modelling was employed to test a great amount data collected from participants across various educational streams, the limitation of which was that the effect sizes could not be established. In the current study, we narrowed down our focus on one educational stream, rescored the texts with a more elaborate rubric, and manually coded the texts for an array of complexity, accuracy, and fluency measures. We also employed new analyses which allowed calculation of effect sizes. Our purpose was to investigate whether explicit and implicit types of instruction impact L2 learners' writing performance differently in terms of overall writing quality, linguistic complexity, accuracy, and fluency after two years of instruction.

2 Literature review

2.1 Instructed Second Language Acquisition

The explicit/implicit dichotomy has been a core concept in SLA (Andringa & Rebuschat, 2015; Ellis, 2009; Hulstijn, 2005) since the distinction between acquisition and learning was made (Krashen, 1982). Language acquisition refers to the process in which the learner acquires the target language subconsciously; language learning, on the other hand, is related to the learner knowing about the target language. Krashen and Terrell (1983) have proposed "the natural approach", according to which the role of L2 instruction is to provide comprehensible input for language acquisition to take place while providing learners with metalinguistic knowledge is ineffective and can impede L2 development. Opponents of the natural approach, however, argue that providing L2 learners with comprehensible input alone may not suffice for desirable levels of attainment in the target language (Hammerly, 1987; Harley, 1989). Students receiving abundant L2 input may attain a high level of fluency in the target language, but continue to make grammatical errors. Moreover, communicative interaction, input processing, comprehensive output, and feedback may all have an important role to play in the language classroom, because they contribute to the probability of learners noticing and internalizing the linguistic structures (Gass et al., 1998; Lightbown & Spada, 1990; Pica, 1994; Swain, 1985). Meaningful input, then, mainly serves to create a context where intervention facilitates language

acquisition. Therefore, L2 instruction should concentrate on providing enough opportunities for the “external manipulation of learners’ focal attention” (Norris & Ortega, 2000, p. 420).

However, how “explicit” instruction should be in order to promote noticing and foster L2 development has been repeatedly debated. Scholars provided several definitions to distinguish different styles of L2 instruction. Long (1991) categorizes instruction into three types: Focus on Meaning, Focus on Forms, and Focus on Form. A Focus on Meaning instruction is compatible with Krashen’s acquisition hypothesis, insisting that L2 acquisition takes place incidentally in the practice of learners being exposed to authentic and comprehensible input. In contrast, Focus on Forms deals with linguistic features in isolation; learners, with their attention directed to forms, learn the constructions externally and systematically. Finally, Focus on Form integrates several properties of the former two. Here, language is learnt in communicative contexts in which meaningful input is provided while the learners’ attention is directed to certain linguistic forms which may hinder L2 progress. This categorization provides a well-founded theoretical framework for studies on L2 instruction; however, in a real language classroom, it is impractical to solely follow one type of the three.

Spada (1997) refers to a variety of L2 instruction approaches as “form-focused instruction”, which occur in a meaning-based context with the attention drawn to linguistic forms “in either spontaneous or predetermined ways” (p. 73). In essence, Spada’s terminology includes both Focus on Forms and Focus on Form, and the two instructional styles are further distinguished as “isolated” and “integrated” form-focused instruction respectively (Spada & Lightbown, 2008). Form-focused instruction thus became an umbrella term through which a broader range of instructional activities can be included. For the convenience of ISLA studies, however, more concise terms – explicit versus implicit instruction – have been proposed and used. Explicit instruction requires focal attention to forms and provides relevant explanations of the grammatical rules whereas implicit instruction provides forms without discussing them further (Hulstijn, 2005). The concepts of L2 instruction have evolved in the past decades, together with an increase in empirical studies that attempted to compare the effectiveness of various instructional approaches.

2.2 The effectiveness debate

Given the variations on the operationalization of instructional effectiveness in SLA studies (Norris & Ortega, 2000), it is not surprising that mixed results have been found. Norris and Ortega (2000), for instance, examined the effectiveness of explicit versus implicit L2 instruction by conducting a meta-analysis in which 77 research reports were selected after extensive evaluation. They classified the instructional treatments as being explicit or implicit and adopted Cohen’s *d* to calculate the effect size for comparison. The results show that explicit instruction displayed a clear advantage ($d = 1.13$) over implicit

instruction ($d = 0.54$), with further confidence intervals qualifying the difference was valid. Nevertheless, the authors have noted that the findings should be interpreted with caution because of biased testing measures of the learning outcomes, brief interventions, and the methodological variance among the reviewed studies.

Following Norris and Ortega's (2000) work, more meta-analyses were conducted to update the knowledge in ISLA. Spada and Tomita (2010) examined 30 empirical studies and aimed to compare the efficacy of explicit and implicit instruction on learning simple and complex L2 structures. Goo et al. (2015), on the other hand, investigated 34 studies published from 1993 to 2011 and paid special attention to the potential factors that might mediate instructional effects. Both studies endeavored to compare the effectiveness of explicit and implicit instruction and suggest that explicit instruction works better than implicit instruction. Yet, the conclusion seems premature because the empirical studies reviewed in these two meta-analyses were hampered by the same problems pointed out earlier by Norris and Ortega (2000): Measures favored explicit knowledge and the treatments were brief.

Such bias has been avoided in recent studies conducted in the Netherlands. In a three-year investigation, Rouse-Malpat et al. (2019) compared the effects of structure-based (i.e. explicit) instruction and dynamic usage-based (i.e. implicit) instruction on secondary learners' acquisition of L2 French. The structure-based instruction included a traditional textbook with grammatical explanations, whereas the dynamic usage-based treatment avoided grammatical rules but used the Accelerative Integrated Method ("What is AIM", n.d.) which provided storytelling, music, and gestures to facilitate understanding. Free-response written data were analyzed and the results show that usage-based instruction leads to higher linguistic complexity both in morphosyntax and lexicon. These findings correspond to an earlier and similar study conducted by Rouse-Malpat and Verspoor (2012). They conducted a two-year investigation to compare instructional effects of Focus on Meaning and Focus on Form through free-response oral data. The study shows that meaning-focused instruction helped the learners to develop equal accuracy, but higher fluency, in their L2 French oral production.

Another longitudinal study was carried out by Piggott (2019), who investigated the effects of delaying and reducing explicit grammar instruction to adolescent Dutch learners of English. In this study, two cohorts of students ($N = 222$ for the explicit group; $N = 241$ for the implicit group) were observed during their first two years of secondary education. The explicit group were taught with a communicative language teaching approach in which explicit rule explanation and grammar drills were included; the implicit group were taught by the same teachers with the same textbooks, but all pages with explicit rule explanation and drills had been eliminated for two academic years. Piggott's findings are in line with the evidence documented in the abovementioned Rouse-Malpat and Verspoor (2012) and Rouse-Malpat et al. (2019) studies: Learners receiving implicit instruction show a more complex and fluent L2 performance than their counterparts do both in speaking and writing.

Thus far, recent longitudinal investigations seem to argue for advantageous effects of implicit instruction on promoting complex and fluent L2 writing performance. Moreover, in terms of overall writing quality and accuracy, long term investigations did not find explicit instruction to be more effective, which is at variance with the conclusions drawn from previous meta-syntheses (e.g., Norris & Ortega, 2000; Spada & Tomita, 2010). As suggested in these meta-syntheses, length of intervention and outcome measures might have largely contributed to the different findings in instructed SLA studies.

2.3 Length of intervention, outcome measures, and aspects of proficiency measures

Regarding the length of intervention, Ortega and Iberri-Shea (2005) argue that as L2 development takes time, longitudinal research can tap into issues that cannot possibly be solved by short-term observations. For example, longitudinal studies can provide a lens to identify developmental characteristics at varying stages of L2 development, which facilitates learners in attaining desired learning outcomes across different contexts (Ortega & Byrnes, 2009). Ortega and Iberri-Shea (2005) further suggest that the length of observation should be scaled on institutional time. As far as measurement is concerned, learning achievement has often been expressed in terms of metalinguistic knowledge with constrained tests (e.g., grammaticality judgement tests). Metalinguistic knowledge can be learned quickly and lends itself well to short-term interventions. Constrained tests thus have been widely criticized for being biased and free-response tests have been called for (Norris & Ortega, 2000). In fact, empirical studies that have used free-response data did not find explicit instruction to be more effective than implicit instruction (Macaro & Masterman, 2006; Piggott, 2019; Rouse-Malpat et al., 2019, Rouse-Malpat & Verspoor, 2012). Taken together, long-term interventions and unconstrained tests are required in order to obtain more robust findings in L2 type-of-instruction research.

As L2 proficiency is a complex construct and multi-dimensional, different measures provide different yet equally meaningful insights into L2 development. Holistic human scoring presents a rounded view of learners' language competence in that a human mind can weigh different strengths and weaknesses of a text simultaneously and has been considered very informative in L2 studies (Hou et al., 2016; Polio, 2001; Verspoor et al., 2012). Meanwhile, the multidimensionality of L2 performance can be investigated by looking at separate sub-components (Ellis & Barkhuizen, 2005; Skehan, 1998). For example, quantifiable measures tapping into the complexity, accuracy and fluency (CAF) of L2 performance have long been recognized and used in SLA research (Wolfe-Quintero et al., 1998). Complexity, from a linguistic perspective, concerns mostly syntactic and lexical constructions; accuracy concerns the degree of conformity to target language norms; fluency concerns the speed of L2 processing in given contexts (Abdel Latif, 2013; Bulté & Housen, 2014; Housen & Kuiken, 2009; Larsen-Freeman, 2009). CAF measures have been used both as descriptive guidelines for learners to assess their language proficiency and as indices to gauge L2 progress (Housen et al., 2012). In the current study, both holistic

ratings and CAF measures were employed to compare the learners' written production under explicit or implicit instructional conditions.

In terms of which specific CAF measures are reliable and suitable for our learners, references were taken from Verspoor et al. (2012). This is because our participants were similar to the L2 learners investigated in Verspoor et al. regarding the learners' L1 background and educational context. Verspoor and colleagues contribute to a comprehensive evaluation of 64 measures to assess L2 writing development ranging from beginner to intermediate levels. Among these measures, we chose mean length of T-unit (MLT) and ratio of dependent clause per T-unit (DC/T) to investigate broad features of syntactic complexity (pp. 260–261), the ratio of Present Simple Tense (%PST) to measure verbal complexity (p. 256), and the Guiraud index (“a Type Token ratio adjusted for text length”, p. 252) to measure lexical richness, given that these measures demonstrated significant correlations to holistic writing quality.

Apart from the abovementioned broad measures, we especially endeavored to investigate writing accuracy. After two years of explicit versus implicit instructional treatment, we expected to see different competence in tense/aspect related verbal accuracy. Previous research has revealed that learners at low levels may exhibit different developmental patterns in their mastery of the contextual *use* of tense/aspect and the corresponding inflections of the *form* (Rousse-Malpat et al., 2019; Spoelman & Verspoor, 2010; Verspoor et al., 2012). We thus operationalized two measures to gauge tense/aspect related verbal errors, of which one looked at the contextual use of the tense/aspect (hereafter, verb use) and the other the inflectional form of the verb phrases (hereafter, verb form). In addition, we also investigated other types of errors, namely, lexicon, grammar, word order, spelling, mechanics and punctuation errors (see Verspoor et al., 2012, p. 253).

Lastly, we used text length to measure fluency. Writing fluency is a complex construct that is difficult to measure (Abdel Latif, 2013). An anonymous reviewer pointed out that text length might not be a sound measure of fluency unless the writing task explicitly asked the writer to write as many words as possible. Indeed, a writer might invest more time into coherence and cohesion, or word choice, instead of writing a longer text. However, as text length clustered together with words per minute in a factor analysis of argumentative essays by Oh (2006; reviewed in Norris & Ortega, 2009), we argue that text length can still be an informative variable under a timed test condition to gauge fluency (Jarvis et al., 2003). Compared to a shorter text, a longer text indicates that the writer is likely to have L2 resources at their disposal more readily and is better able to manipulate these resources for text-based communication (Skehan, 1998).

To sum up, the present study intends to revisit the efficacy debate of explicit versus implicit instruction by following intact classes and applying a free-response writing test to compare the long-term learning outcomes. The data used in this study was taken from a larger project which was previously reviewed (i.e. Piggott, 2019). Previously, it was found that after two years of intervention learners receiving implicit instruction showed steeper growth in writing complexity and fluency than their counterparts did

both in speaking and writing. But the impact on accuracy, especially on the tense/aspect related competence in writing, remained unclear. This was of special interest given that tense/aspect development was placed at the center of the form-focused curricula and that a great deal of time and effort was spent in rule explanation and grammar drills in the explicit teaching condition (Piggott, 2019, p. 70). As our investigation required intensive manual coding of the written texts, we took a subset of the data to reduce the workload: the texts written by learners from the highest scholastic competence stream. The reason for choosing this subgroup was that more proficient L2 writers would use a wider range of linguistic features compared to less proficient ones, and thus we expected to see more variation in the written production produced by students from the highest scholastic stream.

The overarching research question is: After two years of intervention, does explicit instruction lead to better writing proficiency than implicit instruction in terms of overall writing quality, linguistic complexity, accuracy, and fluency? Based on previously reviewed meta-syntheses and longitudinal studies, we hypothesized that:

- (1) Overall, explicit and implicit instruction would lead to equal general writing proficiency as measured by holistic ratings of text quality.
- (2) Implicit instruction would lead to more complex writing in terms of syntactic complexity, lexical richness, and tense/aspect related verbal diversity.
- (3) Explicit instruction would lead to more accurate writing as measured by error ratios.
- (4) Implicit instruction would lead to more fluent writing as measured by text length.

3 Method

3.1 Participants

The participants involved in the present study were two cohorts of students aged 12 to 14 who entered the same secondary school in two successive years. Students who entered school in 2014 comprised the explicit group ($N = 55$), whereas the 2015 cohort formed the implicit group ($N = 59$).¹

The participants' leaving reports for primary education showed that their English proficiency was approximately at A1 to A2 level according to the Common European Framework of Reference (CEFR) before entering secondary education, i.e. at the start of the current period of observation. At the beginning of the observation, a reading and a listening test were administered in class and scored by the English teachers in the participating school, the results of which were used to examine the participants' English proficiency before the observation.

3.2 Instructional treatment

Both groups were instructed within a communicative language teaching framework using the same course books: *More!* (Puchta & Stranks, 2008) for the first year and *Activate!* (Boyd & Barraclough, 2010) for the second year. While the explicit group received instruction that contained both meaning and explanation of grammatical rules, the implicit group was instructed without rule explanation. For the implicit group, the grammar-oriented pages in the textbooks had been torn out and grammatical explanation, drills, and explicit feedback about grammatical rules were avoided both in class and after class. In other words, the instructional difference between the two groups lay in whether metalinguistic knowledge was provided. The groups had the same amount of classroom time (i.e. same amount of instruction) and were both taught in English only by the same teachers. While the explicit group invested time in learning grammatical rules, the implicit group spent time on meaning-focused textbook exercises (see Piggott, 2019, pp. 70–71 for more detail).

3.3 The writing task

Towards the end of the observation, which was in 2016 for the explicit group and in 2017 for the implicit group, the participants were tested on their writing proficiency. The writing tests were organized as formal in-class examinations, in which the students were given 50 minutes to write a letter about their latest holiday experience. The required text length was no less than 140 words. The participants were allowed to write as long as they could within the given time. Dictionary use was not allowed. As tense/aspect had been the grammatical focus in the curriculum, the writing test was designed to elicit the use of varied tense/aspect forms. The exam instructions are given in Appendix 1. The hand-written texts produced by the learners were collected as raw data.

3.4 Treating the data

All the hand-written texts were transcribed by the first author and checked by the second author. The greeting and closing parts of the letters were deleted as they were standard formulae. In order to conceal private information (i.e., names and addresses) and to avoid overrating lexical richness, we replaced proper nouns with *Name* or *Place* wherever that was applicable (cf. Verspoor et al., 2012). Exceptions were country names (e.g., France) and names of capital cities (e.g., Paris) which were kept because they were part of the learning objectives in the curriculum. Finally, texts were converted by the transcription conventions of the Codes of the Human Analysis of Transcripts (CHAT); accordingly, redundant Dutch words and data that could not be recognized by CLAN (MacWhinney, 2014) were cleaned. Such cleaning is listed and explained in Appendix 2.

3.5 Holistic assessment

We employed six trained raters (two professors and four master students of Applied Linguistics) to holistically assess the text quality with a rubric used by Hou et al. (2016). This rubric included not only complexity, accuracy, and fluency, but also idiomaticity and coherence (the CAFIC rubric, see pp. 90–91). We chose this rubric because it could reflect L2 writing quality from a broader linguistic scope.

The training procedure for rating was similar to the one used by Verspoor et al. (2012) and the CAFIC rating scale was used for the holistic scoring. To allow raters to discover for themselves what characteristics contributed to weaker or stronger texts, twelve texts of varied writing quality were selected and used as samples in a trial session. The raters read three texts at a time and judged independently which text was the weakest and which the strongest in each set. Disagreements were discussed until rater agreement was reached. The raters continued to cooperate closely until the 12 samples were ordered according to quality and assigned a score from 1 (lowest) to 5 (highest). These samples served as the benchmarks for rating the remainder (see Appendix 3).

After the trial, the remainder of the samples were randomly divided into three piles and each pile was rated independently by two raters. A valid score was established when a consistent score was given by both raters. In case of discrepancy, a third rater was consulted in order to resolve the disagreement.

3.6 CAF measures

Several CAF measures were adopted to investigate the writing production from multiple dimensions (see 2.3 for the rationale behind the chosen measures). For complexity, we employed four measures: the Mean Length of T-unit (MLT), the Ratio of Dependent Clauses per T-unit (DC/T), the ratio of the presence of Present Simple Tense (%PST) against the tokens of all tense/aspect, and the Guiraud index. For accuracy, we counted the number of errors per 100 words: A high error ratio means low accuracy. We analyzed three types of error ratios, namely, tense/aspect related use and form error ratios and a non-verbal error ratio (i.e. errors irrelevant to tense/aspect). Regarding verb use/form errors, five types of tense/aspect which were used in the learners' writings were included for analyses: Present Simple, Present Continuous, Present Perfect, Past Simple, and Future Simple. Regarding non-verbal errors, six types of errors were included: lexicon, grammar, word order, spelling, mechanics, and punctuation use (Verspoor et al., 2012). For fluency, text length was employed as a measure of writing fluency. The operationalization of the eight CAF measures is summarized in Table 1 below.

Table 1 Summary of CAF measures and their operationalization

Variables	Operationalization
Mean Length of T-unit (MLT)	Number of tokens divided by number of T-units.
Dependent Clause per T-unit (DC/T)	Number of dependent clauses divided by number of T-units.
Ratio of the Present Simple Tense (%PST)	Number of the occurrences of Present Simple tense divided by tokens of tense/aspect.
Guiraud index	Number of lexical types divided by the square root of the number of lexical tokens.
Verbal use error ratio	Number of incorrect <i>use</i> of tense/aspect (including Present Simple, Present Continuous, Present Perfect, Past Simple, and Future Simple) divided by number of tokens, then multiplied by 100
Verbal form error ratio	Number of incorrect verb <i>form</i> of tense/aspect (including Present Simple, Present Continuous, Present Perfect, Past Simple, and Future Simple) divided by number of tokens, then multiplied by 100
Non-verbal error ratio	Number of non-verbal (including lexicon, grammar, word order, spelling, mechanics, and punctuation use) errors divided by number of tokens, then multiplied by 100
Text length	Number of tokens used in each text

3.7 Data coding

Except for text length which could be automatically counted in CLAN, most measures required manual annotation before useful statistics could be extracted and calculated. T-units, correct T-units, clauses, use of tense/aspect, and types of errors were manually annotated in CHAT format, automatically counted in CLAN, and then exported to Excel for calculation. For verb use errors, we coded the errors when the tense/aspect was overused. For example, when the Present Simple was used to talk about the past, it was coded as a Present Simple use error. For verb form errors, we coded the incorrect inflection of the verb phrases based on the writer's intended tense/aspect choice. Coding examples are given in Appendix 4.

The first two authors coded the texts together. The coding conventions were first discussed and agreed upon by the coders; subsequently, a collaborative trial coding of

ten texts was conducted to train the coders and resolve coding discrepancies between them. The remainder was divided and coded independently by the two coders. To ensure inter-coder reliability, all independent coding was examined by the other annotator and disagreements were discussed and resolved.

3.8 Analyses

Due to the normality issue and the scale issue of some datasets, Mann-Whitney U tests were run to compare the two groups for the pre-intervention listening and reading scores, the holistic rating scores, and the three error ratios. A one-way multivariate analysis of covariance (MANCOVA) was run to compare the writing complexity between the two groups, with the pre-intervention listening and reading scores as covariates. A one-way analysis of covariance (ANCOVA) was run to compare the text length between the two groups, with the pre-intervention listening and reading scores as covariates. Because the Mann-Whitney U test was run six times, the α level was adjusted to .008 (0.05 divided by 6) for that test. All statistic tests were conducted with the statistic software SPSS version 26.

4 Results

The results are presented in order of our four hypotheses (see Section 2). For the ease of reading, the results regarding complexity and fluency measures were combined. Firstly, before the intervention the two groups displayed similar English proficiency as measured by the listening and the reading scores; after the intervention the two groups achieved similar overall writing proficiency as measured by the holistic ratings of text quality. Mann-Whitney tests showed that no differences were found for the Pre-listening scores between the explicit group (Mdn = 92.5) and the implicit group (Mdn = 92.5), $U = 1545$, $p = .656$; no differences were found for the Pre-reading scores between the two groups (explicit Mdn = 92.5; implicit Mdn = 90.5), $U = 1580.5$, $p = .811$; and no differences were found for the Holistic Scores between the two groups (explicit Mdn = 3.0; implicit Mdn = 3.0), $U = 1478.5$, $p = .394$. The descriptive statistics of these three measures, as well as the effect size of each test are listed in Table 2 below.

To supplement the comparison of mean holistic scores (see Table 2), we also examined the distribution of the holistic scores across the two groups. As shown in Figure 1 below, while the modes of the holistic scores were 3 for the explicit group and 2 for the implicit group, the implicit group scored more frequently at higher scores (i.e. score 4 and 5) than the explicit group did, which contributed to a higher mean score for the implicit group, although this difference did not reach significance.

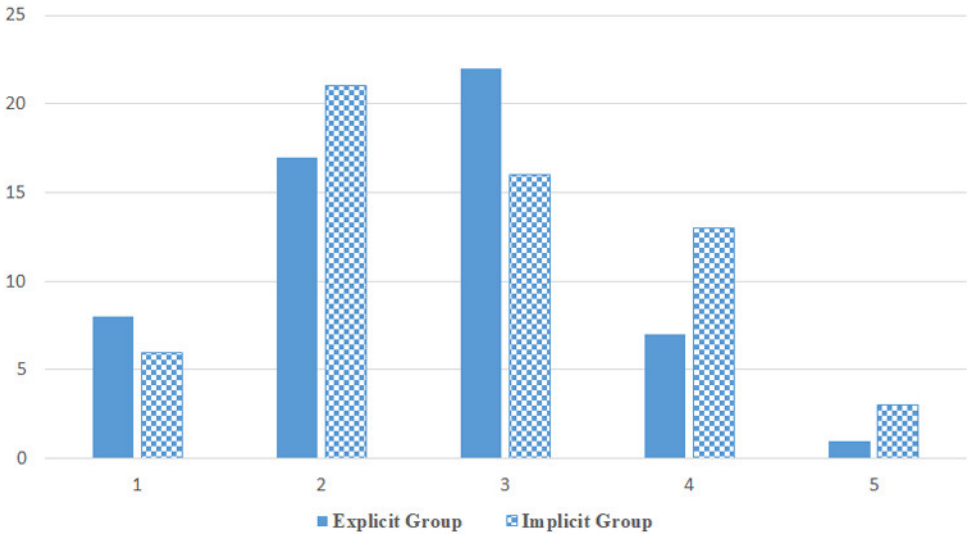


Figure 1 Frequency distribution of the holistic writing scores 1 to 5 across the explicit and implicit groups

Table 2 Descriptive statistics and the effect sizes of testing the proficiency differences between the two groups for the listening and the reading scores before the intervention and for the holistic scores of text quality after the intervention

Measures	Explicit group (N = 55)		Implicit group (N = 59)		Effect size
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
Pre-listening	92.32	5.95	91.82	10.94	$\eta^2 = 0.002$
Pre-reading	90.84	8.13	90.25	7.78	$\eta^2 = 0.001$
Holistic Scores	2.56	0.96	2.76	1.07	$\eta^2 = 0.006$

Secondly, the written texts of the two groups displayed neither differences in linguistic complexity, as measured by MLT, DC/T, %PST, and Guiraud index, nor in writing fluency, as measured by text length. A one-way MANCOVA test indicated no significant differences between the explicit and implicit groups on the combined writing complexity measures after controlling for the pre-intervention listening and reading scores, $F(4, 107) = 0.649, p = .629, \text{Wilks' } \Lambda = .976, \text{partial } \eta^2 = .024$. A one-way ANCOVA test indicated no significant differences between the explicit and implicit groups on text length after controlling for the pre-intervention listening and reading scores, $F(1, 110) = 2.118, p = .148$,

Table 3 Descriptive statistics and the effect sizes of testing the differences between the two groups for the four writing complexity measures and for Text Length

Measures	Explicit group (N = 55)		Implicit group (N = 59)		Effect size
	Mean	SD	Mean	SD	
MLT	8.39	1.40	8.32	1.11	
DC/T	0.19	0.12	0.20	0.10	$\eta^2 = 0.024$
%PST	0.51	0.13	0.55	0.17	
Guiraud	7.26	0.70	7.37	0.75	
Text Length	191.62	63.87	208.93	60.32	$\eta^2 = 0.019$

Table 4 Descriptive statistics and the effect sizes of testing the accuracy differences between the two groups for the verb use error ratio, the verb form error ratio, and the non-verbal error ratio

Measures	Explicit group (N = 55)		Implicit group (N = 59)		Effect size
	Mean	SD	Mean	SD	
Verb use error ratio	2.00	1.28	0.50	0.81	$\eta^2 = 0.414^*$
Verb form error ratio	0.26	0.42	0.76	0.77	$\eta^2 = 0.141^*$
Non-verbal error ratio	10.01	4.71	10.42	4.28	$\eta^2 = 0.003$

* Difference is significant at the .001 level (2-tailed)

partial $\eta^2 = .019$. The descriptive statistics of the complexity measures and text length, as well as the effect sizes of the two tests are listed in Table 3 above.

Lastly, in terms of accuracy Mann-Whitney tests indicated group differences for use and for form errors related to tense/aspect, but in different directions. The error ratio in the use of tense/aspect was greater for the explicit group (Mdn = 1.92) than for the implicit group (Mdn < 0.001), $U = 429$, $p < .001$, with a large effect size $\eta^2 = 0.414$, whereas the error ratio in tense/aspect forms was greater for the implicit group (Mdn = 0.57) than for the explicit group (Mdn < 0.001), $U = 956.5$, $p < .001$, with a large effect size $\eta^2 = 0.141$. No differences were found for non-verbal error ratio between the explicit group (Mdn = 9.4) and the implicit group (Mdn = 9.31), $U = 1519.5$, $p = .559$. The descriptive statistics of the three accuracy measures and the effect size of each test are given in Table 4 above.

5 Discussion and conclusion

Applying a free-response writing task to measure long-term learning outcomes, the current study aimed to revisit the efficacy debate of explicit versus implicit L2 instruction. We compared two very similar groups in terms of L1 background, educational context, scholastic aptitude and beginning level of L2 English. The young learners were instructed by the same teachers with the same textbooks. The only difference was that all the pages dealing with explicit grammar had been torn out of the books for the implicit learners. After two years of instruction, the learners were compared on a timed writing task in terms of general performance (holistic scoring) and an array of CAF measures. Except for the first hypothesis (i.e., explicit and implicit instruction would lead to equal writing quality measured by holistic ratings), the results do not support our hypotheses regarding CAF measures.

To summarize, despite the different instructional methods, the two groups developed similar overall writing quality, linguistic complexity, and writing fluency. Even though differences were found in writing accuracy measured as error ratios, neither instructional method showed a clear advantage over the other. While the explicit group showed higher accuracy in tense/aspect related verbal inflection, the implicit group were better at choosing the right tense/aspect to express meaning. In terms of non-verbal accuracy (i.e. accuracy in lexicon, grammar, word order, spelling, mechanics, and punctuation use), the two groups were equally accurate. All in all, the two types of instruction did not seem to make much of a difference in the learners' L2 writing development over the two years of intervention.

Regarding complexity, our findings are in line with Macaro and Masterman (2006), which suggests that explicit instruction does not contribute to the diversity or complexity in the learners' expression. Similarly, Andringa et al. (2011) found that explicit and implicit instruction contribute equally to the use of the target constructions after a computer-assisted L2 Dutch intervention. However, this finding is at variance with Piggott (2019), which included a larger sample of participants and included both the first- and the second-year writing data into a multilevel modelling analysis. The analysis revealed a significant interaction for Time*Group with the implicit group showing a steeper growth in complexity measures between Year 1 and Year 2. The current study focused on a homogeneous subgroup (participants from the highest scholastic stream), and compared only the second-year writing data, with the pre-intervention English proficiency (i.e. listening and reading scores) being controlled for. The results of a one-way multivariate analysis of covariance did not show difference between the explicit and implicit groups in terms of combined writing complexity measures. It might be that adding another data point (Piggott's study; the first-year writing data) revealed different developmental patterns of the two groups within the course of two years, and this difference might have levelled off in the second-year data (current study). Another factor, which might better explain the similar learning outcomes between the explicit

and implicit groups of learners, may lie in the English exposure outside the classroom. We will come back to this point later.

Regarding writing fluency as measured by text length, explicit and implicit types of instruction seem to be equally effective. This finding does not support our hypothesis or previous longitudinal investigations, e.g., Piggott (2019) and Rousse-Malpat et al. (2019), whose findings suggest implicit instruction to be more effective in promoting writing fluency. When interpreting the current finding, we need to apply some caution. Firstly, text length might not be an ideal measure for writing fluency especially when the task did not explicitly encourage the writers to write longer texts. Furthermore, although text length has been commonly used as a fluency measure, some researchers found high correlations between text length and other highly valued writing qualities, suggesting that text length has strong impact on both holistic and multi-trait scoring (Lee et al., 2010). Text length thus may better serve as a broad proficiency measure instead of a fluency measure. Measures at the phrasal level may be more informative for assessing writing fluency.

Moreover, as mentioned earlier, L2 exposure outside the classroom might have played a role in the learners' learning outcomes. This speculation may partially explain the different findings between the current study and Rousse-Malpat and colleague's (2019) three-year investigation. Unlike the current study which focused on Dutch young learners learning L2 English, Rousse-Malpat et al. focused on similar learners learning L2 French, and found advantageous effects of implicit instruction on both writing fluency and complexity. A crucial difference between the different populations may be the amount of out-of-school exposure to the target language, which is much higher for English (e.g., through television programs, games, and social media), than for other foreign languages such as French (Peters et al., 2019). Furthermore, as shown in Peters (2018), out-of-school exposure explains more variance than other factors do (e.g., the length of instruction) in the English vocabulary knowledge of Dutch-speaking teenagers. We thus suspect that whereas explicit/implicit instruction may contribute to significantly different learning outcomes of Dutch learners of L2 French, type of instruction may not matter as much to Dutch learners of L2 English, whose English attainment is strongly influenced by out-of-class exposure.

Finally, intriguing differences were found in the writing accuracy of the explicit and implicit groups. The two groups displayed equal accuracy in the non-verbal error ratio but different areas of strength in tense/aspect related verbal error ratios. While the explicit group were more accurate in verb forms, the implicit group were more accurate in verb use. These findings are in line with the differences in accuracy that Tilma (2014) found in a longitudinal study of Finnish as an L2, where there were no differences between the implicit and explicit learners in overall case accuracy rate after ten months of intervention, but here, too, the implicit learners made relatively fewer use errors. Taking a dynamic usage-based perspective on L2 development (Verspoor, 2017), we can explain this in terms of the associative learning of form-use-meaning-mappings. Whereas explicit instruction

directs the learners' attention mainly to form, implicit instruction – through frequency of exposure to meaningful utterances in contexts – allows the learner to associate the form with the appropriate contextual use.

It is worth noting that the differences in the overall verb use/form error ratios of the two groups might be largely due to the two groups' differing mastery of the Present Simple, given that the Present Simple constituted more than half of the tense/aspect tokens in the writings of both groups (as measured by %PST; see Table 2). In fact, when taking a look at the datasets, tense aspect types like the Present Continuous, the Present Perfect, and the Future Simple occurred very infrequently; as a result, for these tense aspect types the error ratios were highly skewed to zero. Although the writing task was designed to elicit a wide range of tense aspect forms, both groups seemed to favor the Present Simple in writing. This is in line with Verspoor et al. (2012), whose findings suggest that learners at lower levels tend to mainly rely on the Present Simple tense, and a more balanced use of different tense/aspect occurs only at higher proficiency levels.

Given the large weight of the Present Simple in the total use of different tense/aspect, our findings in verbal accuracy suggest that the explicit group seemed to overuse the Present Simple more than the implicit group did; however, when the Present Simple was used, the explicit group were better able to deliver the correct forms than the implicit group. It makes sense that explicit grammar instruction could have trained the explicit group to pay attention to form issues, with explicit learners erring less in verb forms as a result. This corresponds with the conclusions from the majority of type-of-instruction SLA studies in which grammatical accuracy measures were the main focus (Norris & Ortega, 2000; Spada & Tomita, 2010; Goo et al., 2015). On the other hand, a competition for attentional resources might occur when the learner had to choose both the right tense/aspect to express meaning and the correct form corresponding to the chosen tense/aspect. As the explicit group would give more attention to producing a form correctly, they might not pay enough attention to choosing the correct use, (over)using their most familiar tense aspect instead – the Present Simple. The implicit group, on the contrary, might have experienced less pressure to produce correct forms but learned to directly focus on meaning. The instruction might have helped the implicit group to be more aware of the context and make a more accurate choice of which tense/aspect to use.

To conclude, our longitudinal study shows that explicit and implicit types of instruction have a similar impact on the Dutch-speaking secondary school learners of L2 English in terms of their general writing proficiency, linguistic complexity, non-verbal accuracy and writing fluency. As far as tense/aspect related writing accuracy is concerned, the two types of instructional methods demonstrate different strengths and weaknesses. While explicit instruction promotes an accurate command of verb form, implicit instruction fosters a contextually appropriate use of tense and aspect.

Implications drawn from this study are that L2 instruction does make a difference for language development as it directly taps into different language learning patterns and

knowledge (Norris & Ortega, 2000). In addition, although varying types of instruction would lead to similar effectiveness with respect to overall proficiency, the effectiveness differs when it comes to specific linguistic aspects. However, the deeply rooted beliefs in the importance of explicit grammar instruction, especially for accuracy concerns, may be overstated. For L2 classroom teaching, therefore, teachers can be confident that students would develop good L2 attainment even when explicit grammar instruction would not be provided – at least in a context with a substantial amount of out-of-school exposure to the target language.

The present study is not without limitations. Firstly, we only used one writing task to assess writing proficiency, while overall writing proficiency should ideally be measured using multiple writing tasks (Schoonen et al., 2011) to elicit a wider range of language use. Secondly, the number of measures taken to examine L2 writing proficiency was limited; for fluency in particular, text length may be too broad to provide sufficient insight into writing fluency. It was also possible that the effectiveness of instruction was underestimated due to a possible ceiling effect in some of the measures used. For an expansion of this study and future research, we suggest that more fine-grained measures should be employed to obtain a comprehensive understanding of L2 writing. Furthermore, it would be interesting to engage learners of different age and proficiency groups in a classroom setting to probe into the effects of types of instruction. Finally, the impact of out-of-school exposure versus the impact of instruction method could warrant further investigation.

Acknowledgements

The data analyzed in this study was part of a doctoral project funded by the Dudoc-Alfa Program (the Netherlands); the scholarship was awarded to Leslie Piggott. Without this fund and the generous participation of many teachers and students, a series of studies including this one would not have been possible. We would like to thank our professional raters for their time and support. We are also grateful to the anonymous reviewers, the editor, and the applied linguistics support teams both in the University of Groningen and Utrecht University for their constructive comments and suggestions at various stages of this study.

Notes

- 1 There were in total 60 students in the implicit group, but one student lacked the pre-intervention proficiency scores, thus was excluded in the analysis.

References

- Abdel Latif, M.M.M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, 34(1), 99–105. <https://doi.org/10.1093/applin/ams073>
- Andringa, S., De Gloppe, K., & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning*, 61(3), 868–903. <https://doi.org/10.1111/j.1467-9922.2010.00623.x>
- Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in Second Language Acquisition*, 37(2), 185–196. <https://doi.org/10.1017/S027226311500008X>
- Boyd, E., & Barraclough, C. (2010). *Activate! A2 student's book*. Pearson Longman.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26(1), 42–65. <https://doi.org/10.1016/j.jslw.2014.09.005>
- DeKeyser, R.M., & Botana, G.P. (Eds.). (2019). *Doing SLA research with implications for the classroom: Reconciling methodological demands and pedagogical applicability*. John Benjamins Publishing Company.
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Multilingual Matters.
- Ellis, R., & Barkhuizen, G.P. (2005). *Analysing learner language*. Oxford University Press.
- Gass, S.M., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition: Introduction to the special issue. *Modern Language Journal*, 82(3), 299–307. <https://www.jstor.org/stable/329956>
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–483). John Benjamins Publishing Company.
- Hammerly, H. (1987). The immersion approach: Litmus test of second language acquisition through classroom communication. *The Modern Language Journal*, 71(4), 395–401. <https://doi.org/10.2307/328469>
- Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics*, 10(3), 331–360. <https://doi.org/10.1093/applin/10.3.331>
- Hou, J., Verspoor, M., & Loerts, H. (2016). An exploratory study into the dynamics of Chinese L2 writing development. *Dutch Journal of Applied Linguistics*, 5(1), 65–96. <https://doi.org/10.1075/dujal.5.1.04loe>
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10.1093/applin/amp048>
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.
- Hulstijn, J.H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-

- language learning: Introduction. *Studies in Second Language Acquisition*, 27(2), 129–140. <https://doi.org/10.1017/S0272263105050084>
- Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12(4), 377–403. <https://doi.org/10.1016/j.jslw.2003.09.001>
- Krashen, S.D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Krashen, S.D., & Terrell, T.D. (1983). *The natural approach: Language acquisition in the classroom*. Pergamon Press.
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 579–589. <https://doi.org/10.1093/applin/amp043>
- Lee, Y., Gentile, C., & Kantor, R. (2010). Toward automated multi-trait scoring of essays: Investigating links among holistic, analytic, and text feature scores. *Applied Linguistics*, 31(3), 391–417. <https://doi.org/10.1093/applin/amp040>
- Lightbown, P.M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, 12(4), 429–448. <https://www.jstor.org/stable/44488335>
- Long, M.H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). John Benjamins Publishing Company.
- Macaro, E., & Masterman, L. (2006). Does intensive explicit grammar instruction make all the difference? *Language Teaching Research*, 10(3), 297–327. <https://doi.org/10.1191/1362168806lr1970a>
- MacWhinney, B. (2014). *The CHILDES project: Tools for analysing talk, volume I: Transcription format and programs*. Lawrence Erlbaum Associates.
- Norris, J.M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Norris, J.M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–577. <https://doi.org/10.1093/applin/amp044>
- Oh, S. (2006). *Investigating the relationship between fluency measures and second language writing placement test decisions*. [Unpublished master's thesis, University of Hawai'i, Honolulu, the United States]. <http://hdl.handle.net/10125/20203>
- Ortega, L., & Byrnes, H. (2009). *The longitudinal study of advanced L2 capacities*. Routledge.
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26–45. <https://doi.org/10.1017/S0267190505000024>
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL – International Journal of Applied Linguistics*, 169(1), 142–168. <https://doi.org/10.1075/itl.00010.pet>
- Peters, E., Noreillie, A., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and

- out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, 69(3), 747–782. <https://doi.org/10.1111/lang.12351>
- Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, 44(3), 493–527. <https://doi.org/10.1111/j.1467-1770.1994.tb01115.x>
- Piggott, L. (2019). *First meaning then form: A longitudinal study on the effects of delaying and reducing form-focused instruction for young adolescent EFL learners*. LOT (Landelijke Onderzoekschool Taalwetenschap)
- Polio, C. (2001). Research methodology in second language writing research: The case of text-based studies. In T. Silva, & P.K. Matsuda (Eds.), *On second language writing* (pp. 91–115). Lawrence Erlbaum Associates, Inc., Publishers.
- Puchta, H., & Stranks, J. (2008). *More! 1 student's book*. Cambridge University Press.
- Rousse-Malpat, A., Steinkrauss, R., & Verspoor, M. (2019). Structure-based or dynamic usage-based instruction: Long-term effects on (morpho)syntactic and lexical complexity in writing samples. *Instructed Second Language Acquisition*, 3(2), 181–205. <https://doi.org/10.1558/isla.38054>
- Rousse-Malpat, A., & Verspoor, M. (2012). Measuring effectiveness in Focus on Form versus Focus on Meaning. *Dutch Journal of Applied Linguistics*, 1(2), 263–276. <https://doi.org/10.1075/dujal.12.07rou>
- Schoonen, R., Van Gelderen, A., Stoel, R.D., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, 61(1), 31–79. <https://doi.org/10.1111/j.1467-9922.2010.00590.x>
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30(2), 73–87. <https://doi.org/10.1017/S0261444800012799>
- Spada, N., & Lightbown, P.M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly*, 42(2), 181–207. <https://doi.org/10.1002/j.1545-7249.2008.tb00115.x>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(4), 532–553. <https://doi.org/10.1093/applin/amq001>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S.M. Gass, & C.G. Madden (Eds.), *Input in second language acquisition* (pp. 165–179). Newbury House.
- Tilma, C. (2014). The dynamics of foreign versus second language development in Finnish writing. *Jyväskylä Studies in Humanities* (233).
- Verspoor, M. (2017). Complex dynamic systems theory and L2 pedagogy. In L. Ortega, & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 143–162). John Benjamins Publishing Company.

- Verspoor, M., Schmid, M.S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239–263. <https://doi.org/10.1016/j.jslw.2012.03.007>
- What is AIM? (n.d.). <https://www.aumlanguagelarning.com/what-is-aim/>.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press.

Appendix 1: The writing test used for data collecting

Time: 50 minutes

No dictionary

Write a postcard

Think about your last holiday.

Now pretend you're still on this holiday.

You are writing a postcard to your best friend at home.

Include the following information in your postcard:

- Waar ben je?
- Hoe ben je daar gekomen (auto/vliegtuig etc.)?
- Waar verblijf je?
- Omschrijf de kamer waarin je slaapt. Wat staat er allemaal in.
- Schrijf over een activiteit die je al hebt gedaan en wat je nog gaat doen.
- Vertel wat je tot nu toe het leukst vond van je vakantie.

Don't forget a correct greeting and ending for your letter.

Write a minimum of 140 words.

Appendix 2: Data cleaning when converting texts to CHAT files

Decision	Reason	Example
Delete Dutch translations in brackets.	redundant	My little sister and I jumped atmediately (meteen) in the swimming pool.
Delete interjections	Cannot be recognized, also redundant	omg -I can't believe I'm really here.
Delete headers or subtitles in the texts.	Cannot be read as utterances	1. Austria
Change ellipsis (...) into comma or full stop according to the context.	CLAN cannot recognize ellipsis	I don't want to go home yet ... But the summer is almost gone.
Delete brackets used in the texts; add a comma or period to the utterances if needed.	CLAN cannot analyse text in brackets	I'm in the USA (like you didn't now).
When a sentence ends with a colon, change the column to a period.	Utterances must end with a period in CLAN	I'm gonna tell you how it is here:

Appendix 3: Benchmarks of the holistic scores 1 to 5 and the description of the text quality for each score

Score	Benchmark	Description
1	I'm at the Place. I'm going to it with a plane of KLM. I'm living and sleeping there in a villa before me and my mother, father, brother and sister. The room where I sleep is big, there a 2 beds, 1 before me and 1 before my brother. Right and left of the beds are nighttables before your telefoon, for example. There is 1 closet with many place before me and my brother. I was going to the houses of Place, in the light and in the night, because then are the houses lighted with really nice lights. I'm going to a really nice beach when the sun is shining. Because than is needing the weather sunny. The nicest what I have done is swimming in us swimming pool fore me and my mother, father, brother and sister alone. That was really nice because is was a private swimming pool before us.	Simple and short sentences; mainly use present or past simple tenses; fundamental errors; use many Dutch words, very few if any idiomatic phrases; shot texts, or relatively longer but with a lot of repetition; difficult to follow the story
2	How are you doing? With me it's fine. I'm in Italy, and it's realy hot here! We came here with the car, I hated the long trip because I had to sit in the middle ... We are staying at a campsite, and it's very pretty here. The campsite is on a small mountain not far from the sea, there are some verry large trees here as well. We sleep in a sort of tent from the campsite. It isn't big, and the rooms are sepreat with a thin sort of blanket. I sleep in an small bed next to my brother's bed. There is not that much spase so there are only two beds in there. I've bin to the sea two times and it was a lot of fun! The second time I went to the sea there were a lot of waves and we couldn't swimm far, but that was ok! I sat in a wave and it splashed over me, it was quite fun. About two days we are going to go to a little town not far frome here, we want to see the big church there and I want to buy an magnet for my magnet collection. The funniest thing I did was, I think, playing in the sea with the waves and it was fun that I made a new friend! For now I think it's enough.	Mostly simple sentences, some complex sentences; attempt at past, present, continuous tenses; some creative use at lexicon; many errors induced by attempt at complexity, but do not impeded understanding; longer texts; many Dutchism; attempt at storytelling, but no real flow
3	I miss you and home because its very boring here. I'm now in Germany with my parents and my brother, my sister is in the Netherlands with my grandparents. Yesterday morning we drove to Place and we slept in a hotel in the city. I slept in a little room without a tv or any other electric stuf and my bed is PINK. We didn't do anything that was fun but ... tomorow we go to the game between Name and Name. I hope that Namegonna win because when they win they are the chaimpon. Were staying to next week and then were going back to home. I hope that you have a better holiday then me. I also hope that you have better weather then us because we only had rain yet. I see you soon!	Some complex sentences; lexicon command rather basic; simple tenses mostly correct; some errors, but no fundamental errors; a few idiomatic expressions; telling a rather solid story

(cont.)

Score	Benchmark	Description
4	I am on holiday in Place. We came with our car. It was around 250 km and it took us like 3,5 hour to get here. We are staying in a really nice hotel, but it is not very big. Our room is not very big either. There are four beds, a table and two chairs. There is no television. Yesterday we did something I really enjoyed. We were cycling inside a cave in Place. We had to go 50 m under the ground. It was really cold inside the cave, so we had to wear some thick clothes. we also had a lot of light on our bikes, because it was really dark. Tommorrow we are going to a cave in Place in Belgium. I am really looking forward to it! My favourite activity was cycling inside a cave in Place, because I really like cycling and never did something like that. How is your holiday going? I hope you are enjoying it! Write me back soon!	Complex sentences; various use of tenses; no major errors; some good idiomatic expressions; solid story telling
5	I'm on an exciting holiday in the great city of Place on the beautiful Australian island. I went down under with an airplane. Now I am staying in an hotel with my parents and my brother and sister. I have to share a room with my brother, the room is very big, especially for two persons. In the room we have a tv, a bed, a sofa and a bathroom with a shower and a toilet. On the second day we went to a place where we could bungee jump. My little brother and my mother would not bungee jump. Two or three seconds before I jumped I saw how high it was. That was scary, but when I finally jumped it was good fun. Tomorrow we will go to a great stadium in Place to watch a rugby game between Australia and New-Zealand. That's gonna be cool. In the second week we will drive to Place to visit the Place and the rest of the city. After that week we will fly back to the Netherlands. I'll see you soon.	Some complexity; good sentence structure; good lexical command; no major errors; lots of idiomatic phrases; solid story and nice flow

Appendix 4: Examples of manual coding

Target measures	Example	Coding
T-unit	in the bathroom is a bath!	T
Sentence type and clauses	we're staying in an beautyfull hotel <u>that is super close to the most important street in Place: Name.</u>	Complex; relative clause
Tense aspect	we're staying in an beautyfull hotel that is super close to the most important street in Place: Name.	Present continuous; present simple
Non-verbal errors	we're staying in <u>an</u> [*] <u>beautyfull</u> [: beautiful] [*] hotel that is super close to the most important street in Place: Name.	Grammar; spelling
Verb use error	today we go [*] to the shopping mall to buy new clothes.	Present simple use error
Verb form error	I thinks [*] he liked me.	Present simple form error