JESLA | Journal of the European Second Language Association

Long-term effects of structure-based versus dynamic usage-based instructional programs for French

RESEARCH

WIM GOMBERT
MEREL KEIJZER
MARJOLIJN VERSPOOR

*Author affiliations can be found in the back matter of this article

WHITE ROSE UNIVERSITY PRESS
Universities of Leeds, Sheffield & York

ABSTRACT

This study compares the effects of two instructional programs on speaking, listening, reading and writing skills after six years of instruction. The first is based on dynamic usage-based (DUB) principles, with a great deal of exposure and meaningful interaction and implicit attention to form. The second program is based on structure-based (SB) principles, with a great deal of explicit attention to form and forms. Rousse-Malpat et al. (2022) showed that after three years, a DUB program is more effective than an SB program on speaking and writing skills, but most teachers in their study switched to a SB approach because they were convinced this is necessary to meet Dutch central exams (Voogel, 2018). One teacher, however, continued with the DUB approach and compared several cohorts of SB and DUB students on both the standardized reading and listening tests and speaking and writing skills. The results of the current study show that the SB and DUB students perform similarly on reading, but DUB students outperform the SB on listening. In writing, the SB and DUB students perform similarly on holistic scores and accuracy, but the DUB students outperform the SB students in sentence length, text length, and the use of formulaic language. As expected, the DUB students outperform the SB students by far on speaking. The findings suggest that a communicative program with an emphasis on exposure, interaction and speaking does not have to be at the expense of reading and writing skills.

CORRESPONDING AUTHOR:
Marjolijn Verspoor
University of Pannonia, Hungary; University of Groningen, the Netherlands
m.h.verspoor@rug.nl

KEYWORDS:
instructional programs; dynamic usage-based; structure based; listening; speaking; reading; writing

TO CITE THIS ARTICLE:
Gombert, W., Keijzer, M., & Verspoor, M. (2024). Long-term effects of structure-based versus dynamic usage-based instructional programs for French. *Journal of the European Second Language Association*, 8(1), 18–33. DOI: https://doi.org/10.22599/jesla.118

DOI: 10.22599/jesla.118

1. INTRODUCTION

Rousse-Malpat et al. (2022) showed that second-language (L2) learners of French in Dutch pre-university education, who followed a communicative language program with lots of exposure and meaningful interaction, performed significantly better on speaking and writing skills after three years of instruction than students who had received a version of Communicative Language Teaching (CLT) with a great deal of attention to explicit grammar. The latter CLT program was in line with Structure-Based (SB) principles in that it focused on explicit rule learning to avoid errors, all taught through the medium of the first language (L1). The other CLT program was in line with Dynamic Usage-Based (DUB) principles in that it involved a great deal of exposure, imitation, repetition, learning the meaning of every single word through gestures and presenting whole chunks of language, all without explicit grammar teaching but with corrective feedback when needed. However, despite the positive results after the first three years, two of the three schools involved changed to an SB program for the last three years, mainly because teachers are convinced that students have to "know about" the language to do well on the Dutch central exam, which consists of a centralized multiple-choice reading test and a decentralized listening comprehension test.

However, one school that participated in the Rousse-Malpat et al. (2022) study continued with DUB principles in years 4 through 6 (the Dutch pre-university secondary school curriculum comprising 6 years) to test whether explicit teaching and training in the spirit of SB principles are truly necessary to prepare students well for the final exams. In the current study, SB and DUB L1 Dutch learners of French from the same school and taught by the same teacher in years 4 through 6 are compared at the end of their studies on reading, listening, writing and speaking.

2. BACKGROUND

2.1. L2 FRENCH TEACHING IN THE NETHERLANDS

Lightbown and Spada (2013) point out that approaches with a strong SB component remain widespread among L2 teaching practices all over the world. Specifically focusing on a Dutch context, Graus and Coppen (2016) investigated student and teacher beliefs in the Netherlands regarding the role that grammar and grammar instruction should play in the English L2 classroom and found that participants (both teachers and students) considered explicit, systematic, and isolated grammar instruction a necessary condition not only for linguistic correctness but also for advanced communicative competence. West and Verspoor (2016) examined L2 teaching practices in the Netherlands through classroom observations and found the main characteristics to be an explicit focus on grammar, a frequent use of the L1 as a medium to teach (about) the L2, the use of translation, learning vocabulary with translation equivalents and an emphasis on written language. Popma (1997) and Hermans-Nymark (2006) concluded that commonly used CLT coursebooks in the Netherlands generally reflect a SB design. The graded acquisition of selected structural and lexical items is seen as necessary instruments for successful communication and constitute the backbone of most coursebooks, which often adopt an explicit and deductive approach to grammar, emphasizing lexical and grammatical accuracy. The same is true for French as L2 programs, but at the same time many teachers regret that the focus on grammar and reading may be at the expense of the development of oral skills (speaking and listening), as very little class time is left to practise such skills (Voogel, 2018).

2.2. SB VERSUS DUB THEORY

To illustrate the difference between an SB and a DUB theoretical perspective, we take one sentence from a short French narrative (with idiomatic English translation) as an example.

(1) *Il était une fois une maman cochonne qui avait trois petits cochons.*
Once upon a time there was a mama pig who had three little pigs.
Il (Subject-pronoun)
était (predicator-third person singular-past tense)
une fois (adverbial-noun phrase-feminine noun)

DOI: 10.22599/jesla.118

une maman cochonne [*qui/avait/trois petits cochons*] (Subject Attributer-noun phrase (head feminine noun) modified by a relative clause)

From a SB perspective, the sentence could be deconstructed and broken down into major constituents and analysed further from the syntactic to the morphological level, focusing on constructs such as gender, agreement and tense. Implicit to such a view of language is the consideration that a speaker builds up such a sentence by applying grammatical rules while producing it and that, over time and as L2 proficiency increases, this process becomes (more) automatic (cf. DeKeyser, 1997).

In fact, a large body of SLA research seems to support the need of an SB approach with explicit instruction. Several meta-analyses (e.g., Norris & Ortega, 2000; De Graaff & Housen, 2009; Spada & Tomita, 2010) showed that explicit types of instruction were more effective than implicit types of instruction in terms of L2 development. However, more recently, Kang et al. (2019) meta-analysed 54 empirical studies conducted between 1980 and 2015 and found that explicit and implicit instruction were equally effective in immediate post-tests, but implicit instruction was more effective in delayed post-tests. They suggest that the larger relative number of studies using free-response measures might have contributed to this result, supporting Doughty's (2003) concern that testing instruments in the earlier studies have favoured explicit treatments.

Even though SB views do not deny the existence of meaning and language use, the focus is first and foremost on grammatical form, with meaning and use added as separate components. In contrast, a DUB view starts from "constructions" that are first and foremost meaningful units, as shown in (2).

(2) *Il était une fois une maman cochonne qui avait trois petits cochons*.
Il était une fois (a fixed phrase that is used to introduce a fairy tale)
une maman cochonne (a being)
qui avait (expressing some possessive relationship)
trois petits cochons. (some beings)

Usage-based theory emphasizes input frequency and associative learning, reflective of learners using simple learning processes to statistically generalize over masses of input data (Ellis & Wulff, 2015), as well as output and practice. DUB theory is based on a combination of complex dynamic systems theory (Larsen-Freeman, 2007) and usage-based linguistic theory (see Schmid, 2020, for details). Complex dynamic systems theory emphasizes the individual developmental trajectories that show fluctuations and variability throughout the language-development trajectory (Verspoor, 2017).

The premise underlying DUB approaches (Verspoor, 2017) to language learning is that a speaker uses sequences of forms that have been used and that have been encountered in similar contexts (use) with a similar meaning and are learned by association, in particular syntagmatic association, which is basically the co-occurrence of words. The form-use-meaning mappings that are encountered most frequently and saliently are the ones that are typically learned first, become entrenched in the mind and are eventually produced automatically. Some of these sequences are rather fixed (as in chunks or other multi-word sequences), but others have open slots, and the construction can form a template for new-to-be-acquired sequences and constructions (as in verb-argument constructions).

Although DUB theory had not been established when CLT approaches emerged in the 1970s, its principles are very much in line with versions of communicative approaches that focus on meaning or form (Long, 2000). As in Krashen's theory (Krashen, 1992), the foundation of DUB theories of L2 development lies in repeated exposure to meaningful and contextualized language in mostly implicit instructional designs. A difference is that in Krashen (1992), the main focus is on the quality of the input (comprehensibility), but DUB theories emphasize iteration in both exposure and use so that constructions, including whole phrases and chunks, reoccur regularly. The fact that more L2 exposure and use have an effect on the acquisition of such chunks was found by Gustafsson and Verspoor (2017), who showed that high-input learners developed a significantly greater range of chunk types and a greater proportion of chunk-words per text compared to low-input learners.

DOI: 10.22599/jesla.118

Unfortunately, other than Rousse-Malpat et al. (2022), there are not many long-term classroom experiments (cf. DeKeyser & Botano, 2019) supporting DUB approaches with equal amounts of instruction in the L2 and implicit grammar teaching. One exception is Piggott et al. (2020) involving 416 Dutch learners of L2 English, who investigated the effectiveness of a two-year intervention with explicit grammar instruction versus an intervention without such explicit grammar instruction but with more authentic target exposure instead. The results showed that the effectiveness of the intervention method was directly associated with different aspects of language performance: After one year, the CLT group with explicit grammar outperformed the CLT group with implicit grammar on language accuracy measures, but this difference disappeared in the second year. From the very start, the CLT group with implicit grammar was more effective in terms of oral complexity and fluency measures, operationalized as text length and chunks.

3. THE CURRENT STUDY

The aim of the current study is to compare two groups of learners with different instructional approaches based on SB or DUB principles after six years of schooling. This study, thus, crucially adds an unparalleled longitudinal design in a landscape where mostly short-term interventions are compared, allowing for a better understanding of the differential outcomes of SB versus DUB approaches to language teaching. We focus on all four language skills (reading, listening, speaking and writing). The general assumption underlying the study is that students do better at the particular skills focused on in their respective classes. The research questions and hypotheses are as follows:

1. How do the SB and DUB students compare on a standardized final reading exam?

   The hypothesis is that the SB students outperform the DUB students, as a great deal of time in the final years is spent on reading exam texts and training to answer the multiple-choice questions.
2. How do the SB and DUB students compare on a standardized final listening exam?

   The hypothesis is that the DUB students outperform the SB students, as they had been exposed more to authentic spoken French from year 1 during their classes, even though the SB students had practised more regarding the format of the listening exam and strategies to employ.
3. How do the SB and DUB students compare on a free-response writing test?

   The hypothesis is that the SB students outperform the DUB students in accuracy, as they had received a great deal of explicit instruction throughout their French classes and had received feedback on accuracy when they practised formal letter writing. On the other hand, the DUB students are predicted to be more fluent and use more formulaic language, as they had been exposed to and had produced a great deal of written and spoken texts.
4. How do the SB and DUB students compare on a free-response speaking test?

   The hypothesis is that the DUB students outperform the SB students, as they had practised speaking freely from year 1 and the SB students had had very little speaking practice.

3.1. METHOD

The study design is cross-sectional comparing two groups after six years of pre-university SB or DUB instruction.

3.1.1. Participants

In order to measure effectiveness after six years of SB or DUB instruction, 133 students (36 men; 97 women) from five different cohorts who opted to continue with L2 French as an optional subject in years 4–6 of their secondary schooling were included in this study (SB: N = 55-SB, DUB: N = 73). Their L1 was Dutch, and all learners received instruction at the same pre-university educational level. At the age of 12, all participants had a similar scholastic aptitude, as reflected by a score obtained by the Dutch national curriculum test (so-called *Cito*) at the end of primary education. All students started their six years of instruction at the age of 12 in 2009 to 2013 as absolute beginners of French and completed their final exams at the age of 18

DOI: 10.22599/jesla.118

in 2015– to 2019 at a level of B2 (Common European Framework of Reference for Languages), based on final curriculum assessment norms. Students with dyslexia were excluded from the current study.

Because of the detailed hand-coding work involved in the written materials, the number of learners were limited to two different cohort groups (the last SB year 2016 and first DUB year 2017). In essence, this was convenience sampling, as the writing study was the first to be conducted. The SB group consisted of 24 learners (5 men; 19 women) who started learning French in 2010 and graduated in 2016, and the DUB group consisted of 32 learners (6 men; 26 fwomen) who started in 2011 and graduated in 2017. All learners' writing abilities in French were tested in their final year.

3.1.2. SB versus DUB teaching methods

The SB cohorts in the current study were taught using two typical coursebooks used in the Dutch secondary school context for French as a foreign language: *Grandes Lignes* (Bakker et al., 2005) and *Libre Service* (Breek et al., 2003). These coursebooks are called methods because they contain ready-to-use instructional materials that help the teacher execute the procedures that have been designed in accordance with the theoretical principles of a SB approach. *Grandes Lignes* was used during the first three years and *Libre Service* in the final three years for the SB-taught students. Each unit consists of a short text to be read for comprehension, a list of expressions and words from the text with their Dutch translations, several exercises in which students are asked to interact with each other using (pre-formulated) questions based on the vocabulary items and text and a substantial number of exercises in which grammar is offered and practised, with a strong focus on verb forms. In each unit, additionally, a great deal of time is spent on activities targeting the development of comprehension strategies (through reading and listening exercises). These reading and listening texts provide exposure to linguistic items like (thematic) vocabulary and grammar rules. Following these activities, certain linguistic items (words and rules) are thus highlighted and practised, and isolated word lists have to be memorized. The acquired knowledge is regularly tested, often by means of discrete-point tests. Most activities on comprehension strategies and linguistic knowledge are conducted in the L1 (Dutch), as it is considered necessary for the students' comprehension and henceforth is thought to make learning more effective (van Compernolle, 2015).

The SB students in our study also were required to write formal letters, which were graded and discussed mainly in terms of linguistic accuracy. On the whole, L2 exposure and active use in the SB condition were rather limited and most in-class teaching, and interaction was also done through the medium of the L1 and classroom procedures, which not only aimed at developing explicit (grammatical and lexical) knowledge but also at explicitly training comprehension strategies, as the central exam is limited to testing reading comprehension.

The DUB cohorts were taught using the Accelerative Integrative Methodology (AIM), a French as a Foreign Language method created by Wendy Maxwell (www.aimlanguagelearning.nl). Unhappy with the method she used as a French teacher in Anglophone Canada in which learners were not actually practising the language very much, she developed this method, which could be considered an exposure-rich version of CLT and in line with DUB principles in that it provides a great deal of meaningful interaction in the target language and uses visual cues to accompany forms as scaffolds. It focuses on form-use-meaning mappings (i.e., short phrases with a specific meaning in a specific context, see section 2), builds in a great deal of repetition through playful drills and focuses on form in the broadest sense: Pronunciation, multi-word expressions and other common sequences. In the class drills and raps, the teacher may focus on a particular form and correct it where needed but this often happens on the basis of visual cues (i.e., a gesture that is associated with a given item such as a word). AIM methods were originally designed for younger learners and thus typically do not span beyond the first three years of secondary school. Therefore, AIM extended (AIMe) was created based on similar AIM principles to be used in the final three years of pre-university education. As the DUB learners had achieved enough basic vocabulary with the help of gestures in their first three years, they were not really needed anymore for comprehending new words. Moreover, the lessons now consisted of meaningful, multi-modal authentic exposure by means of video materials from French news channels to be watched several times for content at home and were discussed in detail in class. The students were also asked to write essays based on the video materials,

DOI: 10.22599/jesla.118

which were graded and discussed mainly in terms of argumentation, sentence complexity and word choice, starkly contrasting with the SB group, where the focus was mainly on accuracy.

3.1.3. Amount of instruction

In both the SB and DUB condition, the total amount of instruction time was the same: Two 50-minute lessons per week in the first three years and three 50-minute lessons per week in the final three years. In addition, students completed homework assignments, providing an additional 110 hours of input (roughly 20 minutes of homework per 50-minute class). As exposure to French outside the classroom is mostly minimal to non-existent in the Netherlands, the total amount of instruction time over the six years can be estimated at 730 hours.

Also in the last three years, the amount of time spent on homework was the same for both groups, but the type of input and exercises that they did at home were different. The SB-taught students mainly practised the grammatical rules presented in class and were asked to complete additional reading comprehension exercises. The DUB students, by contrast, were asked to listen to French media and read authentic French magazines without completing reading comprehension questions or grammar drills.

However, as traditional SB instruction (on the basis of the course books indicated earlier) did not include free writing or speaking activities at all and we needed to make sure the students in both SB and DUB conditions would be able to talk and write about similar topics, a 30-hour intervention was designed in which the learners in both conditions had an equal amount of L2 exposure and speaking and writing practice. Because this intervention with DUB features took only about 4% of the total of the SB instructional time, it was thought to affect the outcomes of the study minimally.

The intervention program involved seven academic topics such as migrants, tattoos, and abortion that were chosen because the participants were thought to have an opinion on these topics and be able to talk and write about them. During 6 to 7 lessons, a topic was introduced using a video-documentary of two minutes on average per topic. This introduction was followed by listening and reading exercises, which entailed repeated exposure to the language used in the video documentary, prompting the learners to focus on the vocabulary needed to speak about the topic later on. Gradually, exercises became more productive. Learners were encouraged to think of arguments for and against a given view related to the topic under discussion and to use these arguments as part of several guided tasks, making sure all students were able to understand and produce meaningful content relating to these topics. Finally, free-response tasks like a debate, a press conference and a discussion were organised to prepare students to produce language spontaneously during the final speaking and writing tests.

3.1.4. Tests and scoring: Listening and speaking

In the Netherlands, there are required central exams developed by *Cito* (the national testing agency) in the final year of secondary school for reading and listening. The reading exams consist of closed-type questions (true/false, cloze, etc. but mostly multiple choice) and a few open-ended questions. The final listening tests consist exclusively of multiple-choice questions. The tests are not the same for all cohorts, but as Cito has an extensive procedure in place that includes pre-testing exam questions and conducting statistical analyses on the questions to make sure exam scores represent the same knowledge across years (to ensure safeguarding the national exams), the raw sores are normalized and official exam grades are provided (Cito, no date). The official scores were taken as the basis for our analysis. It should be noted, however, that these procedures ensure that by and large the level of difficulty of the national exams for French are highly comparable across different years.

3.1.5. Tests and scoring: Writing

A week before the writing exam took place, the students were asked which four topics of the seven from the 30-hour intervention were their favourites, and they were then asked to prepare for those four in terms of argumentation and expressions to be used. At the time of the test, two of those four topics were presented to the students (students did not know which two would be presented) and students were able to select one. We felt that this presented a good

DOI: 10.22599/jesla.118

mix between being able to prepare and having to spontaneously produce written language. As this procedure was kept constant for all groups, we felt that the difference in outcome effects could still be ascribed to one of the two teaching approaches.

Students were asked to write an essay of a minimum of 200 words within 45 minutes on the topic of their choice. They wrote the essay in the computer lab at school and handed in their assignment electronically. Supportive tools (e.g., a spelling- and grammar checker) were not turned off during the test, but they were hardly used given the time restrictions (the teacher was able to monitor all screens by using specialized software). The anonymised essays were assessed in four ways: Through holistic ratings by expert teachers, by means of machine-mediated morphosyntactic profiling, by means of analytical complexity, accuracy and fluency (CAF) measures (see below) and by means of hand-coded formulaic language.

To rate the texts holistically, the same comparative judgment method was used as in Verspoor et al. (2012). A group of nine French teachers were asked to rank 10 texts in terms of general proficiency in several rounds until consensus was reached on ratings. These texts were then used as benchmarks. The remainder of the texts were scored by three raters each. Agreement among the raters was high ($r = 0.893$, $p < 0.001$).

To assess the texts on morphosyntactic stages of development, Direkt Profil (DP), one of the few corpus tools available for L2 French, developed at the University of Lund in Sweden (Granfeldt et al., 2018), was used. The software bases its analysis on 142 different analytic text measures and profiles the morphosyntactic content of any (learner) text including subject-verb agreement, tense use, number of conjunctions, etc. Three different algorithms (see Granfeldt et al., 2006 for details) produce a profile based on six morphosyntactic stages of development, from beginner to native speaker, in turn identified and defined by Bartning and Schlyter (2004) and the results were validated by Granfeldt and Ågren (2014).

As part of the automatized morphosyntactic profiling, the texts were analysed in terms of a number of commonly used CAF measures. For complexity, tense use, Guiraud Index and sentence length were chosen. For tense use, the relative use of tenses other than the present tense was seen as a measure of verb phrase complexity (Granfeldt & Ågren, 2014). Tense use was computed on a 10-point scale, ranging from 0 (only present tense) to 10 (only other tenses), where a score of 6 implied that 6 out of 10 tenses were tenses other than the present tense. The Guiraud Index was chosen as a measure of lexical diversity for texts containing more than 200 words (van Hout & Vermeer, 2007). Sentence length is seen as a broad measure of sentence complexity (Wolf-Quintero et al., 1998).

Accuracy measures were subject-verb agreement (SVA) and determiner-noun Agreement (DNA), as they are expected to contribute significantly to accuracy in L2 French (Ågren et al., 2012). Both SVA and DNA were calculated on a scale ranging from 0 (no agreement) to 10 (100% agreement). Text length, operationalized as the total number of tokens in the text, was taken as a rough fluency measure. The online vocabulary program Vocabprofilers (Cobb, 2018) was used to calculate the Guiraud Index, average sentence length and text length.

Finally, to measure formulaic language use, all chunks (also referred to as formulaic sequences or multi-word sequences) were hand-coded. Even though chunks may be easy to define, they are difficult to operationalize as both relative frequency and association strength play a role, and they are difficult to distinguish from non-formulaic multi-word sequences. Following earlier work, we used the chunk categorization as employed by Verspoor et al. (2012), which in turn was based on Moon (1997, pp. 44–47), with two additional chunk types identified in other studies. The same categorisation was used by Piggott (2019) and Hou et al. (2018). Table 1 provides an overview of the categories with examples.

Partially schematic chunks include structures and complements, which require slots to be filled. Fully fixed chunks include compounds, particles, collocations, fixed phrases and discourse chunks, which do not have open slots.

To ensure the appropriateness of this chunk categorization for intermediate-level French, first a random sample of five SB and five DUB essays were selected and coded along the same lines. The operationalization of formulaic sequences as developed for English revealed to be quite suitable for French.

DOI: 10.22599/jesla.118

LABEL	DEFINITION	EXAMPLES*
Partially schematic chunks		
Structures	Fixed expression with slot-fillers	*j'ai vu une vidéo* (I have watched a video), *trois jours par semaine* (three days a week), *il faut être prudent,* (one has to be careful), *il y a des personnes qui disent que* (there are people who say that)
Complements	Verb with a complement (infinitives, gerunds, nominal sentences, or reflexives)	*tu veux* (you want), *on a vu que* (one has seen that), *on dit que* (one says that), *c'est une bonne idée de* (it is a good idea to), *je pense que* (I think that).
Fully fixed chunks		
Compounds and collocations	Fixed combinations of nouns, adjectives, prepositions, or particles; collocating nouns, adjectives, verbs, adverbs, prepositions and/or pronouns	*point de vue* (opinion), *tout le monde* (everyone), *espérance de vie* (life expectancy), *quelque chose* (something), *le marché noir* (black market), *aujourd'hui* (today), *quelqu'un* (someone), *plusieurs fois* (several times), *chaque jour* (every day), *gagnait l'argent* (earn money), *tomber enceinte* (to become pregnant), *commet un crime* (commit a crime), *a besoin d'argent* (to need money),
Particles	Verbs or nouns with prepositions or particles, including phrasal verbs	*Beaucoup de* (many), *dans le futur* (in the future), *risque de* (risk of), *loin du* (far from), *difficile à* (difficult to), *aspirent à* (aspire to), *à la télévision* (on television), *lutter contre* (to fight against)
Fixed Phrases	Conventional combinations of words, often idiomatic, usually consisting of more than two words	*il y a* (there is), *autant que possible* (as many as possible), *tout le temps* (all the time), *c'est une bonne idée* (it is a good idea), *je ne sais pas* (I don't know), *je suis contre* (I am opposed)
Discourse	Any form of chunk with a discourse function	*d'une part* (firstly), *par exemple* (for example), *parce que* (because), *tout d'abord* (first of all),

*The examples are taken from the present study, in order to provide French examples, where English examples were provided in the original source.

Table 1 Categorization of chunks, adapted from Verspoor et al. (2012, p. 250).

After the texts had been cleaned by removing personal information, each essay was coded for chunks by two independent raters. All differences in coding were discussed and resolved. Coding was done using a Visual Basic for Applications (VBA) script, a Microsoft Office application to create macros to automate repetitive word- and data-processing functions.

The problem with quantifying chunks is that longer chunks may have a negative effect on the total number of chunks. It is also difficult to count embedded chunks. Therefore, Gustafsson and Verspoor (2017) considered a variety of ways to measure chunk usage and found that simply counting the number of chunks did not reflect the actual amount of authentic target language use, but chunk coverage did. Chunk coverage is operationalized as the number of words occurring in chunks divided by the total number of words in text.

3.1.6. Tests and scoring: Speaking

To assess speaking skills, the so-called Student Oral Proficiency Assessment (SOPA) protocol was used (see Thompson et al., 2002, for details). It was originally designed for younger children (10–12) and was adapted for the older learners as follows. First, the interviewer asked personal questions to both students (Task 1). Then Student A and B alternately presented one of the subjects in one minute and were asked informative questions (Task 2 and 3). Tasks 1 through 3 were mainly intended as warm-ups to make the students feel at ease. As the questions could easily be prepared in advance at this level, most students did well on them. In Task 4, the students were invited to give their opinion on their own subject and on the subject of the other student. As this task involved open questions, required more extensive language use and could not be prepared in advance, it discriminated the most in terms of proficiency levels of the students. Therefore, the proficiency ratings were based mainly on the students' performance during this task.

DOI: 10.22599/jesla.118

To assess oral proficiency, the four subscales on the Center of Applied Linguistics Oral Proficiency Exam and Student Oral Proficiency Assessment Rating Scale (COPA/SOPA-RS) (oral fluency, grammar, vocabulary and listening comprehension) with nine levels each, were used to represent four aspects of speaking, and each level was assigned a score ranging from 1 to 9, which was considered a numeric variable (cf. Breiner-Sanders et al., 2000). As a result, each participant received four scores and a total score, which was the sum of all four individual scores. The total scores as well as the scores on the four aspects of speaking were analysed as five dependent variables.

To guarantee reliability of ratings across cohorts over the years, benchmarks were established in a pilot study on the basis of a cohort that is not part of the reported study and these benchmarks were established as follows. First, two experts rated all performances independently and achieved full agreement through discussion. Then, from this set, six students were selected with high, mid and low scores. The performance of each student was scrutinised for language features, and their speaking scores were compared with the students' other grades in reading, writing and overall class grades. As the ratings of these performances correlated highly with linguistic features and other scores, the final scores were thought to be valid and reliable. Thus, these performances were used as benchmarks to (re)train the raters before rating a new cohort of learners and new performances were calibrated with these benchmarks in mind.

3.1.7. Analyses

To compare the respective L2 measures, we used SPSS for Windows 27.01 (SPSS, Inc. Chicago, IL) to conduct a series of independent samples t-tests per language domain (listening, reading, speaking and writing) with group as the independent variable (two levels: SB and DUB) and the various language-domain measures outlined above as dependent variables. In cases of violated assumptions, we employed Mann-Whitney U tests (see Gombert, 2022, for specific details).

3.2. RESULTS

Table 2 shows the mean reading and listening scores per condition (descriptive statistics) as well as the comparison results across the two groups (inferential statistics). The independent samples t-tests showed no significant differences between the groups on reading skills (p = 0.143), with a small effect size (d = 0.26) and a significant difference between the groups on listening in favour of the DUB students (p < .001), with a large effect size (d = 0.86; see Plonsky & Oswald, 2014, for effect sizes).

Table 2 Reading and listening scores.

READING AND LISTENING	SB PROGRAM *N* = 55 MEAN *(SD)*	DUB PROGRAM *N* = 73 MEAN *(SD)*	COHEN'S *D*	SIGNIFICANCE (2-TAILED) *P*-VALUE
Reading scores (Max. 10)	6.36 (1.28)	6.67 (1.07)	.26	.143
Listening scores (Max. 10)	5.96 (1.32)	7.06 (1.24)	.86	<.001

Table 3 presents the holistic scores and the level of writing proficiency provided by *Direkt Profil*. The differences were not significant and effect sizes were low.

Table 3 Writing scores and morphosyntactic profiling.

WRITING	SB PROGRAM *N* = 24 MEAN *(SD)*	DUB PROGRAM *N* = 32 MEAN *(SD)*	COHEN'S *D*	SIGNIFICANCE (2-TAILED) *P*-VALUE
Holistic scores by expert teachers (scores 1–4)	2.50 (0.95)	2.58 (0.92)	.09	.757
Morphosyntactic profiling by Direkt Profil (scores 1–6)	4.05 (1.22)	4.25 (1.14)	.17	.534

Table 4 provides an overview of the results on CAF measures that we used in our study. As far as accuracy is concerned, there were no significant differences between the groups. However, the DUB students produced significantly longer sentences and produced longer texts overall, with medium effect sizes.

CAF	SB PROGRAM *N* = 24 MEAN *(SD)*	DUB PROGRAM *N* = 32 MEAN *(SD)*	COHEN'S D	SIGNIFICANCE (2-TAILED) *P*-VALUE
Complexity				
Guiraud Index	8.97 (1.12)	8.80 (1.00)	.16	.541
Tense Use Ratio	2.68 (1.93)	2.43 (1.05)	.16	.584
Average Sentence Length	14.72 (3.73)	17.33 (2.78)	.79	.001
Accuracy				
Subject-Verb Agreement	7.33 (1.63)	7.90 (1.28)	.39	.147
Determiner-Noun Agreement	9.15 (0.83)	8.83 (0.99)	.35	.213
Fluency				
Text Length	293 (75)	356 (128)	0.60	0.044

Table 4 CAF measures.

Table 5 presents the overall chunk coverage per instructional group. DUB students showed a higher chunk coverage (M = 44.09, SD = 7.40) than SB students (M = 38.12, SD = 8.83), with a large effect size.

CHUNKS	SB PROGRAM *N* = 24	DUB PROGRAM *N* = 32	COHEN'S *D*	*P*-VALUE
	Mean *(SD)*	Mean *(SD)*	.73	$p < .05$
	38.12 *(8.83)*	44.09 (7.40)		

Table 5 Chunk coverage.

Table 6 shows the overall scores on oral proficiency. This test showed a significantly higher mean for the DUB program than for the SB program with a very large effect size.

ORAL PROFICIENCY	SB PROGRAM *N* = 55	DUB PROGRAM *N* = 73	COHEN'S *D*	SIGNIFICANCE (2-TAILED) *P*-VALUE
	Mean (*SD*)	Mean (*SD*)		
	18.2 (4.1)	26.5 (4.0)	2.05	$p < .001$

Table 6 Scores on oral proficiency.

Table 7 presents the scores on the four sub-aspects of oral proficiency. With regard to these sub-scores, some assumptions were violated, and a non-parametric test was necessary. As the scores were not normally distributed, median scores had to be used for comparing the groups. All median scores were significantly higher for DUB students than for SB students with large effect sizes ranging from d = 1.31 to d = 2.21 on all four aspects of speaking.

SUB ASPECTS ORAL PROFICIENCY	SB PROGRAM *N* = 55 (SCORES 1–9)MEDIAN (MEAN/SD)	DUB PROGRAM *N* = 73 (SCORES 1–9)MEDIAN (MEAN/SD)	MANN WHITNEY *U*	STAND. SCORE *Z*	COHEN'S *D*	ASYMPTOTIC SIGNIFICANCE (2-TAILED) *P*-VALUE
Fluency	5.0 (4.8/1.2)	7.0 (6.9/0.9)	3673	8.2	2.01	.001
Grammar	5.0 (4.6/1.1)	6.0 (6.3/1.2)	3322	6.5	1.31	.001
Vocabulary	4.0 (4.5/1.1)	7.0 (6.5/1.2)	3533	7.5	1.71	.001
Listening	4.0 (4.2/1.2)	7.0 (6.9/1.1)	3749	8.5	2.21	.001

Table 7 Scores on four sub aspects of oral proficiency.

DOI: 10.22599/jesla.118

4. DISCUSSION

In the current study, several cohorts of SB and DUB students from the same school with the same teacher in years 4 through 6 were compared at the end of their studies on the final standardized national reading and listening exams and free-response writing and speaking tests. In order to enable the SB students to participate in the free-response tasks and to make sure the SB and DUB groups had equal access to the topics, students in both conditions received the same 30-hour DUB inspired intervention with a great deal of exposure and speaking practice. This intervention could have affected the outcomes of the SB group, especially in their performance on the free-response tasks. However, since the main goal was to see whether DUB students were able to do as well as SB students on their final reading exams and would be as accurate as them in writing, this potential benefit to SB students was not seen as a problem.

As far as the first research question and reading were concerned, we hypothesized that the SB students would outperform the DUB students, as a great deal of time in their final school years was spent on reading exam texts and training to answer the multiple-choice formatted questions. However, despite extensive training on exam tests, the SB did not do better. It is possible that the DUB learners did equally well because they had had relatively more L2 French exposure (both in written and spoken form), and even though they had not trained on reading comprehension tests, they had to comprehend the texts and videos because they had to discuss them extensively in class. This connects with a suggestion by Swain (2000, pp. 221–222), where she in turn refers to work by Netten and Spain (1989):

> Producing French may force learners to pay more attention to (or to notice) how the language is used to express one's intended meaning than does comprehending it .... Supporting evidence for this claim comes from an observational study conducted in three grade-two FI classes (Netten and Spain 1989). Of the three classes, one class (Class A) had a low average scholastic ability score (54th percentile) relative to the other two classes, yet performed unexpectedly well on a test of French reading comprehension and much better than Class C, whose average scholastic ability score was much higher (73rd percentile). Observations in these classes revealed that, in Class A, students "...were constantly using, and experimenting with, the second language as they engaged in communications of an academic and social nature with their peers and the teacher...," whereas in Class C, students "...had limited opportunities to use the second language to engage in real communication acts" (Netten & Spain, 1989, p. 494).

As far as the second research question pertaining listening is concerned, the DUB students outperformed the SB students, as we expected. Although the SB students had a high degree of L2 written exposure in the last three years to prepare for the reading exam and must have gained enough lexical knowledge for reading comprehension (reflected in a similar reading score across the two groups), it did not facilitate their listening comprehension development. Four aspects may have played a role in this result. First, in spite of the 30-hour intervention study, the DUB and SB programs differed substantially in the amount of L2 oral exposure provided. In the DUB program, maximal target language use in all six years and immersion-like activities with videos through online learning systems at home in the final three years provided far more input repetition for auditory form-use-meaning mappings to become entrenched compared to the minimal exposure their SB-taught peers received. Second, although segmentation skills (considered to contribute to the development of listening skills) were not explicitly practised in either the DUB and SB programs, the DUB program provided a great amount of exposure to oral French in the first three years, with a great deal of repetition. In the last three years, videos with French subtitles were provided and students were invited to read along. The repeated and simultaneous processing of oral and written forms is likely to have induced segmentation skills (cf. Vandegrift, 2007). Third, the need for phonological memory skills is expected to largely depend on the speed of meaning retrieval. As argued by different authors (cf. Gustafsson & Verspoor, 2017; Lewis, 1993; Pawley & Syder, 1983; Tang, 2013; Wray & Perkins, 2000), meaning retrieval will greatly benefit from a more entrenched auditory vocabulary and the mastery of chunks that can be the result of abundant L2 exposure in this DUB program. Fourth, we believe that students' listening anxiety caused by the ephemerous and timed character of auditory

DOI: 10.22599/jesla.118

input, especially during listening tests, is expected to be reduced considerably as students develop a certain listening ease caused in particular by target language practice. In the DUB program, students were expected to speak only French in class, which also helped maintain form-use-meaning mappings in everyday conversation, and after six years, listening to French L2 is business as usual and anxiety could be greatly reduced as a result of routine mechanisms.

As far as the third research question concerning the free-response writing test was concerned, we expected the SB students to outperform the DUB students in accuracy because they had received a great deal of instruction focusing on accuracy, especially in verb forms. This hypothesis was not confirmed. There were no differences in holistic scores, DP scores (which takes errors into consideration), nor in two highly frequent specific accuracy measures, including verb forms. Though not in line with our expectations, they are in line with earlier studies by Rousse-Malpat and Verspoor (2012) and Piggott (2019), who had already shown that implicitly taught students may take longer to achieve accuracy but only when given sufficient time to catch up with their SB-taught peers. In both studies, implicitly taught learners were less accurate than SB students after year 1, but after year 2 there were no differences. We see the same picture emerging in our results.

In contrast, the DUB students were predicted to be more fluent and use more formulaic language as they had been exposed to whole chunks and phrases and had produced a great deal of written and spoken texts. This hypothesis was confirmed: The DUB learners wrote significantly longer texts, wrote longer sentences and used significantly more formulaic language. This is in line with Piggot's findings (2019) that high-exposure students generally write longer texts and use longer sentences with more chunks. Our finding is also in line with results obtained in Gustafson and Verspoor (2017), who showed that high-exposure students use relatively more formulaic language.

For the final research question pertaining to free-response speaking, we expected the DUB students to outperform the SB students. The fact that this was borne out of the results for many aspects related to speaking did not come as a surprise, as the DUB learners had practised speaking extensively for six years, and the SB students had focused more on reading and writing skills. Apparently, the 30-hour DUB intervention that the SB students had received to practise the topics and the vocabulary involved was not enough to compensate for the general lack of fluency.

The results convincingly show that a DUB approach can prepare students adequately for their final exams in terms of reading and writing skills. In other words, the DUB program enabled learners to pass their exams, while also allowing them to develop communicative competence in the target foreign language French. This contradicts the claim that the development of writing skills, and accuracy in particular, necessitates explicit instruction on grammar, which was first motivated by SLA research (Norris & Ortega, 2000; Spada & Tomita, 2010) and was soon widely adopted by L2 teachers. Apparently, an implicit approach to the teaching of grammar, in our case within an AIM approach with a carefully built-up program with lots of repetition and corrective feedback, is as effective as an explicit approach to develop writing skills after six years. It has the added value that, without time spent on explicit grammar, there is ample time to promote spoken and written fluency skills.

5. CONCLUSION

Although the results of this study show that a DUB approach is more effective than a more traditional SB approach in developing listening and speaking skills and equally effective in the development reading and writing skills, several limitations need to be considered. As classrooms are dynamic environments involving numerous psychological and social processes, long-term classroom research usually involves many variables which cannot be controlled for. Henceforth, experimental control, which is a prerequisite for generalization of the findings, is impossible with this kind of research. Secondly, although this study is unique in the sense that it investigates the effects of a 6-year program, similar studies are needed to overcome the scientific problem reported above. Both the nature of this research and the lack of replication should limit the scope of this study and caution the reader to generalize any of these findings.

However, although generalization of this study's findings might not be possible due to the lack of experimental control, this study's ecological validity is high, as it investigates real programs in a real school with real students, using widely used tests that are part of the curriculum and not a simple intervention designed solely for the sake of the experiment. Ecological validity and experimental control, thus, seem to be incompatible by default in long-term classroom research. Although generalizability is indeed not possible, pedagogical implications can still be well-defined by practitioners and policymakers, who can interpret the findings against the backdrop of their own situations and determine what is relevant and meaningful to them (see also Spada, 2019). In the light of this "particularizability", coined by Clarke (1994), the relevance of this study cannot be denied for practitioners and policymakers who might be inspired to reconsider current practice in L2 teaching, as well as for researchers who might be inspired to replicate this study or to investigate L2 teaching in the same direction.

ETHICS AND CONSENT

This study was conducted at the faculty of Arts of the University of Groningen before it adopted a university-wide Research Data policy that necessitated a research data management plan and ethical approval. This means that no ethical approval had been obtained prior to data collection. However, the authors of this paper ensured that they conducted the study ethically. Written active informed consent from the parents of the pupils involved in the study was obtained before the start of the study.

ACKNOWLEDGEMENTS

The work reported in this paper was part of the first author's dissertation; the writing outcomes have been previously published in Gombert et al. (2022). We would like to thank two anonymous reviewers of an earlier version of this paper for their helpful suggestions. Our manuscript improved greatly because of their input, for which we are very thankful.

FUNDING INFORMATION

The dissertation work was made possible by Dudoc Alfa funding awarded to G. Gombert (grant number: DA2–2017-04), which is hereby gratefully acknowledged.

AUTHOR AFFILIATIONS

Wim Gombert orcid.org/0000-0002-7826-0692
University of Groningen, the Netherlands

Merel Keijzer orcid.org/0000-0001-9041-8563
University of Groningen, the Netherlands

Marjolijn Verspoor orcid.org/0000-0001-8828-8405
University of Pannonia, Hungary; University of Groningen, the Netherlands

REFERENCES

Ågren, M., Granfeldt, J., & **Schlyter, S.** (2012). The growth of complexity and accuracy in L2 French. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 95–119). John Benjamins. DOI: https://doi.org/10.1075/lllt.32.05agr

Bakker, T., Beijer, S., Brouwer, J., Heringa, J., & **Morice, I.** (2005). *Grandes lignes* [Broad outlines] (3rd ed.). Noordhoff.

Bartning, I., & **Schlyter, S.** (2004). Itinéraires acquisitionnels et stades de développement en français L2. [Developmental trajectories and stages in French as an L2.] *Journal of French Language Studies*, *14*(3), 281–299. DOI: https://doi.org/10.1017/S0959269504001802

Breek, L., Polinder, K., Groenen, G., & **Van der Drift, M.** (2003). *Libre Service* [Self Service] (2nd ed.). Thieme Meulenhoff.

Breiner-Sanders, K. E., Lowe, P., Miles, J., & **Swender, E.** (2000). ACTFL proficiency guidelines-Speaking (Revised 1999). *Foreign Language Annals, 33*(1), 13. DOI: https://doi.org/10.1111/j.1944-9720.2000.tb00885.x

Cito. (n.d.). Normeren [Standardizing]. Retrieved March 4, 2024 from https://cito.nl/centrale-toetsen-en-examens/centrale-examens-vo/voor-docenten/examenproces/wat-is-normeren/

Clarke, M. A. (1994). The dysfunctions of the theory/practice discourse. *TESOL Quarterly, 28*(1), 9–26. DOI: https://doi.org/10.2307/3587196

Cobb, T. (n.d.). *Compleat Web VP v.2*. Retrieved February 15, 2018 from https://www.lextutor.ca/vp/comp/

De Graaff, R., & **Housen, A.** (2009). Investigating the effects and effectiveness of L2 instruction. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 726–755). Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444315783

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*(2), 195–221. DOI: https://doi.org/10.1017/S0272263197002040

DeKeyser, R. M., & **Botano, G. P.** (2019). Current research on instructed second language learning: A bird's eye view. In R. M. DeKeyser & G. P. Botana, (Eds.), *Doing SLA research with implications for the classroom: Reconciling methodological demands and pedagogical applicability* (pp. 1–8). John Benjamins. DOI: https://doi.org/10.1075/lllt.52

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Blackwell. DOI: https://doi.org/10.1002/9780470756492.ch10

Ellis, N. C., & **Wulff, S.** (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 75–94). Routledge. DOI: https://doi.org/10.4324/9780203628942

Gombert, W. (2022). *From "learning to use" towards "using to learn"?: Long-term effects of structure-based versus dynamic usage-based programs for French.* [Unpublished Doctoral Dissertation]. University of Groningen.

Gombert, W., Keijzer, M., & **Verspoor, M.** (2022). Structure-based versus Dynamic usage-based instruction: L2 French writing skills after six years of instruction in high school. *Dutch Journal of Applied Linguistics, 11*. DOI: https://doi.org/10.51751/dujal10935

Granfeldt, J., & **Ågren, M.** (2014). SLA developmental stages and teachers' assessment of written French: Exploring Direkt Profil as a diagnostic assessment tool. *Language Testing, 31*(3), 285–305. DOI: https://doi.org/10.1177/0265532214526178

Granfeldt, J., Nugues, P., Persson, E., Persson, L., Kostadinov, F., Ågren, M., & **Schlyter, S.** Direkt Profil version 2.2 [computer program]. Retrieved February 15, 2018 from http://profil.sol.lu.se/profil/logon.jsp

Granfeldt, J., Nugues, P., Persson, E., Thulin, J., Ågren, M., & **Schlyter, S.** (2006). CEFLE and Direkt Profil: A new computer learner corpus in French L2 and a system for grammatical profiling. In P. Nugues, M. Ågren, J. Thulin, E. Persson, & S. Schlyter (Eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation* (pp. 565–570). ELRA.

Graus, J., & **Coppen, P. A.** (2016). Student teacher beliefs on grammar instruction. *Language Teaching Research*, *20*(5), 571–599. DOI: https://doi.org/10.1177/1362168815603237

Gustafsson, H., & **Verspoor, M.** (2017). Development of chunks in Dutch L2 learners of English. In J. Evers-Vermeul & E. Tribushinina (Eds.), *Usage-based approaches to language acquisition and language teaching* (pp. 235–264). de Gruyter. DOI: https://doi.org/10.1515/9781501505492-011

Hermans-Nymark, L. D. (2006). *English in the EFL classroom: Why not? Classroom discourse patterns and teachers' beliefs.* [Unpublished Doctoral dissertation]). RU Radboud Universiteit Nijmegen.

Hou, J., Loerts, H., & **Verspoor, M. H.** (2018). Chunk use and development in advanced Chinese L2 learners of English. *Language Teaching Research*, *22*(2), 148–168. DOI: https://doi.org/10.1177/1362168816662290

Kang, E. Y., Sok, S., & **Han, Z.** (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research, 23*(4), 428–453. DOI: https://doi.org/10.1177/1362168818776671

Krashen, S. (1992). The input hypothesis: An update. In J. Alatis (Ed.), *Georgetown University Round Table in Languages and Linguistics* (pp. 409–431). Georgetown University Press.

Larsen-Freeman, D. (2007). On the complementarity of chaos/complexity theory and dynamic systems theory in understanding the second language acquisition process. *Bilingualism: Language and cognition*, *10*(1), 35–37. DOI: https://doi.org/10.1017/S136672890600277X

Lewis, M. (1993). *The lexical approach* (Vol. 1). Language teaching publications.

Lightbown, P. M., & **Spada, N.** (2013). *How languages are learned* (4th ed.). Oxford University Press.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (pp. 179–192). John Benjamins. DOI: https://doi.org/10.1075/z.96.11lon

Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40–63). Cambridge University Press.

Norris, J. M., & **Ortega, L.** (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning, 50*(3), 417–528. DOI: https://doi.org/10.1111/0023-8333.00136

Pawley, A., & **Syder, F. H.** (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–225). Longman.

Piggott, L., Tribushinina, E., & **de Graaff, R.** (2020). The Icing on the Cake? Effects of explicit form-focused instruction after two tears of implicit EFL learning. In W. Lowie, M. Michel, A. Rousse-Malpat, M. Keijzer & R. Steinkrauss (Eds.), *Usage-based dynamics in second language development* (pp. 249–270). Blue Ridge Summit. DOI: https://doi.org/10.21832/9781788925259-014

Piggott, L. E. B. (2019). *First meaning then form: a longitudinal study on the effects of delaying and reducing explicit form-focused instruction for young adolescent EFL learners* [Unpublished Doctoral Dissertation]. University of Utrecht.

Plonsky, L., & **Oswald, F. L.** (2014). How big is "big"? Interpreting effect sizes in L2 research: effect sizes in L2 research. *Language Learning, 64*(4), 878–912. DOI: https://doi.org/10.1111/lang.12079

Popma, G. (1997). Mag het ietsje minder zijn? Het gebruik van Nederlands in de Engelse les. (May it be a little less? The use of Dutch in English lessons.) *Levende Talen Magazine, 84*(517), 70–73.

Rousse-Malpat, A., Steinkrauss, R., Wieling, M., & **Verspoor, M.** (2022). Communicative language teaching: Structure-based or Dynamic usage-based? *Journal of the European Second Language Association*, *6*(1), 1–14. DOI: https://doi.org/10.22599/jesla.86

Rousse-Malpat, A., & **Verspoor, M. H.** (2012). Measuring effectiveness in focus on form versus focus on meaning. *Dutch Journal of Applied Linguistics, 1*(2), 263–276. DOI: https://doi.org/10.1075/dujal.1.2.07rou

Netten, J. E., & **Spain, W. H.** (1989). Student-teacher interaction patterns in the French immersion classroom: Implications for levels of achievement in French language proficiency. *Canadian Modern Language Review, 45*, 485–501.

Schmid H. J. (2020). *The Dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198814771.001.0001

Spada, N. (2019). Balancing methodological rigor and pedagogical relevance. In R. M. DeKeyser, & G. P. Botana (Eds.), *Doing SLA research with implications for the classroom: Reconciling methodological demands and pedagogical applicability* (pp. 201–215). John Benjamins. DOI: https://doi.org/10.1075/lllt.52.10spa

Spada, N., & **Tomita, Y.** (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263–308. DOI: https://doi.org/10.1111/j.1467-9922.2010.00562.x

Swain, M. (2000). French immersion research in Canada: Recent contributions to SLA and applied linguistics. *Annual Review of Applied Linguistics*, *20*, 199–212. DOI: https://doi.org/10.1017/S0267190500200123

Tang, J. (2013). Input of chunks and its effects on L2 learners' listening competency. *Theory and Practice in Language Studies, 3*(7), 1264–1269. DOI: https://doi.org/10.4304/tpls.3.7.1264-1269

Thompson, L., Kenyon, D., & **Rhodes, N.** (2002). A validation study of the student oral proficiency assessment (SOPA). Iowa State University.

Van Compernolle, R. A. (2015). *Interaction and second language development*. John Benjamins. DOI: https://doi.org/10.1075/lllt.44

van Hout, R., & **Vermeer, A.** (2007). Comparing measures of lexical richness. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 93–115). Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511667268.008

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191–210. DOI: https://doi.org/10.1017/S0261444807004338

Verspoor, M. (2017). Complex dynamic systems theory and L2 pedagogy: Lessons to be learned. In L. Ortega & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman*. John Benjamins. DOI: https://doi.org/10.1075/lllt.48.08ver

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. DOI: https://doi.org/10.1016/j.jslw.2012.03.007

Voogel, M. (2018). Bon ton of boring? De ontwikkeling van het Frans in onderwijs en uitgeverij in Nederland Bon ton of boring [Bon to or boring? The development of French in education and publishing in the Netherlands] [Doctoral Dissertation]. Pallas Publications.

West, L., & **Verspoor, M.** (2016). An impression of foreign language teaching approaches in the Netherlands. *Levende Talen Tijdschrift, 17*(4), 26–36.

Wolf-Quintero, K., Inagaki, S., & **Kim, H.-Y.** (1998). Second language development in writing: Measures of fluency, accuracy, & complexity. Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.

Wray, A., & **Perkins, M. R.** (2000). The functions of formulaic language: An integrated model. *Language & Communication, 20*(1), 1–28. DOI: https://doi.org/10.1016/S0271-5309(99)00015-4

DOI: 10.22599/jesla.118

TO CITE THIS ARTICLE:
Gombert, W., Keijzer, M., & Verspoor, M. (2024). Long-term effects of structure-based versus dynamic usage-based instructional programs for French. *Journal of the European Second Language Association*, 8(1), 18–33. DOI: https://doi.org/10.22599/jesla.118

Submitted: 09 October 2023
Accepted: 02 April 2024
Published: 21 May 2024

Journal of the European Second Language Association, is a peer-reviewed open access journal published by White Rose University Press.

WHITE ROSE UNIVERSITY PRESS
Universities of Leeds, Sheffield & York