



Communicative language teaching: Structure-Based or Dynamic Usage-Based?

AUDREY ROUSSE-MALPAT 

RASMUS STEINKRAUSS 

MARTIJN WIELING 

MARJOLIJN VERSPOOR 

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

RESEARCH

WHITE ROSE
UNIVERSITY PRESS
Universities of Leeds, Sheffield & York

ABSTRACT

Although communicative language teaching (CLT) was thought to have revolutionized classroom practice, there are “weak” and “strong” versions (Howatt, 1984). Most foreign language classrooms in the world still favor weak versions with structure-based (SB) views on language (Lightbown & Spada, 2013), and practice in the Netherlands is not much different (West & Verspoor, 2016). However, a small group of teachers in the Netherlands started teaching French as a second language with a strong CLT program in line with Dynamic Usage-Based (DUB) principles. Rather than focusing on rule learning and explicit grammar teaching to avoid errors, the DUB program takes the dynamics of second-language development into consideration and focuses on the three key elements of usage-based theory: frequency, salience and contingency. These translate into a great deal of exposure, repetition, learning the meaning of every single word through gestures, and presenting whole chunks of language, all without explicit grammar teaching. This study aims to compare the effects of the SB and DUB instructional programs after three years. We traced the second-language development of 229 junior high school students (aged 12 to 15) learning French in the Netherlands over three years. The participants took three oral tests over the course of three years (568 interviews) and wrote seven narratives on the same topic (1511 narratives). As expected, the DUB approach, which is in line with a strong CLT version, was more effective in achieving proficiency in both speaking and writing and equally effective in achieving accuracy.

CORRESPONDING AUTHOR:

Audrey Rousse-Malpat

University of Groningen, NL

a.rousse-malpat@rug.nl

KEYWORDS:

instruction type;
Communicative Language Teaching; Dynamic Usage-Based; structure-based;
L2 French; L2 teaching effectiveness

TO CITE THIS ARTICLE:

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), 20–33. DOI: <https://doi.org/10.22599/jesla.86>

1. INTRODUCTION

Communicative Language Teaching (CLT) has been in vogue since Hymes (1971) pointed out that communicative competence, rather than grammatical competence, was needed to become a proficient second-language (L2) speaker. Two main versions of CLT exist, which Howatt (1984) explained as follows:

The “weak” version stresses the importance of providing learners with opportunities to use their English for communicative purposes and, characteristically, attempts to integrate such activities into a wider program of language teaching. Efforts are made to ensure the communicative activities relate to the purpose of the course as specified in the syllabus, hence the importance of proposals to include semantic as well as purely structural features in a syllabus design. The “strong” version of communicative teaching advances the claim that language is acquired through communication, so that it is not merely a question of activating an existing but inert knowledge of the language, but of stimulating the development of the language system itself. If the former could be described as “learning to use” English, the latter entails “using English to learn it” (p. 279).

Lightbown and Spada (2013) reviewed empirical studies and concluded that an approach based on communicative principles has the best chance to be effective in language teaching. This means language should be used meaningfully, with a large amount of input – preferably as authentic or meaningful as possible – and with some attention to form, very much in line with a Focus on Meaning or Focus on Form approach (Long, 2000). Examples that they give of effective approaches include “strong” versions of CLT such as the reading comprehension approach, in which learners read and listen to books, and a content and language integrated learning approach, in which L2 learners are taught subject content (such as science) in the target language. The conclusion that strong versions of CLT are much more effective than structure-based (SB) programs is very much in line with the general consensus about what effective language teaching should look like (Long, 2000). Still, Lightbown and Spada (2013) argue that the use of SB teaching methods, which may be considered “weak” versions of CLT, remain widespread.

Such “weak” versions of CLT are also common in the Netherlands (West & Verspoor, 2016) with a heavy emphasis on explicit language teaching, often in the first language (L1), which is reminiscent of the Focus on Forms approach (Long, 2000). However, about 10 years ago, a group of teachers who were disappointed with the little French their L2 learners were able to speak after 6 years of instruction imported a “strong” CLT method from Canada – the Accelerated Integrated Method (AIM) (Maxwell, 2001). This instructional method is based on story telling scripts with a great deal of repetition of whole phrases and gestures to make meanings of words clear and has no explicit grammar lessons, which is in line with Dynamic Usage-Based (DUB) principles. In the next section, we discuss the theoretical differences between the two approaches.

2. SB VERSUS DUB APPROACHES

According to Verspoor (2017), the difference between an SB and a DUB approach is the way language itself is conceptualized. An SB-inspired approach assumes language is a complex system in which different sub-components (such as syntax and lexicon) interact predictably according to rules. In contrast, a DUB-inspired approach assumes language is a dynamic system in which all components are interconnected and can change unpredictably; there is no fundamental difference between syntax and lexicon. Moreover, language learning dynamically interacts with cognitive, affective, and contextual factors. These different views have implications for how language should be presented and instructed and the behavior that the learners should aim for.

SB instruction refers to approaches that view language learning as rule-driven (Lightbown & Spada, 2013) and often have elements of a Grammar-Translation approach (Richards & Rodgers, 2014). Generally, it is assumed that learning and applying grammatical rules is needed in becoming proficient in an L2. In order to avoid fossilization, the focus is on accuracy of the grammatical forms presented from simple to complex. Other aspects of language – vocabulary, formulaic phrases, pronunciation and intonation, pragmatic use and so on – are usually taught as discrete units. Commonly used CLT textbook approaches in the Netherlands

are in line with an SB approach in that they include explicit grammar sections (Tammenga-Helmantel & Majjala, 2019).

DUB instruction refers to approaches that view language learning as habit-driven. The term DUB combines Complex Dynamic Systems Theory and usage-based perspectives. DUB accentuates the view that language development is experiential, individually owned and non-linear and that some sub-systems or skills are learned before others, such as lexicon before syntax (Verspoor et al., 2012). Most importantly, making errors is part and parcel of the developmental process and variability in the use of structures (which includes making errors) is needed to progress (Verspoor, 2017). Complex Dynamic Systems Theory is in line with usage-based approaches (Langacker, 2000), where frequency, salience and contingency are key concepts. Frequency effects have been attested in many L1 and L2 studies and are a consequence of repeated exposure to meaningful input and language use (cf. Ellis, 2002; Tomasello, 2003).

In usage-based approaches, it is assumed that linguistic constructions (pairings of form and meaning within a pragmatic context) are learned through association as they are “heard and used frequently and therefore entrenched, which is the result of habit formation, routinization or automatization” (Verspoor & Schmitt, 2013, p. 354). We suggest that salience can be achieved by trying to create joint attention, for example by emphasizing and carefully articulating a part of a construction, pointing out a grammatical form, or adding a gesture or other visual clue. Contingency can be achieved by offering language in phrases or longer multi-word sequences so learners hear which words belong together and by offering similar linguistic constructions in similar communicative contexts so that form-meaning associations can emerge. Gestures and visuals can also be considered contingent symbols.

Through exposure and use, learners can detect regular patterns through general learning mechanisms such as perception, association, categorization and schematization (Schmid, 2020). Because a DUB approach focuses on meaningful exposure and language use, the principles align with strong versions of a CLT approach in that it relies on communication to stimulate the development of the language system itself (Howatt, 1984). The AIM method used in the current study aligns with these principles in that the target language is used exclusively with a lot of repetition of phrases throughout the year and implicit learning of grammar (Rousse-Malpat et al., 2021).

Despite the fact that strong versions of CLT are recommended (Lightbown & Spada, 2013), we may wonder why SB methods, with their strong explicit component on morphosyntax, are still so very common in foreign-language teaching. Graus and Coppen (2017) showed that at teacher training colleges and in real practice, there is still a strong belief that explicit grammar instruction is a prerequisite for successful L2 learning. There is also evidence pointing to a positive effect of explicit grammar instruction (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). However, Doughty (2003) pointed out that the effects of explicit instruction may have been overestimated because research designs often favor explicit types of instruction and use proficiency measures relying on “constrained, constructed responses” (e.g., fill the blanks, metalinguistic judgement responses) (Spada, 2011, p. 228) and studying brief treatments only (Spada & Tomita, 2010). The problem in comparing explicit and implicit instruction is that implicitly taught learners have to discover the language patterns on their own, and this process may require relatively more hours of exposure. For example, Rousse-Malpat and Verspoor (2012) showed that after one year, learners taught with an SB approach with explicit instruction were more accurate than DUB learners, but this difference disappeared after two years. Therefore, studies looking at brief periods of instruction and predominantly at grammatical accuracy might be biased in favor of explicit instructional settings.

To address these research concerns, DeKeyser (2003) suggests conducting more realistic experiments “in actual classrooms, with much larger fragments of language” where students are learning to achieve communicative skills rather than “just learning for the sake of the experiment” (p. 337). The tests should thus focus on general language skills (speaking, listening, reading and/or writing), with less focus on grammar and more on general linguistic abilities (DeKeyser, 2017).

In all, there are very few long-term, systematic and empirical studies that compare the effects of “strong” or “weak” CLT teaching approaches in terms of general, communicative linguistic abilities. We are aware of a few and briefly review them here. First, Bourdages and Vignola (2009) conducted semi-structured interviews with AIM and non-AIM learners in their first year

of instruction (the same DUB teaching approach as in the current study). The authors found “few significant differences between the AIM group and the non-AIM group” (Bourdages & Vignola, 2009, p. 731). But as Cummins (2014) pointed out, the researchers focused only on the fact that both groups of early-stage learners were making similar grammatical errors; they did not point out that AIM students were more fluent and continued to speak French rather than English when attempting to express themselves.

Second, Verspoor and Hong (2013) investigated whether the teaching of English as a foreign language could be improved at a Vietnamese university as the task-based inspired method with SB principles was not felt to be effective. The semester course included a high degree of meaningful input provided by means of a popular English movie. There was no explicit explanation of grammar rules, but a great deal of implicit focused on form-use-meaning pairs at all levels. The control group was taught with a task-based inspired course developed by the university English teachers, which focused on form, interaction and output, with relatively little authentic English input. The results demonstrated that although both groups improved in English proficiency, the experimental group had significantly higher gain scores on the receptive General English Proficiency test and on the productive writing test. Subsequently, Irshad et al. (2019) replicated Verspoor and Hong (2013) in Sri Lanka, using the same movie-based method (different film) with the same battery of tests. However, they added a Computer Assisted Language Learning condition and found this condition to be most effective, perhaps because the learners could pace themselves.

In another study, Rousse-Malpat and Verspoor (2012) compared the effectiveness of their SB method to the new DUB-inspired AIM method. Four groups of students were traced in their first two years involving two teachers, each of whom taught one SB course and one AIM course. Results showed that when effectiveness was understood as general proficiency, the DUB group significantly outperformed the SB group. However, the SB group was more accurate after one year. This difference in accuracy disappeared after two years.

Finally, Gombert, Keijzer et al. (submitted) studied learners' writing after six years in the SB and DUB programs. The students' texts were scored both holistically with human raters giving an overall proficiency score and analytically with various hand-counted complexity, accuracy and fluency measures. The holistic scores showed no differences, but in some hand-counted measures, the DUB students outperformed the SB learners in sentence length and text length, but there were no differences in accuracy. In a related study, Gombert, Vandendorpe et al. (submitted) found that the DUB students used significantly more chunks (multi-word or formulaic sequences) than their SB counterparts.

Although the field of second language acquisition has long commended “strong” versions of CLT in foreign-language teaching, which would be in line with a DUB view of language, “weak” versions dominate because teachers believe that an SB approach with explicit grammar teaching is needed to avoid fossilization. Therefore, this study seeks to provide empirical evidence in this continuing debate. The research question was as follows: Which type of instructional program – SB or DUB – is more effective after three years in terms of general oral and written proficiency in L2 French? To answer this question, we compared two CLT instructional programs, either based on SB or DUB principles over the course of three academic years.

3. METHOD

3.1. CONTEXT AND TEACHERS

The study was conducted in the Netherlands at five different schools. Secondary education in the Netherlands is streamed according to academic level, which is tested at the end of primary school by the Cito test (Steerenberg, 1996), with the highest level preparing for university and lower levels for professional institutes.

There is no extramural French in the Netherlands and children are not exposed to French on a regular basis like they are to English. There is no set curriculum for the teaching of foreign languages, only targeted final learning outcomes of a Common European Framework of Reference for Languages (CEFR) level in the final year. School teams may decide on their own methods and testing procedures; some will use tests provided by the method, others will develop their own tests. The majority of foreign-language classes are taught with textbook methods as the ones used in the current study.

The AIM method used in this study is based on stories and was originally created for elementary school children, but as Dutch learners of L2 French are true beginners when they enter high school, the method has shown to be appropriate for them too.

In total, 14 teachers were involved in the study. In an attempt to control for teacher effects as much as possible, we selected teachers by means of interviews and questionnaires given to both teachers and learners. All teachers had a high level of French proficiency (B2/C1 according to the CEFR; Council of Europe, 2001), were content with the method they were teaching with, had more than five years of teaching experience and had a good rating by their students on teaching qualities. Some groups kept the same teacher throughout the three years; others had a new teacher every school year.

3.2. SCHOLASTIC ABILITY

At the age of 12, Dutch students take a national Cito test, a scholastic ability test consisting of reading and mathematics questions (which are also quite language oriented) with scores up to 550. Based on teachers' recommendations and these scores, students are streamed into about ability levels of high school. We examined the three highest levels of high school. Bilingual schools (Gymnasia) that have Greek and Latin are the top level. Voortgezet Wetenschappelijk Onderwijs is the second highest level and is also university preparatory. The third highest is Hoger Algemeen Vormend Onderwijs, which prepares students for professional courses, including teacher training institutes. These academic levels were controlled for in the current study, as in Verspoor et al. (2015) who found Cito scores to be good predictors for foreign-language achievement.

3.3. INSTRUCTIONAL GROUPS

Participants were placed in classes by their respective schools, but they did not choose in advance the method with which they would be learning French. We distinguished two groups of participants: Those learning French with an SB textbook called *Grandes Lignes* or *D'accord* and those learning French with a DUB method (AIM). All participants kept learning with the same teaching method throughout the three years of the study.

3.4. EXPOSURE

The SB method was expected to provide little L2 exposure and the DUB method was expected to provide a great deal of L2 exposure. To verify these assumptions, we estimated the number of teaching hours each participant had during three years and the percentage of French the learners heard, read, spoke or wrote during these hours. The percentage was calculated according to the answers the teachers gave in the background questionnaire, actual classroom observations (two per teacher) and verification by the teachers. Table 1 shows that students in SB programs were exposed to French between 40 and 60% of the time, whereas students in DUB programs were exposed to French 90% of the time.

SCHOOL	PROGRAM	% OF L2 EXPOSURE	YEAR 1	YEAR 2	YEAR 3
School 1	DUB	90	63	63	63
School 2	SB	60	48	48	48
School 3	DUB	95	90	60	45
School 4	SB	40	32	16	32
School 5	DUB	90	90	90	90
	SB	50	50	50	50

Table 1 Percentage of L2 exposure (in hours L2 instruction per year).

3.5. PARTICIPANTS

Originally the study had 309 learners, but those who did not have the same instructional method for three years and those who had had extramural French exposure or a type of learning disorder were excluded. The study included 229 Dutch high school L2 learners of French aged 13 at the beginning of the study enrolled in five different schools. Ninety-two were in SB classes

and 137 in DUB classes. Most of them were Dutch monolinguals. Some were bilingual ($n = 5$) but considered themselves to have a native level of Dutch. At the beginning of the study, no learners had previous knowledge of French, so they were true beginners.

3.6. TEACHING METHODS

In the SB condition, the teachers used communicative textbooks commonly used in the Netherlands: *Grandes Lignes* (Brouwer et al., 2014) and *D'Accord* (Ebbinge-Planchon et al., 2013). The chapters are organized around topics such as family, school, sports and holidays. Learners are first presented with a reading text containing target grammatical rules and vocabulary. Then, students are asked to answer questions, usually in the L1, about the meaning of the text. The book offers explicit explanation of the grammatical rules in the L1. Grammatical rules are displayed from simple, such as articles and the present tense, to more complex, such as possessive pronouns and the past perfect tense.

In principle, the SB method could allow for a great deal of L2 exposure (as one of the SB teachers was able to provide), but our classroom observations showed that much time was spent in the L1 on the grammatical rules explicitly addressed by the teacher and then practiced in small groups of learners. Vocabulary was given in the form of a word list or a chunk list called *phrases clés* (key sentences) with their translation into Dutch. The textbook exercises would allow for spontaneous interaction in the L2, but classroom observation showed that the interaction between learners was usually prepared in advance, often in the L1.

SB lessons were designed around the activities in the book. They usually started with the correction of the homework as a whole class activity or in small groups, followed by a new item that needed to be mastered (listening exercise, reading or grammar). The teacher gave corrective feedback to learners individually or to the whole class. Sometimes, a recast was used, but most of the time, the teacher relied on the L1 for explanation. Teachers also had designed their own activities, and we observed some learners practicing a role play. Learners were tested on receptive skills (listening and reading) and had to apply grammar rules to fill in the gap to test their comprehension of the grammar. Some teachers used tests provided by the book publishers, others designed their own tests. Tests usually contained only closed questions with a strong focus on morphological accuracy.

In the DUB condition, the teacher used the AIM (Maxwell, 2001). The method consists of booklets with various stories, fairy tales early on and later short narratives about travelling, school, friends and family. Classroom management talk and the story scripts make use of L2 pared-down language, which helps make contingency issues consistent (for example, different verb forms are avoided early on by using the indefinite pronoun *on* ('one' or 'we') instead of *nous*, *vous* and *ils* ('we', 'you-plural' or 'you-formal' and 'they', respectively). Using clear visuals and a gesture for each word, teachers read out the scripts in meaningful chunks with the learners sitting in a circle around the teacher. Learners repeat the chunks and gestures, usually in chorus. As the first few lines of the first story illustrate in (1), the story provides a great deal of built-in repetition and chunks of language with similar patterns (*voici* five times, *cochon(s)* four times, *il + verb* seven times).

(1) *Voici l'histoire des trois petits cochons.* ('This is the story of the three little pigs.')

Voici le premier petit cochon. ('This is the first little pig.')

Il joue de la guitare et il est gentil. ('He plays the guitar and is very sweet.')

Voici le deuxième petit cochon. ('This is the second little pig.')

Il travaille un peu et il aime la musique. ('He works a little and he likes music.')

Voici le troisième petit cochon. ('This is the third little pig.')

Il danse et chante et il est fantastique. ('He dances and sings and he is fantastic.')

Voici le loup. Il est méchant. ('This is the wolf. He is mean.').

It takes about six months to go through one story with fast-paced, 10-minute activities in each lesson (e.g., drilling of chunks previously learned, group activities with role-play, singing songs, introduction of new parts of the story). The goal is to entrench whole chunks that learners will later be able to reuse in other contexts. There is focus on form: When learners mispronounce a word or make another error, the teacher will draw attention to the form and then provide the whole chunk again.

Both teachers and learners use only the target language in the classroom from the first day, both in speaking and writing. Learners are not tested until they start writing, six months after the beginning of the lessons. They are tested on their general receptive skills (listening and reading) and productive skills (speaking and writing) in the form of comprehension tests or free-production tasks.

3.7. PROFICIENCY ASSESSMENT IN FRENCH

3.7.1. Oral proficiency test

To be able to remain consistent in our scoring of oral proficiency over the years, we used a standardized, validated oral proficiency test (Rhodes, 1996; Thompson et al., 2002). Developed in 1991 by the Center for Applied Linguistics for Spanish students of English, the student oral proficiency assessment (SOPA) method has been increasingly used to test students with other language backgrounds. One advantage is that the SOPA can be used independent of the mode of language instruction as it is not based on a specific curriculum but on everyday functions of language. It focuses on what participants can do with the language when they need to express ideas in real or simulated situations. The setting and the tasks are designed to decrease the anxiety level of the L2 learners as much as possible and give them room to show their highest proficiency level.

The SOPA interviews were held at the end of each school year (after 10, 22 and 34 months of instruction) in a setting with two students sitting opposite the two researchers. Interviewing students in pairs prevented them from becoming too nervous as they were allowed to help each other out. The aim was to be able to interview and rate two students at the same time. One researcher was the interviewer and the other rated the students' performance (see Rouse-Malpat & Verspoor, 2012, for an example of rating sheets).

The pairs of students were decided upon by the teachers who were asked to match students according to proficiency and personality. Each interview was videotaped and took about 20 minutes, followed by a five-minute session to assess the learners.

Learners were asked to perform three tasks increasing in difficulty. At all times, students were made to feel comfortable with compliments and encouragement and when the learner's ceiling level had been reached, the interviewer went back to an easier task so that the learners left feeling they had done well. At the end, learners received a candy reward.

3.7.2. Writing proficiency task

To test writing proficiency, we used free writings in the form of short, simple narratives based on topics discussed in class. We limited the genre to narratives to keep the genre constant over time. The narratives were written in the classroom in about 20 to 30 minutes. Learners were instructed to write a text according to the criteria in (2).

- (2) the story is interesting to the reader;
- the author should use as many words as they could;
- the reader can easily follow the story.

We asked teachers to stress the fact that the writing assignments would not be graded on grammar but on content. The learners had to write as well as they could to help the reader to understand the story. Learners were not allowed to consult a dictionary or get any other help. To keep the learners motivated to write throughout the study, teachers were asked to grade the assignments or to award bonus points. The writing started five months after the first French lesson and continued during three school years. Table 2 provides details on the writing proficiency task.

3.8. SCORING

3.8.1. Oral proficiency

The scoring of the SOPA tests was done by well-trained and certified SOPA assessors. Immediately after the testing session, the interviewer and the rater discussed and mutually agreed on the scores for oral fluency, vocabulary and listening comprehension, using a standardized rubric (for the rubric, see Rouse-Malpat & Verspoor, 2012) based on the developmental stages of

YEAR	# OF MONTHS OF INSTRUCTION	TOPIC	# OF PARTICIPANTS	
			SB	DUB
Year 1	4	Talk about yourself!	138	174
	9	Padma (DUB) / Jane (SB) likes to travel! Tell her story!	108	173
Year 2	15	Tell about a day at the beach!	23	47
	18	It is prom night! Tell the story!	68	118
Year 3	26	It is the first day (DUB) / week (SB) of school! Tell about how it went!	89	134
	28	What is your favorite book, film or series? Tell the story!	105	133
	34	Let's talk about the future! How do you see your life in 20 years?	78	123

Table 2 Written Assignments per Condition.

language learners ranging from fixed formulae, unsuccessful creative language to successful conventionalized ways of saying things. The testing sessions were videotaped so scores could be double checked and transcribed in case they were needed for further analysis.

3.8.2. Writing proficiency

To rate the writing samples, an approach similar to the SOPA evaluation was taken. As no rubrics existed for writing at the beginning level, we created our own. We collected written samples on different topics for the study reported in Rousse-Malpat et al. (2012), and we asked five experts in L2 French to rate 39 assignments according to proficiency level. They were instructed to look at the texts holistically, especially for overall meaning-making and coherence. They could consider vocabulary and grammar, but not focus only on those aspects. The experts compared their ratings and discussed differences until consensus was reached. Once texts were assigned a level, the experts were asked to explain and agree on what the common characteristics of the samples in each level were, which led to a rubric.

This rubric was validated by asking two other raters to score 475 assignments from the three first writing tasks that were collected in the current project. An interrater reliability analysis using the Spearman Rho correlation was performed to determine consistency among the two raters, not only for all 475 evaluated texts, but also separately for each writing task in order to see whether there was variation across three different writing tasks. The mean interrater reliability for the raters was found to be $\rho = 0.842$, which shows a strong agreement across raters. The lowest value of interrater agreement was found for the third writing task ($\rho = 0.74$) which was still considered acceptable.

3.9. ANALYSIS

The oral and written data were analyzed using mixed-effects regression modeling with R (version 3.4.1 with lme4 package; Bates et al., 2015). The holistic oral and written scores of each student were the dependent variable. We gave this variable the name Grade in the model. There was one oral score per year (three in total). The written scores (two in year 1, two in year 2 and three in year 3) were averaged in the model for each year (three in total).

The significance of several predictors of interest was assessed: Program (DUB versus SB), task type (oral versus written) and time of testing (Year 1, Year 2, Year 3). Time of testing was centered in such a way that year 2 was set to 0, year 1 was set to -1 and year 3 was set to the value 1. We included random intercepts for participant, level, class and teacher and random slopes for the predictors of interest was assessed. In other words, we added predictors of interest in a model one by one and every time we compared the new model with the old using ANOVA. We included a random intercept or slope whenever model comparison (using the ANOVA function) indicated its inclusion was significant (with $p < .05$). Including random slopes and intercepts is important in order to avoid type-I errors in assessing the influence of the predictors of interest (Baayen, 2008).

4. RESULTS

4.1. DESCRIPTIVES

Table 3 and Figure 1 and show how each group developed over time in the two separate skills, speaking and writing. Both skills developed more in the DUB group than in the SB group.

TASK AND INSTRUCTIONAL CONDITION		YEAR 1				YEAR 2				YEAR 3			
		M	SD	MIN	MAX	M	SD	MIN	MAX	M	SD	MIN	MAX
Oral	DUB	2.00	0.79	1	4	2.76	1.01	1	5	3.54	0.90	2	6
	SB	1.32	0.51	1	3	1.90	0.74	1	4	2.40	0.72	1	5
Written	DUB	1.55	0.54	0	4,5	3.26	0.68	1	5	3.21	0.56	2	6
	SB	1.26	0.49	0	4	2.62	0.60	1	4	2.56	0.64	0.5	4

Table 3 Descriptive statistics.

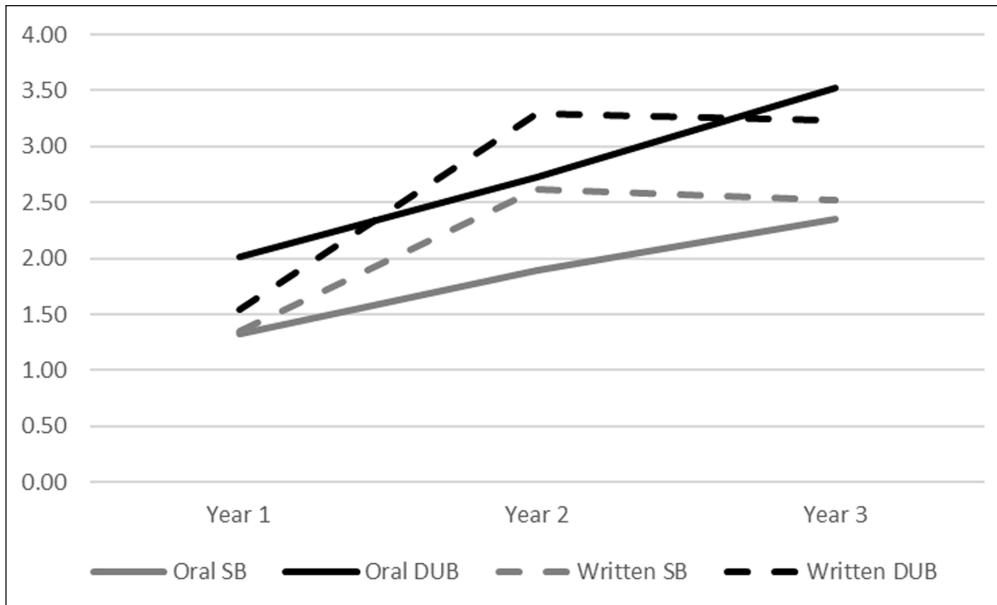


Figure 1 Development of each group over time.

4.2. MIXED-EFFECTS MODEL

Using the procedure of multiple pairwise comparisons, we determined that the best random intercepts to use for our models were participant, class and level. Teacher did not appear to be a significant random intercept. We also needed to include by-participant and by-class random slopes for year (i.e., the moment of testing), which showed that there was much individual and class-related variability among the participants throughout the years.

Regarding the fixed effects, only the factors year and program were significant; task type was not significant. The predictors and random effects were included using model comparison, and the following model specification was determined to be optimal: $grade \sim program * year + (1 + year | participant) + (1 + year | class) + (1 | level)$.

As [Table 4](#) and [Figure 2](#) show, there is a clear interaction between program and year. They show that all students improved over the years, but that the improvement was greater for the DUB group. An R^2 of 62.0 percent was found for the model, meaning that 62 percent of the variance was explained by the random and the fixed effects. An R^2 of 41.9 percent was found for the fixed effects alone, meaning that 41.9 percent of the variance was explained by the fixed effects (predictors) of this model. [Table 4](#) details the results of the model including significant effects only. Estimates with an absolute t -value greater than 2 may be considered significant.

EFFECTS	ESTIMATE	STD. ERROR	T-VALUE
(Intercept)	2.68	0.10	26.16
Program (SB)	-0.85	0.12	-7.02
Year (-1, 0, 1)	0.84	0.07	12.72
SB:Year	-0.37	0.09	-4.01

Table 4 Mixed-effects model.

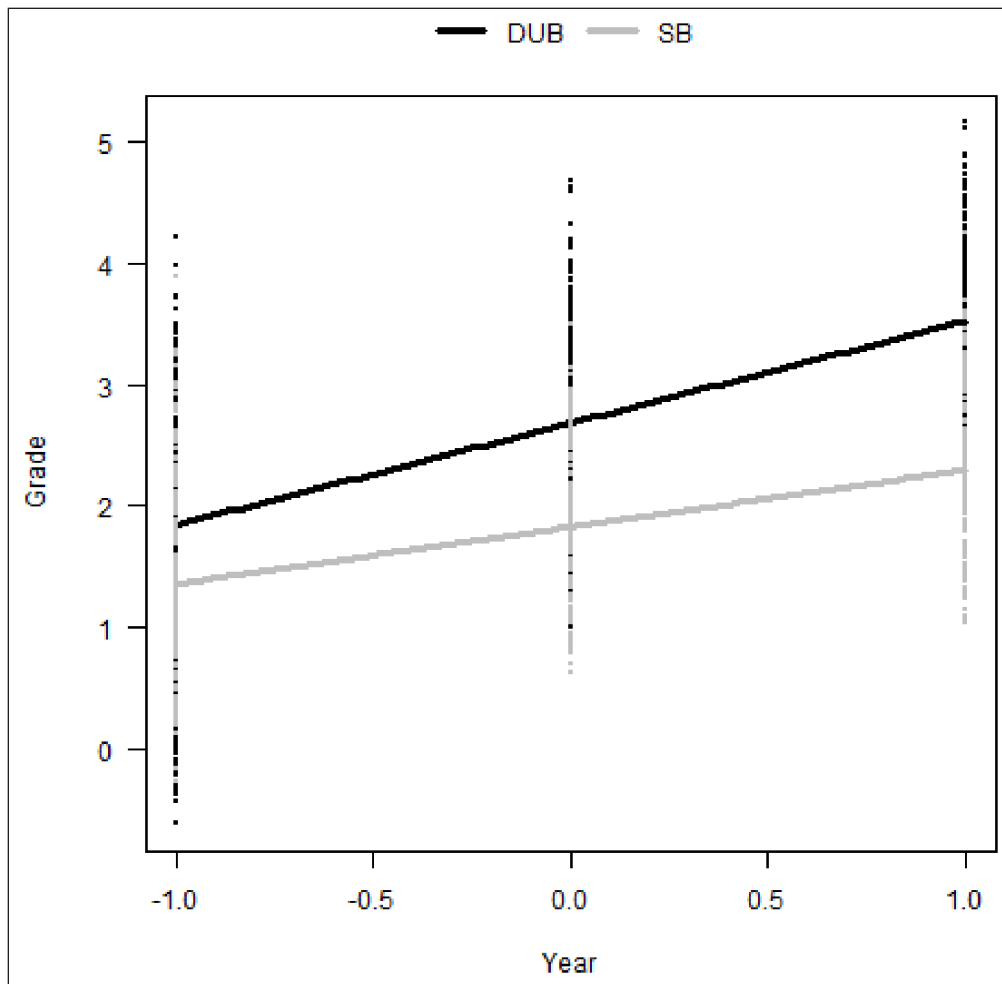


Figure 2 Interaction between year and program.

5. DISCUSSION

The current study has responded to DeKeyser's (2003) suggestion for more ecologically valid research in classrooms, with free response data, which tests learners' communicative skills rather than knowledge about grammar. We followed 229 students over three years in their respective L2 instructional programs. We compared a "weak" and a "strong" version of CLT programs used in the Netherlands: One program taught according to SB principles with a great deal of explicit grammar taught in the L1, and one in line with DUB principles, called AIM, in which the target language was used exclusively and grammar was taught implicitly. We tested oral and written skills after one, two and three years of instruction. Our research question was which instructional program was more effective in terms of general proficiency operationalized as free-response oral and written skills.

In line with Lightbown and Spada's (2013) conclusion and Cummins' (2014) reinterpreted findings of Bourdages and Vignola (2009), our results show that the strong communicative, meaningful approach with a great deal of exposure and interaction is more effective. Our findings are corroborated by earlier findings (e.g., Irshad et al, 2019; Verspoor & Hong, 2013) regarding DUB programs, which focus on meaning rather than form and included scaffolded comprehensible input with a great deal of repetition built in. We can say that AIM, the DUB program, was more effective in the development of L2 oral and writing skills after one, two and three years of instruction compared to the traditional SB programs used in the Netherlands. This may be very much in line with what most applied linguists would expect, as exposure has long been recognized as a driving force in L2 acquisition, but it is very much against common beliefs and practices held by teachers in the Netherlands (see Graus & Coppen, 2017) and teachers' common practice (see West & Verspoor, 2016).

As proponents of "strong" CLT versions would argue, the DUB program was probably more effective than the SB program because of the great amount of meaningful L2 exposure in a meaningful context. However, using the L2 exclusively from the beginning for three years with young learners is not an easy feat, but the AIM method facilitates this well. The teachers were able to speak comprehensible French during their lessons because of the carefully scripted method, with pared down language, lots of playful drilling and repetition, and the use of iconic gestures. Moreover, as language was offered in whole utterances, learners were able to pick up

not only words, but especially short phrases and clauses in their entirety. Just hearing and using phrases or chunks over and over again (it took almost one academic year to tell the *Three Little Pigs* story) helps to form strong associations between meaning and form and between the words themselves, so the whole sequence becomes entrenched (see Schmid, 2020). Moreover, each word (lexical and grammatical) has a gesture, so strong multi-modal associations are made. Despite the lack of creativity that some of the early activities seem to have, the data clearly show that learners were able to use constructions rather creatively in the testing contexts early on.

Teachers in the SB program varied a great deal in the amount of L2 they provided in the classroom, mainly because they used the L1 to explain the grammar rules and vocabulary was presented in lists. We suspect that these textbook methods do not include enough playful exercises that allow for repetition of words and phrases over time, so they do not become entrenched as they do in the DUB approach. However, there is no reason a creative teacher could not use a textbook method and spend relatively more time on the reading texts and listening exercises, asking learners to repeat whole chunks and making sure they understand every word. They could also make sure that they revisit and test the same vocabulary and chunks over time.

A “strong” CLT version relies on meaningful L2 exposure and interaction, but the question is whether such meaningful interaction alone contributed to the positive effects of the DUB program. A few studies suggest that it may also be due to the specific instructional program. In Bourdages and Vignola (2009), the two groups had equal amounts of L2 French exposure and interaction and the AIM group outperformed the non-AIM group in general proficiency measures and were the same in accuracy. One SB teacher in the current dataset spoke L2 French almost exclusively and both the oral and writing results of her class were compared to those of a DUB group with a similar scholastic level. The DUB group still outperformed this SB group in many respects, including accuracy in oral skills. Rousse-Malpat et al. (2021) showed that the DUB method led to greater speech rate, greater grammatical complexity and higher accuracy of the present tense and L2 use in speaking. For writing, Rousse-Malpat et al. (2019) demonstrated that the DUB method led to greater fluency and complexity at various morphosyntactic levels and also to a greater use of short routines in writing.

However, our findings are not in line with the conclusions from various meta-studies (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010), which found that explicit conditions are generally more effective. This is not surprising, as Norris and Ortega (2000) pointed out that their results were based on many brief interventions in which the tests were often biased because they were constrained tests focusing on morphosyntax and were often limited to accuracy. And, as Doughty (2003) has pointed out, if certain variables, especially exposure, had been controlled for better, explicit instruction would not have been found more effective. Furthermore, Andringa and Schultz (2016) have shown that controlling for the amount of exposure would have changed the results in Spada and Tomita (2010). Thus, if anything, our study has shown that instructional effectiveness needs to be studied over time, for at least one academic year and preferably more, mainly because learning to use a foreign language meaningfully takes time. In an implicit learning condition learners have to discover the patterns themselves, so a substantial amount of exposure will have to be provided before learners become accurate. As pointed out by VanPatten (1990), learners probably first focus on meaning and only later have attentional resources to recognize regular patterns in the language.

Besides allowing for more time in testing interventions, researchers should avoid constrained tests focusing on morphosyntax and focus instead on general language skills (speaking, listening, reading and/or writing), with less focus on grammar and more on general linguistic abilities (DeKeyser, 2017), so as not to bias for explicit learning. As far as tests are concerned, we used free response data in the form of oral interviews and brief narrative texts for these absolute beginners. This holistic approach took meaning-making in all its aspects into account. We can recommend using the SOPA method, as it is a validated test with excellent training modules. Moreover, it is not dependent on any particular instructional program. We observed that the young learners enjoyed taking the tests in pairs, felt very comfortable and were eager to participate. Moreover, once testers were trained, we found that it was a rather efficient way to test for oral skills. But most importantly, it helped us obtain reliable general proficiency scores at these lower levels. Setting up a similar rubric for the writing products helped us make the fine-grained differences necessary to discriminate between our own learners at these beginning levels that another scale such as the CEFR cannot make.

Of course, there are limitations to the study. Our SOPA results could not be independently rated due to time constraints. Another point is that we were not able to trace motivation over time as learners and teachers became less interested in filling out the forms over time. However, learners in the DUB program with the fast-paced activities seemed to be much more engaged than their SB peers.

Furthermore, in classroom studies, there are many other variables that cannot be controlled for, such as the amount of repetition or the use of gestures. Therefore, we would also welcome controlled laboratory studies that would explore issues such as the role of repetition and the contribution of the use of visuals and gestures. Laboratory studies would be valuable because the current study compared two programs that were inherently very different, and we do not know precisely what affected the differences. However, we were able to show that explicitness of instruction is not a prerequisite for greater effectiveness.

Finally, our results have also shown variation among groups. In both conditions, some groups did better than others, and within classes, some learners did better than others. Of course, taking a dynamic perspective, we would expect teachers and learners to be different. Some teachers or students may feel more comfortable or creative than others in a particular program. However, even if we could control for every factor, a dynamic perspective would predict variability in learners and variation among learners as individual learner characteristics will dynamically interact with cognitive, affective and contextual factors.

6. CONCLUSION

The current study examined the effects of instruction on L2 learning after three years. The approach traditionally used in the Netherlands is a “weak” CLT version in which learners are “learning to use” with an SB approach to language. The other is a “strong” CLT version in which learners are “using to learn” with a DUB approach to language. We showed that, over time, the DUB program was more effective than the SB program on general oral and written skills in L2 French. The DUB program offered the most L2 exposure by creating a greater number of L2 learning events in drills and repetition of frequently occurring patterns in a scripted and scaffolded input. The DUB program focused on its concepts of frequency, salience and contingency. Moreover, the embodiment found in the gestures and playful activities may have facilitated language development. Finally, the DUB learners were as accurate as the SB students, implying that explicit attention to grammar is not needed to become accurate.

The findings also suggest that we need to inform our teachers and textbook writers better. As a field, we need to explain the advantages of “strong” CLT programs and realize that there are different perspectives on language and how they may be learned. We should emphasize that meaningful exposure is the key to L2, that making errors is part and parcel of the developmental process, and that explicit grammar is not needed to become proficient.

AIM may be an excellent example of an instructional approach in line with DUB principles, but the disadvantage is that teachers need to spend quite a bit of time and money on training as they have to memorize the stories and gestures. Another disadvantage is that the method was created for young children and may work differently for older teenagers and adults. However, the field of second language acquisition is full of other good instructional approaches that bring exposure and meaningful interaction into the classroom, such as Content Language Integrated Learning and Task-Based Language learning, and we hope such approaches find their way to language classrooms.

NOTES

The assessment rubric for writing proficiency is available via the IRIS database: <https://www.iris-database.org/iris/app/home/detail?id=york:940373>.

ETHICS AND CONSENT

This study has been conducted before 2015 when the faculty of Arts of the University of Groningen first adopted a university-wide Research Data policy. This means that no ethical committee has been consulted to collect the data. However, the authors of this paper ensure that they conducted the study ethically. Written informed consent from the parents of the pupils involved in the study was obtained before the start of the study.

This study is part of a PhD project funded by NWO (Nederlandse Wetenschappelijke Organisatie).

AUTHOR AFFILIATIONS

Audrey Rousse-Malpat  orcid.org/0000-0002-4563-3085

University of Groningen, NL

Rasmus Steinkrauss  orcid.org/0000-0002-3643-8704

University of Groningen, NL

Martijn Wieling  orcid.org/0000-0003-0434-1526

University of Groningen, NL

Marjolijn Verspoor  orcid.org/0000-0001-8828-8405

University of Pannonia, HU

REFERENCES

- Andringa, S., & Schultz, K.** (2016, September 24). *Exposure confounds in form-focused instruction research: A meta-reanalysis of Spada and Tomita (2010)* [Conference presentation]. New York, NY, United States: Second Language Research Forum.
- Bates, D., Mächler, M., Bolker, B., & Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Baayen, R. H.** (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511801686>
- Bourdages, J., & Vignola, M. J.** (2009). Evaluation de l'habileté de communication orale chez des élèves de l'élémentaire utilisant AIM [Assessment of oral communication skills in elementary school students using AIM]. *Canadian Modern Language Review*, 65, 731–755. DOI: <https://doi.org/10.3138/cmlr.65.5.731>
- Brouwer, J., Christiaans, F., Hoogerwerf-Beijer, S., Hugon, G., & Morice, I.** (2014). *Grandes Lignes 5^{de} édition: Frans in de onderbouw* [French in junior high]. Noordhoff Uitgevers.
- Council of Europe.** (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Cummins, J.** (2014). To what extent are Canadian second language policies evidence-based? Reflections on the intersections of research and policy. *Frontiers in Psychology*, 5, 1–10. DOI: <https://doi.org/10.3389/fpsyg.2014.00358>
- DeKeyser, R. M.** (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Blackwell. DOI: <https://doi.org/10.1002/9780470756492.ch11>
- DeKeyser, R. M.** (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). Routledge. DOI: <https://doi.org/10.4324/9781315676968-2>
- 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.** (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. DOI: <https://doi.org/10.1017/S0272263102002024>
- Ebbinge-Planchon, L., Gaillot, M., Jansen, H., & Swed, Z.** (2013). *D'accord!1 havo/vwo / 1 vwo* (3rd ed.) [Alright! General and pre-university education]. Malmberg.
- Gombert, W., Keijzer, M., & Verspoor, M.** (submitted). *Structure-based versus dynamic usage-based instruction: L2 French writing skills after six years of instruction in high school*. Manuscript submitted for publication.
- Gombert, W., Vandendorpe, L., Keijzer, M., & Verspoor, M.** (submitted). *Chunks, complexity, and fluency in L2 French writing*. Manuscript submitted for publication.
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M.** (2015). Implicit and explicit instruction in L2 learning. *Implicit and Explicit Learning of Languages*, 48, 443–482. DOI: <https://doi.org/10.1075/sibil.48.18goo>
- Graus, J., & Coppen, P. A.** (2017). The interface between student teacher grammar cognitions and learner-oriented cognitions. *The Modern Language Journal*, 101(4), 643–668. DOI: <https://doi.org/10.1111/modl.12427>
- Howatt, A.** (1984). *A history of English language teaching*. Oxford University Press.
- Hymes, D.** (1971). *On communicative competence*. University of Pennsylvania Press.
- Irshad, M., Keijzer, M., Wieling, M., & Verspoor, M.** (2019). Effectiveness of a dynamic usage based computer assisted language program. *Dutch Journal of Applied Linguistics*, 8(2), 137–162. DOI: <https://doi.org/10.1075/dujal.16018.irs>

- Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63). CSLI.
- 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>
- Maxwell, W. (2001). *Evaluating the effectiveness of the accelerative integrated method for teaching French as a second language* [Unpublished Master thesis]. University of London Institute in Paris.
- 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>
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge University Press. DOI: <https://doi.org/10.1017/9781009024532>
- Rhodes, N. C. (1996). *Alternative assessment for immersion students: The student oral proficiency assessment (SOPA)*. Center for Applied Linguistics.
- Rousse-Malpat, A., Steinkrauss, R., Koote, L. & Verspoor, M. (2021). Parlez-vous français? Effects of structure-based vs. Dynamic usage-based approaches on oral proficiency. *Language Teaching Research*. Advance online publication. DOI: <https://doi.org/10.1177/13621688211040298>
- Rousse-Malpat, A., Steinkrauss, R., & Verspoor, M. (2019). Structure-based or dynamic usage-based instruction. *Instructed Second Language Acquisition*, 3(2), 181–205. DOI: <https://doi.org/10.1558/isla.38054>
- 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/dujdl.1.2.07rou>
- Rousse-Malpat, A., Verspoor, M., & Visser, S. (2012). Frans leren met AIM in het voortgezet onderwijs. Een onderzoek naar de effecten van AIM-didactiek op schrijven in het Frans van brugklasleerlingen [Learning French with AIM in Junior High. A study on the effects of AIM on L2 French writing]. *Levende Talen Tijdschrift*, 13(3), 3–14.
- 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. (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching*, 44(2), 225–236. DOI: <https://doi.org/10.1017/S0261444810000224>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature. *Language Learning*, 60, 263–308. DOI: <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Steenberg, L. (1996). *The Cito test of oral proficiency: Validity, reliability and practicality*. Arnhem, The Netherlands: Cito.
- Tammenga-Helmantel, M., & Majjala, M. (2019). The position of grammar in Finnish, Dutch, and global course books for German as a foreign language. *Language teaching research*, 23(5), 562–583. DOI: <https://doi.org/10.1177/1362168817752542>
- Thompson, L., Kenyon, D., & Rhodes, N. (2002). *A validation study of the student oral proficiency assessment (SOPA)*. Iowa State University.
- Tomasello, M. (2003). *Constructing a language*. Harvard University Press. DOI: <https://doi.org/10.2307/j.ctv26070v8>
- VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12(3), 287–301. DOI: <https://doi.org/10.1017/S0272263100009177>
- 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. DOI: <https://doi.org/10.1075/llt.48.08ver>
- Verspoor, M., de Bot, K., & Xu, X. (2015). The effects of English bilingual education in the Netherlands. *Journal of Immersion and Content-Based Language Education*, 3(1), 4–27. DOI: <https://doi.org/10.1075/jicb.3.1.01ver>
- Verspoor, M., & Hong, N. T. P. (2013). A dynamic usage-based approach to communicative language teaching. *European Journal of Applied Linguistics*, 1(1), 22–54. DOI: <https://doi.org/10.1075/jicb.3.1.01ver>
- Verspoor, M. H., & Schmitt, N. (2013). Language and the lexicon in SLA. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 353–360). Routledge.
- Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage-based perspective on L2 writing development. *Journal of Second Language Writing*, 21(3), 239–263. DOI: <https://doi.org/10.1016/j.jslw.2012.03.007>
- West, L., & Verspoor, M. (2016). An impression of foreign language teaching approaches in the Netherlands. *Levende Talen Tijdschrift*, 17(4), 26–36.

TO CITE THIS ARTICLE:

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), 20–33. DOI: <https://doi.org/10.22599/jesla.86>

Submitted: 28 February 2022

Accepted: 26 May 2022

Published: 12 July 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

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