University of Groningen

Effectiveness of explicit vs. implicit L2 instruction

Rousse-Malpat, Audrey

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

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

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Rousse-Malpat, A. (2019). *Effectiveness of explicit vs. implicit L2 instruction: A longitudinal classroom study on oral and written skills.* [Groningen]: Rijksuniversiteit Groningen.

Chapter 3
Explicit vs. Implicit L2 instruction in the classroom: beyond the dichotomy[2]

1. Introduction

After an abundance of studies carried out in the 1990's on the effects of explicit L2 instruction compared (or not) to implicit L2 instruction, several meta-analyses were published from 2000 onwards (Norris & Ortega, 2000; Spada & Tomita, 2010; Doughty, 2003; Goo, et al., 2015) making the balance on crucial questions remaining in instructed second language acquisition: Is instruction effective? What are the effects of explicit instruction on second language acquisition? Can implicit instruction also be beneficial? Which one is more effective? Which one leads to more fluency? Which one leads to more accuracy?

These meta-analyses involve a large number of definitions on explicit and implicit instruction. Most describe instruction as explicit if "rule explanation comprised part of the instruction or if learners were directly asked to attend to particular forms and try to arrive at metalinguistic generalization on their own" (Norris & Ortega, 2000; p.437) and as implicit if "neither rule presentation nor directions to attend to particular forms were part of a treatment" (Norris & Ortega, 2000; p.437).

The first conclusion in these studies is that compared to mere L2 exposure, L2 instruction is more effective, regardless whether it is explicit or implicit. Secondly, comparing different types of instruction and involving both laboratory and classroom studies, the studies conclude that explicit instruction within a meaning-based approach is more effective than implicit instruction, with relatively large effect sizes and durable positive effects (Norris & Ortega, 2000; Spada & Tomita, 2010).

However, the studies stress the fact that the question on the effect of different types of L2 instruction and their relation to different degrees of L2 exposure is yet to be answered. With an equal amount of L2 exposure, is explicit or implicit instruction more beneficial? Is explicit instruction necessary at the beginning of

---

2   This chapter is based on the following article:
    Rousse-Malpat, A., Steinkrauss, R., Wieling, M. & Verspoor, M. (submitted). Explicit vs. Implicit L2 instruction in the classroom: beyond the dichotomy.

L2 instruction? Or could it be replaced by L2 exposure combined with implicit instruction? Moreover, they add that the current findings may be biased as treatments have been brief, which is a disadvantage for implicit learning, and have used "highly constrained discrete-focus linguistic tasks", which favor explicit instruction (Spada, 2011, p. 228).

To meet such flaws, 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 focus on general language skills (speaking, listening, reading and/or writing) in order to fit the expectations of our modern world, which focuses less on grammar and more on general linguistic abilities (DeKeyser, 2017).

However, we must bear in mind that taking a classroom perspective on the matter would probably go beyond a strict dichotomy between explicit vs. implicit instruction as defined in most meta-analyses (Hulstijn, 2015). We have to accept that the type of instruction is always part of a program that involves other features such as L2 exposure, type of rule presentation, type of feedback, output and interaction.

The current study explores the effects of two teaching programs, one predominantly explicit and the other predominantly implicit, on the development of oral and written skills. We aim to find out which program is the most effective after three years of instruction and whether the amount of guided L2 exposure is perhaps more important than the degree of explicitness. To do so, we traced the L2 development of 229 junior high school students (aged 12 to 15) learning French in the Netherlands during three school years. They took three oral tests over the course of three years (total numbers of interviews n=568) and wrote seven identical narratives (total number of assignments n=1511). The oral tests were scored using the protocol and grid from the SOPA method developed by CAL (Thompson, et al., 2002) and the narratives were scored with a similar grid developed for this particular study. The results were compared according to the type of instruction and to the degree of exposure they had during those three years.

After summarizing the findings on the effectiveness of explicit vs. implicit L2 instruction to date, we argue that although much is already known on the subject, a longitudinal classroom study comparing programs as a complex ecological setting, including information on the degree of exposure could bring extra insights on the effectiveness of L2 teaching.

2. The role of explicit and implicit instruction

Considering the fact that the role of instruction should lead to "getting to a sufficiently high level of automatization to enable second language use that is both fluent and almost completely accurate" (DeKeyser, 2017; p.17), several researchers (DeKeyser, 2007; Ellis, R., 2005) have made a difference between declarative (explicit) knowledge and procedural (implicit) knowledge. Ellis, R. (2005), for instance, wrote a series of guidelines for foreign language teaching in the classroom context. One of his key principles states that "instruction needs to be predominantly directed at developing implicit knowledge of the L2 while not neglecting explicit knowledge with regard to explicit instruction and input" (p.36). So, to be efficient, instruction should bear on both types of knowledge. The question remains whether instruction should first focus on one knowledge and then on the other, or whether it should focus on both at the same time.

Researchers who have investigated this question disagree. DeKeyser (2007) advocates a focus on explicit declarative knowledge at the beginning of acquisition followed by a focus on procedural knowledge at a later stage, following the principles of the Skill Acquisition Theory. To him, language needs to be automatized by repeatedly tapping on declarative knowledge by forming habits that become procedural knowledge. R. Ellis (2006a) also points out the important role of automatization, which he calls implicit knowledge; however, he advises teachers to focus on both types of knowledge at the same time by presenting learners with patterns in frequently repeated input and by giving explicit information at the same time on forms that are not salient or too different from the L1. Some empirical studies, though, do not find any beneficial effects of developing explicit knowledge for the acquisition of some grammatical features (Sanz & Morgan-Short, 2004), whether it is provided before practice, during practice, or at both times.

It thus seems unclear whether explicit instruction has an added value in this process or whether implicit instruction, with enough time and L2 exposure, can achieve this goal as well. Most researchers agree that the main goal of a language program should be to develop implicit knowledge and automatization by using tasks that favor repetition, drills and pattern recognition from input that provides enough cues to the learners to develop explicit knowledge as well. However, the timeline involved in the automatization process of explicitly or implicitly instructed learners might differ greatly and therefore requires longitudinal studies to be investigated. Some even suggest that the benefits of each type of knowledge depend on the learner's level of proficiency. According to Correa (2011; in Gruber, 2012), and Han and Ellis (1998) beginners could benefit from instruction that taps

into both kinds of knowledge, whereas advanced learners could benefit more from instruction that taps particularly into explicit knowledge. Empirical studies to date on the subject have provided mixed results (Robinson, 1996; de Graaff, 1997).

Type of instruction in SLA usually refers to the way information about language forms, almost exclusively pertaining to morphology, are presented and practiced. There are many terms and dichotomies in the literature to refer to form instruction such as Focus on Form(s) vs. Focus on Meaning, inductive vs. deductive, feedback vs. recasts, explicit versus implicit. These terms cover a large range of didactic forms found in instructional programs that present learners to rules: from the most traditional side of the spectrum, called Focus on Forms by Long (1991), which focuses on explaining the rules explicitly to the learners without a communicative context, to the side of the spectrum most strongly resembling a natural acquisition setting, called Focus on Meaning by Long (1991), which focuses on the incidental acquisition of the L2 system with rich and meaningful input. Several studies have investigated the effects of type of instruction.

Day and Shapson (1991) conducted an experiment on French proficiency in a French immersion program. They compared a group who had some focus on grammatical forms to an immersion group with no focus on grammatical forms. In a pre-test/post-test experiment, children were tested in the use of the conditional. Results show that the group with focus on grammatical forms was better at writing and showed the most growth in speaking.

Erlam (2003) investigated the effects of a deductive approach to grammar, which involved rule presentation and explanation, compared to an inductive approach, which involved learners discovering the rules without explicit explanation on the acquisition of direct pronouns in French as a L2. She found that the deductive approach was the most effective. At the same time, she acknowledges that evidence in the literature is conflicting. Some find an advantage for inductive instruction (Herron & Tomasello, 1992), whereas others find the opposite (Robinson, 1996).

Lyster and Saito (2010) investigated the effects of different types of feedback by analyzing the results of 15 classroom-based studies on the topic. They found that feedback did have a positive effect on L2 development but that some types of feedback were more effective than others, particularly for free-constructed responses. The most explicit type of feedback (prompts) appeared to be more effective than the more implicit type of feedback (recast) and in general, feedback was more effective for younger than for older learners.

It thus seems that some kind of attention to form is necessary in effective L2 teaching and that the most explicit kind of instruction helped in acquiring particular language forms, but results are sometimes conflicting. Because of the plethora of terms in the literature and conflicting results, several meta-analyses have been attempting to give a clearer picture of the effectiveness of different types of instruction.

In general terms, meta-analyses on the effects of type of instruction show that explicit instruction is more beneficial than implicit instruction (Norris & Ortega, 2000; Spada & Tomita, 2010; Goo et al., 2015). The first meta-analysis by Norris & Ortega (2000) included quasi-experimental studies from 1980 to 1998 that were categorized according to their degree of explicitness and their type of attention to form. They compared explicit types of instruction (any kind of instruction that has some attention to form) with implicit types of instruction (instruction using mere exposure or no attention to form). They showed that instruction with any kind of focus on form (within communicative activities or as pure grammar instruction) was more effective than implicit instruction on language proficiency in general.

Spada & Tomita (2010) included 30 studies in their meta-analysis and focused on the effectiveness of instruction on complex target features. Their results are in line with Norris & Ortega (2000) as they found an advantage for explicit instruction over implicit instruction for both simple and complex grammatical features. They also suggest that explicit instruction helped to develop the learners' explicit knowledge, which they have been able to automatize in spontaneous speech.

In 2015, Goo et al. revisited Norris & Ortega's data (2000). Eleven studies from Norris & Ortega were included as well as 23 new studies published between 1999 and 2011. They found again that explicit instruction (either given orally and/or in writing) was more effective for oral and written proficiency. Also, they made a difference between feedback and no feedback. No differences were found on the effectiveness of either type of feedback for implicit instruction. However, in the explicit condition, the condition without feedback was better. Regarding spontaneous speech, they found that the explicit condition gave better results. Comparing laboratory with classroom studies, they noted that larger effect sizes were found in laboratory studies.

Thus so far meta-studies give a clear and uniform picture of the differences between explicit and implicit instruction, suggesting that explicit instruction is better, and that explicit instruction might be faster to have an effect than implicit

instruction. However, we must acknowledge that the nature of a meta-analysis is to merge different contexts and different operationalizations of type of instruction, and that their definition of implicit instruction was particularly extreme, including only one end of the spectrum of implicit instruction (actually suggesting no instruction at all). Moreover, the tests in the underlying studies were often constrained, which may bias explicit instruction (Norris and Ortega, 2000).

Some classroom studies, however, involve free-response data, and may give a more detailed picture of the processes involved in each type of instruction on both oral and written skills. They show very clearly that explicit instruction involves other processes and thus has different beneficial effects from implicit instruction. Some studies focus on discrete language features, others use a more holistic approach.

Andringa, de Glopper & Hacquebord (2011) used a grammatical judgement test and free written response data of 81 teenage learners of Dutch receiving either explicit or implicit instruction on two different grammatical structures. It was a longitudinal study with three different measure times. Different results were found for each kind of assessment. The results of the grammatical judgement test showed that explicit instruction was more effective than implicit instruction. However, in the free-response task, explicit and implicit instruction equally led to the use of the grammatical structures. The difference was that explicit instruction was more effective for learners whose L1 was similar to Dutch. For learners whose L1 differed from Dutch on those structures, explicit instruction seemed to have a negative effect.

Hernandez (2011) gives more insight into the role of instruction combined with L2 exposure. Using free oral production data, he compared the effects of explicit instruction combined with input flood (n=36) to input flood only (n=30) and a control group with no intensive input flood treatment (n=25). He focused on discourse markers in the oral narration of an event in the past by 91 L1 English adults following a Spanish course. Results show that both treatments had beneficial effects compared to those of the control group but that the group with instruction combined with input flood used more discourse markers in their speech than the input-only group.

Taking a more holistic approach, Gruber (2012) compared an English program (with a more implicit skills approach) and a German program (with a more explicit structure approach) for L2 learners of French (14-16 years old) on vocabulary size and writing proficiency (lexical richness, accuracy, syntactic complexity).

She included semi-structured interviews, questionnaires and observations and marked the interviews using a grid inspired by CEFR. Results show that explicit instruction was better at accuracy and essay scores and that implicit instruction is better at syntactic complexity and lexical richness.

It thus seems that any kind of L2 instruction combined with L2 exposure can be beneficial to different aspects of the target language, possibly depending on the stage of acquisition. However, the role of L2 exposure in the effects of explicit vs. implicit L2 instruction has received little attention so far.

3. The role of meaningful language use and attention

The role of L2 meaningful language use—which encompasses exposure, input, output, and interaction—in the process of language learning remains unclear, particularly with regard to the question whether it could replace explicit instruction in developing explicit knowledge. Research on the role of input and L2 exposure usually are in line with Krashen's (1981) Input Hypothesis and more recently also with usage-based theories of language acquisition, which hold that language learning, either native or non-native, emerges from the recognition of salient patterns in the input. Learners are assumed to link these forms to meaning, but mere exposure may not be sufficient to maximize grammatical accuracy. Instruction should then support the acquisition of patterns by means of meaningful input in which the patterns are repeated and made salient. The input presented to the learners should thus be adapted to the proficiency level of the learners, enabling the processing of meaning first and then form (Van Patten, 2002). Studies show that providing activities designed to process the input first was more beneficial than providing explicit instruction and then L2 exposure (VanPatten & Oikkenon, 1996).

We saw earlier that input or exposure alone are not enough to effectively teach a second language but output alone is not effective either. Swain & Lapkin (1995) present the Output Hypothesis based on studies in immersion programs in Canada. They claim that learners become aware of a linguistic problem when they produce language, which pushes them to change their output. While talking, learners become aware of their linguistic gaps and engage in a grammatical analysis. In another study, Swain (1985) claims that output pushes the learner to move from involving semantic processes to syntactic processes.

Disagreement exists though on the moment when students should produce output. Swain's (1985) Output Hypothesis states that output should replace part of the input because it provides the practice that learners need to develop language.

De Bot (1996) also sees output as an opportunity for automatization and claims that it can help in the development of fluency. Other researchers question the importance of practicing output early on (DeKeyser and Sokalski, 1996) and advocate drilling and repetition at first to favor automatization and thus acquisition.

Output alone, though, would be useless if it wasn't part of communication, which leads to the importance of interaction. In interaction, learners can talk about the linguistic gap noticed in the output of the others and negotiate meaning. Long's (1983) Interaction Hypothesis suggests that this creates new modified input from which learners can learn new patterns and forms. This way, output transforms into input, the learners can direct their attention to salient forms and notice them, which drives the acquisition process (Gass, 2017; Swain, 1985).

To summarize, we have argued that L2 instruction should promote procedural and automatized language. So far, few studies have actually focused on this particular aspect of L2 development but have focused mainly on grammatical accuracy, often in constrained tests. These studies have generally shown that explicit instruction is more effective than implicit instruction. However, the operationalization of explicit and implicit instruction in many studies has been rather unbalanced, often reducing implicit instruction to no attention to form at all or mere L2 exposure. Only a very limited number of studies have actually investigated the topic in an ecologically valid, longitudinal setting and have controlled for other features of L2 instruction such as the amount of L2 use. The results of the few classroom studies show that both types of instruction have beneficial effects on language acquisition but not in the same way. Explicit instruction may favor accuracy and essay writing, but implicit instruction may benefit complexity and fluency. However, in these particular studies the role of L2 use in terms of exposure remains unclear. Therefore, we believe it is important to take a perspective on the topic that encompasses more than grammatical accuracy in short-term artificial settings: the study needs to be ecologically valid in that it is classroom based and longitudinal. Moreover, it needs to go beyond grammatical accuracy in constrained tasks. Finally, the amount of actual L2 use in the classroom needs to be taken into consideration. To do so, the current study compares two types of instructional programs, one mainly explicit and the other mainly implicit, over the course of three years mainly on procedural, automatized language use by means of free-response oral and written data.

Research questions
1. If effectiveness is operationalized as higher general proficiency in oral and written language use after three years of instruction, measured by (semi)free production tests, which type of instruction is more effective at the beginning of L2 acquisition in a foreign language context?
    a. A predominantly explicit form-focused program?
    b. A predominantly implicit form-focused program?

2. Which predictor is the most important in explaining our results: type of instruction or L2 exposure?

4.   Method

In the current chapter, two instructional programs are compared. Both of our programs give some kind of information about the structure of the target language, but in the current chapter one program is predominantly explicit, where "learners are encouraged to develop metalinguistic awareness of the rule" (Ellis, 2008; p.438) and is predominantly implicit, where learners are encouraged to learn from the input in which language forms are learned inductively and the feedback is limited.

4.1 Participants

The study included 229 (Explicit = 92; Implicit = 137) Dutch high school L2 learners of French aged 13 at the beginning of the study enrolled in 5 different schools (regular and English bilingual schools) which were followed for three years. The learners had no previous knowledge of French, so the learners can be considered absolute beginners.

In total, 14 teachers were involved in the study. To control for teacher effects as much as possible, we pre-selected the teachers by means of interviews and questionnaires given to both teachers and learners. All teachers had a high level of French proficiency (C1 according to the CEFR; Council of Europe, 2001), a high motivation to teach with the method they were teaching with, a considerable amount of experience (more than 5 years of teaching experience), and a good rating by their learners on teaching qualities. Some groups kept the same teacher throughout the three years; others had a new teacher every school year.

The learners were grouped according to scholastic level. In the Netherlands, high school students are streamed according to their scholastic level, which is tested at the end of primary school by the Cito test (Steerenberg, 1996). There are three general scholastic levels: low, intermediate and high. Within the highest

level, however, two further levels can be distinguished: the particularly gifted students (who spend relatively more hours on foreign language learning in bilingual and plus classes) and the very good students. In our study, we had four scholastic levels: intermediate (level 1), high (level 2), very high (level 3) and linguistically gifted (level 4).

Participants were placed in classes by their schools, but they did not choose in advance with which method they would be learning French. We distinguished two groups of participants: those learning French with an explicit program (*Grandes Lignes* or *D'accord*) and those learning French with an implicit program (AIM). All participants kept learning with the same teaching method throughout the three years of the study. Table 7 gives an overview of how the sample was distributed.

Table 7. Background information of the empirical research.

	Program	Teachers in total	Nb of classes in total	Scholastic level	Nb of participants
School 1	Implicit	A	2	4	47
School 2	Explicit	B, C	1	4	24
School 3	Implicit	D, E	2	3	49
School 4	Explicit	F, G	2	3	28
School 5	Implicit	H, I	6	2	12
	Explicit	I, J, K, L	12	2	15
	Implicit	H, I	6	1	29
	Explicit	J, K, L, M, N	15	1	25
Total					**229**

4.2 Teaching programs

In classroom studies, we must go beyond the dichotomy of explicit vs. implicit instruction. The reality of the field shows that teachers combine many features that are often opposed in experimental settings (communicative teaching materials combined with traditional structure instruction or implicit instruction combined with explicit gestures). Therefore, it should be noted that we are comparing programs or curricula as a whole and that features other than "explicit" or "implicit" instruction must be taken into account whilst describing each type of instruction. The "explicit" group had a program that combines communicative task-based materials with a focus on structure and forms. It follows what Spada (1997) would call "Explicit form-focused instruction".

The implicit group had a program that combines communicative language teaching principles, implicit and then inductive grammar instruction with as little feedback as possible, immersion principles, repetition and drills, pushed output and a high focus on meaning, in line with the DUB approach (as defined in *Chapter* 2).

4.2.1 The explicit program

In this study, the explicit group was taught with the two most popular programs for French as a Foreign language in the Netherlands: *Grandes Lignes*[3] and *D'accord*[4]. Both programs are designed by Dutch educational publishers as Communicative Language Teaching methods and have used the can-do statements of the CEFR (in our case A1 to A2) as baselines. They are topic-centered and focus on all competences, but particularly on writing. They offer explicit explanations and scaffolded exercises to learn and practice grammatical rules, and provide a bilingual vocabulary list for each chapter.

4.2.1.1 Topic-centered

Both books are organized around topics such as *Introduce yourself, Your house, Your hobbies, Friends and family, etc…* which are worked out in the listening and reading texts and in the bilingual vocabulary lists that are provided.

4.2.1.2 Focus on all competences

In each chapter, there are several exercises to practice all four competences. The chapters start with a listening exercise, followed by a reading exercise. Then, focus will be put on writing and grammar and finally on speaking and practicing chunks. There is also some attention to pronunciation. At the end of each chapter, there is a final task that combines the four competences.

4.2.1.3 Explicit grammar explanation

The book offers explicit explanation of the grammatical rules in the L1. Grammatical rules are displayed from simple, such as articles (*le/la; un/une*) and the present tense, to more complex, such as possessive pronouns and the past perfect tense.

4.2.1.4 Tests on grammar and competences

Learners are tested on receptive skills (listening and reading) and on their comprehension of the grammar. Some teachers used tests provided by the book publishers, others designed their own tests. Learners prepare for the test at home or in the classroom.

---

3   Noordhoff Uitgevers, Groningen
4   Malmberg Uitgeverij, 's-Hertogenbosch

By means of classroom observations, we were also able to describe how L2 instruction was implemented. Regardless of the book they used, the "explicit" teachers were comparable in the way they were teaching. They had a high focus on reading and writing and used corrective feedback. In most explicit classes, there was a low to medium amount of L2 exposure and use.

4.1.1.5 Focus on reading and writing (group work, whole class)
Lessons are designed around the activities in the book. They usually start with the correction of the homework as whole class activity or in small groups, followed by a new item that needs to be mastered (listening exercise, reading or grammar). The teacher gives corrective feedback to learners individually or to the whole class. Sometimes, a recast is used but most of the time, the teacher relies on the L1 for clarification. Teachers also design their own activities and we observed some learners practicing a small (role) play.

4.2.1 The implicit program
In this study, the "implicit" group was taught with a rather new method in the Netherlands, called the Accelerative Integrated Method (AIM; Maxwell, 2001). In *Chapter* 2, we argued that the method uses Dynamic usage-based principles in language teaching (Verspoor & Hong, 2013; Rousse-Malpat & Verspoor, 2018; Verspoor, 2017) in that it includes a great deal of "chunks" in the input, meaningful language use and a great deal of repetition. With AIM, gestures are used to facilitate comprehension and word retention, which allows the lessons to be in the L2 from day 1.

4.2.2.1 Story-based
Stories are the main source of L2 input in the classes. They are written according to the level of the learners and contain much repetition and pared-down language. In this scripted method, the teacher tells the story to the learners in chunks and the learners repeat the story as in a pattern drilling exercise. One story is usually used for the whole school year. In order to teach AIM, teachers follow a course at the university of Amsterdam or in Canada where they learn the basic principles of the method and the gestures. The lessons given by the teachers are scripted, ensuring uniformity of the program all over the world. A more detailed lesson plan of AIM is provided in *Chapter* 2.

4.2.2.2 Focus on oral skills first, later on writing skills
The stories are presented and practiced orally with the learners. During the first six months, attention is put on oral and listening skills. In a later stage and when learners are already able to maintain a small conversation in the target language,

written language is introduced. This is particularly interesting for French, as the written language differs quite much from the oral language and could influence pronunciation. Exercises move from repeated to creative language use.

4.2.2.3 Inductive approach to grammar

The way grammar is taught changes according to the stage of acquisition in which the learners are. During the first year, grammar is purely implicit. There is little attention to form as all the attention goes to the meaning. From the second year onwards, grammar is taught inductively. There is no pre-planned rule presentation. Learners are encouraged to find the rules themselves. The teacher is not allowed to give the rule until the learners have discovered it themselves. There is almost no feedback, only in the form of recasts or questions addressed to the entire class in the following form: *Do we say "it were beautiful or it was beautiful?"*.

4.2.2.4 Tests on general language competence

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.

4.2.2.5 Use of gestures

For every word used in class, even functional words such as articles, there is an iconic gesture, which the teacher presents with the introduction of every new word and which the learners repeat (teachers can take an on-line course to learn the gestures). Until they are fluent enough to do without, learners practice the language in combination with the gestures. Several studies have shown that gestures enhance comprehension, also in natural speech, because they can compensate for degraded speech (Drijvers & Özyürek, 2017). For L2 learning, the gestures allow the teacher to use the L2 from day 1.

4.3 Amount of exposure

Initially, we did not anticipate a large difference between the explicit and the implicit group in terms of exposure and L2 use. However, while observing classes, we realized that the two programs differed a great deal in this respect and the amount of exposure emerged as an important factor to take into account. The implicit type of instruction led to much more L2 exposure and use than the explicit type of instruction. Within the explicit condition, however, there was also a great deal of variation in terms of L2 exposure and use so this variable was included.

To estimate the amount of L2 exposure, we calculated 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 of the teachers in the background questionnaire, actual classroom observations, and verification by the teacher:

Teacher Questionnaire questions:
- When students arrive in the classroom, I greet them in French.
- I tell them the program of the lesson in French.
- When a student asks a question, I answer in French
- I give instructions in French.
- I explain grammar in French.
- I give homework in French.
- I always speak French in class.

Then, we used the classroom observations to count in minutes how much time was spent in the L2, either when the teachers talked in French or when the learners were practicing one of the four skills in French. We recalculated the minutes as percentages and we averaged the percentages for each class.

We combined our results from the questionnaire with our results from the class observation analysis, and we finally asked the teachers whether they agreed with the percentage we calculated. Since all of them agreed, we were able to calculate the amount of hours spent in the L2 per year. Table 8 gives the distribution.

Table 8. Distribution of hours of L2 exposure per year.

	Program	% of L2 exposure	Year 1	Year 2	Year 3
School 1	Implicit	90	63	63	63
School 2	Explicit	60	48	48	48
School 3	Implicit	95	90	60	45
School 4	Explicit	40	32	16	32
School 5	Implicit	90	90	90	90
	Explicit	50	50	50	50

4.4 Tasks and procedures

4.4.1 The oral proficiency task

To be able to remain consistent in our scoring of oral proficiency over the years, we used a standardized, validated oral proficiency test (Thompson et al, 2002; Rhodes, 1996). Developed in 1991 by the Center for Applied Linguistics (CAL) for Spanish students of English, the 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. The aim is to be able to interview and rate two students at the same time. One of the researchers is the interviewer and the other researcher is the rater. During the interview, s/he fills in the rating sheets that contain some practical information (names, class, date, etc.), but their main purpose is to help to decide on the final assessment of the students' performance. The rating sheet contains a separate box for each task in which remarks and notes can be written down relevant to assessing the oral proficiency. Appendix A contains an example of the rating sheets that we used for this study.

The particular combination of students in pairs were decided upon by the teachers who were asked to match students according to their level of proficiency and personality. We wanted to avoid situations where one student would dominate the conversation. Each interview was video-taped and took about 20 minutes, followed by a 10 minute break in which the two researchers had to come to an agreement on the assessment of the learners.

Learners were asked to perform three tasks increasing in complexity. The first task, meant to put the learners at ease, is called "fruits and vegetables" and is mainly a receptive task, which increases in complexity from pointing at fruits (in our case there were also blocks with colors) to moving them according to commands. Occasionally, learners are asked to identify and name things in the form of single words. When the students do relatively well on this task, the interviewer moves to the next

task called "all about you", which is a semi-structured interview about topics that are familiar to the learners (e.g. school, weather, holidays and sports). Interviewing students in pairs prevents them from becoming too nervous as they are allowed to help each other out. The last task is called "the farm" and is meant to create an environment where more complex commands can be asked. With the help of a 3D farm and characters, the interviewer encourages the learner to tell a story.

At all times, students are made to feel comfortable with compliments and encouragements and when the learner's ceiling level has been reached, the interviewer goes back to an easier task so that the learners feel they have done well. At the end, learners receive a reward in the form of a candy-treat.

4.4.2 The writing proficiency task

To test for writing proficiency, we used free writings in the form of narratives. The free writings were viewed as an example of the communicative and cognitive abilities of the learners, which show writing proficiency at that moment in time. We are aware that the exercise of writing is complex in itself as it demands the activation of many cognitive processes in order to convey a series of ideas. However, it was used successfully in several other studies on language proficiency (Verspoor, Schmid & Xu, 2012; Hong, 2013; Verspoor & Smiskova, 2012; Irshad, 2015), and proved to suit beginners better than other genres (Crowhurst & Piché, 1979).

The narratives were written in the classroom during 20-30 minutes. Learners were instructed to write a text according to the following criteria:

(1) the story is interesting to the reader,
(2) the author should use as many words as they could,
(3) 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.

Most students wrote every other week on topics related to the lessons. However, learners of all groups were asked to write seven narratives on the same topic during the 3 years of observation. We avoided hypothetical situations, which have proven to be difficult because of the use of conditional and subjunctive (Irmsher,

1979; Kirrie, 1979) as much as possible, by stating our instruction in the present tense and by ending the instruction by "Tell the story". The context of the narrative was made clear by a picture and a prompt sentence such as *This is Padma the alien, tell his story* accompanied by a picture of an alien going on a trip. The topics followed the curriculum of the learners so that they would know enough vocabulary to write a small text. The test started 5 months after the first French lesson and continued during three school years. See Table 9 for details.

Table 9. Written assignments per condition.

	Nb of months of instruction	Topic	Nb of participants	
			Explicit	Implicit
Year 1	4	Talk about yourself!	138	174
	9	Padma (implicit group) / Jane(explicit group) 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 (implicit group) / week (explicit group) 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

4.5 Scoring

4.5.1 Oral proficiency

The scoring of the SOPA tests was done by well-trained and certified SOPA/EL-LOPA assessors. Immediately after the testing session, the interviewer and the rater discuss and mutually agree on the scores for oral fluency, vocabulary and listening comprehension, using a standardized rubric (see Table 10) based on the developmental stages of language learners ranging from fixed formulae, unsuccessful creative language to successful conventionalized ways of saying things. The testing sessions were video-taped so scores could be double-checked and transcribed where needed for further quantitative analysis.

Table 10. SOPA descriptors of scores for oral proficiency.

	Oral Fluency	Vocabulary	Oral Comprehension
1	Produces only isolated words (i.e., single-word responses) and/or greetings and polite expressions such as good morning and thank you	-Uses single words in very specific topic areas in predictable contexts. -May use greetings and polite expressions.	-Recognizes single, isolated words, greetings and polite expressions.
2	-In addition to isolated words, uses phrases of two or more words, and/or memorized phrases or sentences (e.g., My name is…, I don't know) in predictable topic areas. -May attempt to create sentences, but is not successful. -Long pauses are common.	-Uses single words, short phrases, greetings, polite expressions, and other memorized expressions on a limited number of topics. -Frequent searches for words are common. May use native language or gestures when attempting to create with language.	Understands predictable questions, statements, and commands in familiar topic areas (with strong contextual support), though at slower than normal rate of speech and/or with repetitions.
3	-Uses memorized expressions with reasonable ease. -Shows emerging signs of creating with the language to communicate ideas. -Creates some sentences successfully, but cannot sustain sentence-level speech.	-Uses vocabulary centering on basic objects, places, and common kinship terms, adequate for minimally elaborating utterances in predictable topic areas. -Use of native language and gestures is common to expand topics.	- Understands simple questions, statements, and commands in familiar topic areas, and some new sentences with strong contextual support. - May require repetition, slower speech, or rephrasing.
4	-Goes beyond memorized expressions to maintain simple conversations at the sentence level by creating with the language, although in a restrictive and reactive manner. -Handles a limited number of everyday social and academic interactions.	-Has basic vocabulary for making statements and asking questions to satisfy basic social and academic needs, but not for explaining or elaborating on them. -Use of some native language is common when vocabulary is lacking.	-Understands familiar and new sentence-level questions and commands in a limited number of content areas with strong contextual support for unfamiliar topics. -Follows conversation at a fairly normal rate.

5	-Maintains simple sentence level conversations. May initiate talk spontaneously without relying on questions or prompts. -Gives simple descriptions successfully. -May attempt longer, more complex sentences. Few, if any, connectors are used.	-Has basic vocabulary, permitting discussions of a personal nature and limited academic topics. Serious gaps exist for discussing topics of general interest. -If precise word is lacking, may use circumlocution ineffectively. May resort to native language.	Understands sentence-level speech in new contexts at a normal rate of speech although slow -downs may be necessary for unfamiliar topics. -Carries out commands without prompting.
6	-Initiates and sustains conversations by using language creatively. -Shows emerging evidence of paragraph-like speech with some connected sentences (e.g., then, so, that, etc.) in descriptions and simple narratives, without actual paragraphs.	-Has a broad enough vocabulary for discussing simple social and academic topics in generalities, but lacks detail. -Sometimes achieves successful circumlocution when precise word is lacking. May use native language occasionally.	- Understands longer stretches of connected speech on a number of topics at a normal rate of speech. -Seldom has problems comprehending everyday topics. (Can request clarification verbally.)

4.5.2 Writing proficiency

To be able 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. Using written samples collected for the study reported in Rousse-Malpat & Verspoor (2012), we asked five experts in French as a second language to sort 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 sortings 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 the following grid:

Table 11. Evaluation grid of writing proficiency.

0	**Nothing / copied only first sentence**
	The student wrote (almost) nothing down.

1	**Dutch, short and/or bizarre texts**
	The text cannot be understood by French readers. There are many Dutch words and sentences. If there is some French in it, it is basic vocabulary (bonjour, au revoir), which is repeated over and over again, and the text is filled with grammatical errors and spelling mistakes. There are structurally unexpected sentences and there is no logical order and content in the story. Most texts at this level do not exceed the length of 3 sentences. Also very short texts with only basic chunks and still grammatical errors go into this category.
	Je suis le foot et le tennis et le dans et le tir à l'arr. Et j'adere le musique.

2	**Emergence of some French sentences (and memorized sentences)**
	Some parts of the text can be understood by French readers. These are mainly memorized sentences. Dutch may still be present in many sentences. The French is mainly basic vocabulary and there are many grammatical errors. There are also many spelling errors, but the spelling is phonetically adequate (when pronounced, it sounds French: sjèze for chaise). The sentences are not linked by linking words. There is some logical order and vague story content. Most texts are longer than 3 sentences.
	Je suis fan de Hef. Il est à rapper. Je suis aussi fan de Joey Diamond. J'écoute souvent Hef avec me mobiel & ipod.

3	**Understandable, but simple French text and emergence of creative language**
	The text is minimally understandable for French readers. The text has clear content and meaning. There may be some Dutch words. The text is written in simple sentences (no complex syntax) and each sentence shows that the student understands the basic sentence structure of subject – verb – complement, although the word order may be erroneous in some sentences. There is no erroneous subject omission. The student still uses basic vocabulary, but uses it mostly in an appropriate and understandable way. The story coherence is minimal, but some linking words may be present.
	Bonjour. Je m'appelle Paul. Je veux va à ma mère. Elle habite à Tokyo au Japan. Et je habite à Toronto au Canada. Alors, je veux va a un avion ou une bateau.

4	**Short and correct French text and/or emergence of more complex French (with some errors)**
	This level contains two types of texts: either short (maximum of 5 sentences) and correct (absence of spelling and grammatical errors, many chunks, authentic), or long and creative with a few more errors in spelling, no more than 3 Dutch words, and correct grammar. There is still some 'directly translated Dutch'. The spelling is quite correct and there is at least one complex sentence and a few linking words between sentences, assuring coherence. Some other verb tenses may be used incidentally.
	Le extraterrestre promené sur le planet. Il va voir la porte noir. Il est sur la porte et il tomb dans un autre planet. Planet Aarde. Il voit il est fantastique maintenant. Il cherche un maison. Mais tout le monde qui voire le extraterrestre crié ou se court. Le extraterrestre est triste. Il ne va bien. Le prochain jour il parte avec un avion de Samabava. Au revoir extraterrestre.

5	**Coherent and more complex French and logical order of events**

The story content is coherent and easy to follow by French readers. Events in the story are linked by logical linking words and there are no structurally unexpected sentences. The story is longer than 6 sentences and contains a lot of chunks and authentic French. The vocabulary is appropriate and the student does not use Dutch words. There may be minor spelling and grammatical errors, but these do not hinder the message of the text. Complex syntax and emergence of various tenses that are applied correctly.

Un jour il est poursuivi par un cochon Volant. Padma va vite a la maison et va dans son soucoupe volant, parce que il pense que le cochon volant ne peut pas le trouvé. Alors le cochon volant trouve Padma. Padma decide de partir vers la terre, comme il est peureux à emmener vers la maison du loup. Il part et tombe dans leau sur la terre. La fin.

6	**Long texts of simple narratives, emergence of paragraphs**

The story starts to have a clear beginning, middle and end with the emergence of paragraphs, although not consequently. It is longer than 10 sentences containing chunks of authentic French, complex sentences made of difference coordinate and relative clauses, as well as many verbs varying in tenses. Several tenses of the past may used, the future tense might emerge. The present tense is correctly conjugated but incorrect forms of the other tenses remain. The learner is able to tell a story by using the language creatively but grammatical inaccuracies remain. Spelling, however becomes consistently more accurate.

Après quelques heures, ils décidaient qu'ils devaient inviter des autres copains pour jouer de football. Maintenant, il y a neuf copains et ils ont fait trois equipes.

Après que la competition finit, les copains sont allés boire du vin. Maintenant, tout le monde a bu beaucoup trop de vin et ils retournent à la pluie où on fait la competition.

This grid was validated by asking two other raters to score 475 assignments that were collected in the current project. An interrater reliability (IR) analysis using the Spearman Rho correlation was performed to determine consistency among the 2 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 IR 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 exercise (0.74).

4.6 Analysis

The oral and written data was analyzed using mixed-effects regression modeling with R (version 3.4.1 with lme4 package). The significance of several predictors of interest was assessed: Program (AIM vs. GL/DAC), Type of task (oral vs. written), Time of testing (Year 1, Year 2, Year 3) and Amount of L2 exposure they received every school year (expressed in hours). 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.

The inclusion of random intercepts for Participant, School, Class and Teacher and random slopes for each of these levels for the predictors of interest was assessed. 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).

Because of the multicollinearity of method and amount of exposure (the AIM method inherently offered more French exposure to the students, see Table 8 above), it was not possible to use both predictors in the same model. Therefore, we built two different models, one assessing Program, Type of task, and Time of testing as predictors, and one assessing Amount of exposure, Type of task, and Time of testing as predictors. In both models, the average of the holistic oral and written score of each student was the dependent variable. The random-effects structure was the same for both models. By subsequently comparing the two models, this approach enabled us to evaluate whether the program (Model 1) or the amount of exposure (Model 2) were better predictors of our data.

5. Results

5.1 Descriptives per year[5]

Table 12. Mean and standard deviation per year and per group.

	Year 1		Year 2		Year 3	
	Mean	SD	Mean	SD	Mean	SD
Explicit	1.56	0.73	2.21	0.87	2.44	0.88
Implicit	2.1	0.99	3.02	0.89	3.38	0.79

5.2 Determining the random effects and significant factors for our models

Using the procedure of multiple pairwise comparisons explained in the previous section, we determined that the best random intercepts to use for our models were Participant, Class and School. 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 shows that there was much individual and class-related variability between the participants throughout the years.

---

5   See appendix C for a complete overview of the descriptives.

Two models with the same random effects were fitted to the data. Regarding the fixed effects, only the factors Year and Program (model 1) or Year and L2 exposure (model 2) turned out to be significant predictors. The factor Type of assignment (oral/written) did not appear to be a significant predictor and it did not interact significantly with other fixed factors in either model.

5.3   Results of model 1: the program as a predictor

Using model comparison including the aforementioned predictors and random effects, the following model specification was determined to be optimal: Grade ~ Program*Year + (1+Year|Participant) + (1+ Year|Class) + (1|School).

Table 13 details the results of model 1 including only significant effects. Estimates with an absolute t-value greater than 2 may be considered significant. As the table shows, there is a clear effect of the program and the year. Overall, the learners 'scores increase over time (Estimate = 0,84). The scores of explicit group, however, increase significantly slower than those of the implicit group (Estimate= -0,37). We can also see that the scores of the explicit program are significantly less high than those of the implicit program (Estimate = -0,85). This is visualized in Figure 3.

Table 13. LMER model focusing on the factor Program to predict the scores of oral and written skills.

	Estimate	Std. Error	T value
(Intercept)	2.68	0.10	26.16
Program (Explicit)	-0.85	0.12	-7.02
Year (-1, 0, 1)[6]	0.84	0.07	12.72
Explicit:Year	-0.37	0.09	-4.01

---

6   The factor "Year" was centered for the purpose of the analysis, we replaced Year 1,2,3 by -1, 0 and 1.

Figure 3. Visualization of the interaction Year and Program.
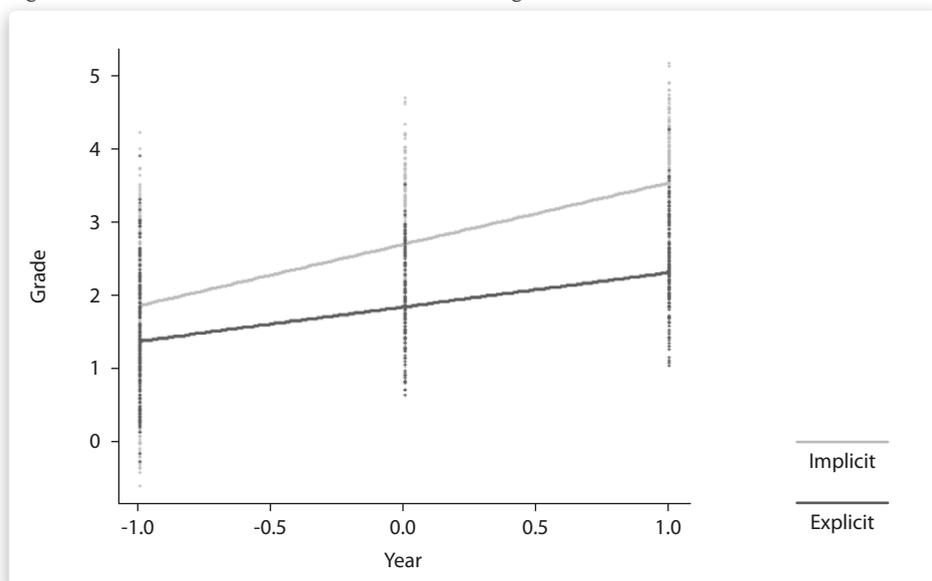
5.4 Results of model 2: hours of L2 exposure as a predictor

Using model comparison including the aforementioned predictors and random effects, the following model specification was determined to be optimal: Grade ~ L2exposure*Year + (1+Year|Participant) + (1+ Year|Class) + (1|School).

Table 14 details the results of model 2. Estimates with an absolute t-value of 2 or above may be considered significant. The results show that L2 exposure is the only significant factor in the model. Contrary to model 1, there is no significant interaction between L2 exposure and year.

Table 14. LMER model focusing on the factor L2 exposure to predict the scores of oral and written skills.

	Estimate	Std. Error	T value
(Intercept)	1.38	0.25	5.56
L2exposure	0.01	0.004	3.82
Year (–1, 0, 1)[7]	0.29	0.23	1.28
L2exposure:Year	0.007	0.003	1.82

---

7   The factor "Year" was centered for the purpose of the analysis, we replaced Year 1,2,3 by -1, 0 and 1.

5.5 Comparison of both models

Model comparison of these two models using anova revealed that Model 1 with the program as a predictor was significantly better ($p < .001$) than the one with the amount of L2 exposure.

Table 15. AIC and BIC and percentage of variance explained by the fixed factors for each model.

	Df	AIC	BIC	R²
Model 1	12	6437*	6508*	41.9%
Model 2	12	6414	6486	31.2%

*p<.001

6. Discussion

The current study has taken up DeKeyser's (2003) suggestion to conduct a realistic experiment in an ecologically valid study, in actual 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, which meant we had to go beyond the traditional dichotomy of explicit vs. implicit L2 instruction. We compared two types of programs used in the Netherlands: one traditional one with a great deal of explicit grammar, and a relatively new one developed in Canada, called AIM, which does include focus on form, but only implicitly. Thus, we compared the effects of a predominantly explicit program with a predominantly implicit program. Moreover, we calculated the effects of L2 exposure. We tested oral and written skills after one, two and three years of instruction. Our main research question was which program - the predominantly explicit or implicit one - was more effective in terms of general proficiency operationalized as free response oral and written skills after three years of instruction. We also wanted to know whether it was mainly L2 exposure and use or the actual program that could account for the effectiveness.

First, we will discuss the tasks we used to elicit oral and written proficiency. We feel that the positive approach we took for testing the oral and written proficiency of our learners, in which meaning-making in all its aspects is taken into account, is helpful in gauging a general proficiency level. Also, setting up our own grid based on our own corpus was useful in that it was able to make the fine grained differences necessary to discriminate between our own learners at these very beginning levels that a standard such as the CEFR cannot make. In our study level 1 and 2 would not have been included in the A1 level and 2, 3, and 4 would all be considered A1. Level 5 and 6 would be a low A2.

Second and without a doubt, we can say that the predominantly implicit form-focused program, AIM, is more effective in the development of L2 oral and written skills after one, two and three years of instruction than the traditional explicit form-focused programs used in the Netherlands. Because AIM is a specific method, which includes scripted lessons, pared-down language, a great deal of playful repetition, the use of iconic gestures to scaffold for meaning, and maximum L2 exposure and use (all in line with Dynamic usage-based (DUB) principles), it is difficult to say whether it is the implicit aspect of this condition only that allowed learners to progress more. However, it is clear from our study that explicit instruction was not more effective in achieving communicative skills.

Our findings are corroborated with earlier findings in more detailed smaller studies based on an earlier data set and the current data set. Rousse-Malpat & Verspoor (2012) showed that as far as accuracy of three morphological items (present tense, gender and negation) is concerned, the explicit group was more accurate after one year, but there were no differences after two years. After three years, *Chapter* 4 shows that the implicit method led to greater speech rate, grammatical complexity and accuracy of the present tense and L2 use in speaking. For writing, *Chapter* 5 shows that the implicit method led to greater fluency and complexity at various morpho-syntactic levels and also to greater use of short routines in writing. So far, we have not carried out a more detailed study in the current dataset on accuracy, but Gombert (forthcoming) showed no differences in accuracy in writing in similar conditions after six years of instruction. These findings together suggest that the implicit learner, who probably focuses on meaning first and then form (Van Patten, 1990), takes longer to achieve the same level of accuracy as the explicit learner.

One reason the implicit program may have been more effective is because of mere exposure. Because of their scripted method, the teachers in the implicit program were able to speak in the L2 exclusively from the beginning, so learners had almost 100% of French during their lessons. Teachers in the explicit condition, on the other hand, varied a great deal in the amount of L2 they provided in the classroom, with one teacher using 60% L2 French in class. Therefore, we decided to compare two models, one with Program as predictor and one with Exposure as predictor. The models showed that both Program and Exposure significantly contributed to the oral and written results. However, the results of our model comparison (Program versus Exposure) showed that the model including L2 exposure accounted for 10% less of the variance than the model including program as predictor. This is in line with other studies that have shown that L2

exposure alone is not enough to be effective on all aspects of the linguistic system. For example, research on the effectiveness of immersion programs in Canada has showed that L2 exposure was particularly effective on fluency, but not on accuracy (Day and Shapson, 1991).

As the model with exposure does not explain the most variance, it is the program that drives the effect. One may wonder which principles of the implicit AIM program lead to this effect that it is not the mere exposure but providing exposure using the AIM method that explains the results better. In *Chapter 4*, we discuss in detail the differences between a structure-based approach, which takes rule learning as a driving force in L2 learning, and a Dynamic usage-based program, which takes making symbolic, syntagmatic and other associations as the driving force in learning an L2. In a Dynamic usage-based program, just hearing and using phrases or chunks (words in similar sequences together) over and over again helps to entrench them. As explained in Rousse-Malpat & Verspoor (2018, *Chapter* 2), the AIM program focuses much on oral skills first with a lot of repetition and drills and succeeds in entrenching the form with the meaning and the pragmatic use of constructions at all levels (morphological, lexical, phrasal, sentential, and discourse levels). The drills and repetition of frequently occurring patterns are possible because the AIM method is very scripted. Moreover, learners are kinesthetically involved by using the iconic gestures, which may help them to remember the words and utterances better. We think that the implicit learners were able to automatize chunks, constructions and conventionalized ways of saying things (Smiskova, Verspoor & Lowie, 2012), which helped them becoming more fluent and increased the number of L2 learning events and use. Despite the lack of creativity that these activities seem to have, the data show that learners were able to use those constructions in another context early on. Apparently the AIM method enabled learners to get to "a sufficiently high level of automatization to enable second language use that is both fluent and almost completely accurate" (DeKeyser, 2017; p.17).

Our findings are not in line with the claim that learners need explicit declarative knowledge before they can develop procedural and automatized knowledge (cf. DeKeyser, 2017). In our case, only the explicit group clearly developed explicit knowledge of rules before starting any kind of proceduralization, but it is unclear whether the implicit group also developed explicit knowledge. We might consider gestures an explicit cue but the nature of the gestures gives more explicit knowledge about lexical rather than morphological or grammatical items. Still, in the program there is a gesture for the feminine and masculine articles, and

if a sound is not pronounced the teacher may gesture as a form of feedback. We do think that the teacher's careful articulation of the prompts (accompanied by gestures) helps to make all sounds more salient and helps learners to notice them, but this is very different from explicit rule presentation and practice happening in explicit instruction. It seems that in the AIM method learners acquire the language more like in a naturalistic L1 manner in which there is a great deal of playful repetition in different contexts too. However, more research is needed to make theoretical claims about the degree of declarative and procedural knowledge in the AIM learners.

Our findings are not in line either with the conclusions from the meta-studies (Norris & Ortega, 2000; Spada & Tomita, 2010; Goo et al., 2015). This is not surprising as for example Norris and Ortega (2000) pointed out themselves that their results were based on many brief interventions in which the tests were often biased as they were constrained tests focusing on morpho-syntax. As Doughty (2003) has pointed out, if certain variables had been controlled for better, explicit instruction would not have been found more effective. Also, Andringa and Schultz (2016) have shown that merely controlling for the amount of exposure would have changed the results in Spada and Tomita (2010).

Of course, there are many limitations to the study. In actual classroom studies there are many variables that cannot be controlled for. For example, gauging the amount of L2 exposure has relied on estimation, which may not have been very precise. Also, we compared two programs which were inherently very different, so we do not know precisely what affected the differences. However, we were able to show that explicitness of instruction was not a prerequisite for greater effectiveness. Our results have also shown variation among the groups. In both conditions, some groups did better than others. In addition, within classes, some learners did better than others. We were not able to account for much of that kind of variation in our model as we only had access to information about teacher and scholastic ability. Despite the fact that classroom studies may be inherently messy, we do hope to see more of them, perhaps as in-depth studies that trace both teacher and student behavior longitudinally to see for example how learners become creative after a period of drill and repetition. Nevertheless, our study can be seen as an attempt to capture the reality of the classroom, to try to understand the link between L2 exposure and type of instruction, and to contribute to the improvement of classroom research as a field.

7.   Conclusion

In this chapter, we aimed at looking at the real dynamics of a classroom regarding the effects of instruction and exposure on L2 learning by redefining the concepts of explicit and implicit instruction to fit better the reality of the field. Our holistic approach to the question led us to make general conclusions on the effectiveness of an explicit vs. an implicit L2 teaching program after three years of instruction. We showed with an ecologically-valid study that, over time, a predominantly implicit program was more effective than a predominantly explicit program on general oral and written skills in the L2. By comparing two models, we showed that even though L2 exposure was a significant predictor to explain our results, the type of program was actually the best predictor. We claimed that the reason why the implicit program was more effective was because it took the most advantage of the L2 exposure by increasing the amount of L2 learning events and use of drills and repetition of frequently occurring patterns in a scripted and scaffolded input.

We also pointed out that other studies clearly show that explicit and implicit instruction involve other processes and thus, have different beneficial effects on language production. The next chapters show the results of two detailed analysis (one on speaking, one on writing) on the effects of both instructional programs. To do so, we used the two most comparable groups in the study in terms of scholastic level and amount of L2 exposure.