Sybille Heinzmann*, Seraina Paul, Robert Hilbe,
Nicole Schallhart and Mirjam Egli Cuenat

Development of productive language skills through language exchange in primary schools in Switzerland – An exploratory intervention study

https://doi.org/10.1515/eujal-2021-0024

Abstract: The present study examined the effect of language exchange activities on the development of productive foreign language skills of primary school pupils. The sample comprised a total of 392 pupils from German- and French-speaking Switzerland, attending 5th and 6th grade. The pupils' speaking and writing skills were examined longitudinally using a quasi-experimental pre-test post-test design with an intervention and control group. The intervention lasted one school year and included two physical reunions with the partner class, as well as preparatory and follow-up activities. The German-speaking pupils' productive skills were assessed at the beginning and end of the school year using communicative tasks. The data was analysed by means of multiple regression analysis. Results show that the intervention had a positive impact on the pupils' speaking skills and their general language competence, as measured by a C-test. However, the intervention had no significant impact on their writing skills. The findings suggest that even short encounters at primary school level with beginners have the potential to spur language development if they are didactically well embedded.

Keywords: language exchange, speaking skills, writing skills, primary school, intervention study

***Corresponding author: Sybille Heinzmann,** St. Gallen University of Teacher Education, Institute for Language Teacher Education, St. Gallen, Switzerland, E-Mail: sybille.heinzmann@phsg.ch
Seraina Paul, St. Gallen University of Teacher Education, Institute for Language Teacher Education, St. Gallen, Switzerland, E-Mail: Seraina.Paul@phsg.ch
Dr. Robert Hilbe, St. Gallen University of Teacher Education, Institute for Language Teacher Education, St. Gallen, Switzerland, E-Mail: robert.hilbe@phsg.ch
Nicole Schallhart, St. Gallen University of Teacher Education, Institute for Language Teacher Education, St. Gallen, Switzerland, E-Mail: nicole.schallhart@phsg.ch
Dr. Mirjam Egli Cuenat, Fachhochschule Nordwestschweiz, Pädagogische Hochschule Standort Brugg, Institut Primarstufe, Brugg, Switzerland, E-Mail: mirjam.egli@fhnw.ch

Zusammenfassung: Die vorliegende Studie untersuchte den Einfluss von Sprachaustauschaktivitäten auf die Entwicklung der produktiven Fremdsprachenkenntnisse von Primarschülerinnen und Primarschülern. Die Stichprobe umfasste insgesamt 392 Schülerinnen und Schüler aus der deutsch- und französischsprachigen Schweiz, die die 5. oder 6. Klasse besuchten. Die Sprech- und Schreibkompetenzen der Schülerinnen und Schüler wurden in einem quasi-experimentellen Prä-Test-Post-Test-Design mit einer Interventions- und einer Kontrollgruppe längsschnittlich untersucht. Die Intervention dauerte ein Schuljahr und umfasste zwei physische Treffen mit der Partnerklasse sowie vorbereitende und nachbereitende Aktivitäten. Die produktiven Fertigkeiten der deutschsprachigen Schülerinnen und Schüler wurden zu Beginn und am Ende des Schuljahres anhand von kommunikativen Aufgaben erhoben. Die Daten wurden mit Hilfe von multiplen Regressionsanalysen analysiert. Die Ergebnisse zeigen, dass sich die Intervention positiv auf die Sprechfertigkeit und die allgemeine Sprachkompetenz, gemessen mittels C-Test, der Schülerinnen und Schüler auswirkte. Die Intervention hatte jedoch keinen signifikanten Einfluss auf deren Schreibfertigkeiten. Die Ergebnisse deuten darauf hin, dass selbst kurze Begegnungen mit Anfängern in der Grundschule das Potenzial haben, die Sprachentwicklung zu fördern, wenn sie didaktisch gut eingebettet sind.

Resumen: El presente estudio examinó el efecto de las actividades de intercambio lingüístico en el desarrollo de las competencias productivas en lengua extranjera de los alumnos de primaria. La muestra estaba formada por un total de 392 alumnos de la Suiza alemana y francesa que cursaban 5º y 6º curso. Se examinaron longitudinalmente las capacidades de expresión oral y escrita de los alumnos mediante un diseño cuasi-experimental de preprueba y posprueba con un grupo de intervención y otro de control. La intervención duró un año escolar e incluyó dos reuniones físicas con la clase asociada, así como actividades de preparación y seguimiento. Las competencias productivas de los alumnos de habla alemana se evaluaron al principio y al final del curso escolar mediante tareas comunicativas. Los datos se analizaron mediante un análisis de regresión múltiple. Los resultados muestran que la intervención tuvo un impacto positivo en la capacidad de hablar de los alumnos y en su competencia lingüística general, medida por una prueba C. Sin embargo, la intervención no tuvo un impacto significativo en sus habilidades de escritura. Los resultados sugieren que incluso los encuentros breves con principiantes en la escuela primaria tienen el potencial de estimular el desarrollo del lenguaje si están bien integrados en la didáctica.

1 Introduction

Language exchange and mobility are considered effective means of promoting language learning because they enable learners of all ages, and especially younger learners, to experience, use, and acquire language in meaningful social contexts, combining extracurricular and intra-curricular learning (Llanes and Muñoz 2013). While mobility programs are popular throughout Europe, especially among tertiary level students and teenagers (e.g. Erasmus & Youth in Action), and while programs for younger learners also exist, not many learners engage in them. Especially at primary school level, exchange activities are rare (Krüger-Potratz et al. 2018).

Switzerland, which is officially quadrilingual and where different cultural and linguistic communities live together in a small area, appears to be a privileged context for various forms of exchange, also at lower levels of schooling. Most Swiss communities are monolingual due to the territorial principle anchored in the Federal Constitution of the Swiss Confederation (Bundesverfassung der Schweizerischen Eidgenossenschaft 1999), but the country's internal cohesion depends heavily on a certain degree of individual multilingualism. The Languages Act, adopted in 2007, based on Article 70 of the Federal Constitution, binds the Confederations and the cantons to encourage mutual understanding between the language communities through domestic exchange (Bundesgesetz über die Landessprachen: Art. 14). Language exchanges as an important field of action in the Swiss education system, promoting multilingual and intercultural skills as well as motivation at all school levels, have recently been supported by a national exchange strategy (Schweizerische Eidgenossenschaft and EDK 2017). The current Swiss curricula, such as *Lehrplan 21* or *Plan d'Etudes Romand*, also recommend language encounters and exchange activities and outline corresponding learning goals (Egli Cuenat et al. 2015).

Given this favourable starting position regarding educational policy, it could be assumed that language exchanges across language regions are common in primary schools in Switzerland. However, this is not yet the case. In the school year of 2015/16, when this study was planned, less than three percent of all Swiss pupils in compulsory education (grades 1–9) benefited from an exchange within the country (Bundesamt für Statistik 2017; ch Stiftung 2017). At the primary level, which is the focus of the present study, participation is even lower, at just under 0.5 percent throughout Switzerland. Among the reason for this low participation rate, the additional workload for teachers associated with an exchange seems to be one of the most weighty ones. They are already very busy in their everyday work and, therefore, not always willing to take on additional work (Borel and Gygax 2011). In the case of primary school pupils, the

young age and the limited language skills of the pupils make things even more difficult.

In light of this scarcity of exchange activities even in contexts of geographical proximity and the challenges associated with their implementation at lower grades, there is a need to develop age-appropriate forms of exchange, which are targeted at intensifying language learning as well as the need to examine their impact. So far, there are only a few studies available on the impact of language exchange projects in Switzerland (Ogay et al. 2007; ch Stiftung 2006; Heinzmann et al. 2014). This is particularly true at the primary level. In general, exchange projects with this age group have not yet been the subject of much international research either. The project presented in this paper contributes to closing this gap[1]. The aim was to develop an exchange setting at primary level of shorter duration between partner classes in German-speaking and French-speaking Switzerland which can be implemented relatively easily and to test its effectiveness with an intervention group and a control group. The learners were 6th grade students. The focus of the study was on curriculum-relevant speaking and writing skills as well as language learning motivation. The development of pupils' target language skills was only assessed in the German-speaking classes. The present paper focuses exclusively on the effect of the exchange activities on the development of these target language skills.

2 Theoretical and empirical underpinnings

Research on language stays and study abroad has always been strongly interested in their effectiveness for language development. This outcome-oriented research has shown that longer stays (usually in the form of study abroad) have the potential to improve the language skills of those involved on a long-term basis (see e.g. Freed 1995; Engle and Engle 2004; Magnan and Back 2007). Most progress can be found in the areas of fluency (Llanes and Muñoz 2009; Segalowitz and Freed 2004), vocabulary (Dewey 2008; Serrano et al. 2011), and pragmatic skills (Regan 1995). As far as writing is concerned, the picture is less clear, however, as findings

1 The project was developed under the direction of Mirjam Egli Cuenat in collaboration with Seraina Paul, Katharina Höchle Meier, Thomas Roderer and Wilfrid Kuster. It was carried out under the direction of Sybille Heinzmann and Seraina Paul in collaboration with Robert Hilbe and Nicole Schallhart as well as teachers and school classes from the cantons of St. Gallen, Vaud, Neuchâtel and Fribourg. The project was financed by the National Competence Centre for Multilingualism and the PH St. Gallen.

are contradictory (Pérez-Vidal and Barquin 2014; Pérez-Vidal and Juan-Garau 2009; Sasaki 2004, Sasaki 2007, Sasaki 2009, and Sasaki 2011).

However, since longer language or study stays in a foreign-language area are mainly undertaken at upper secondary or tertiary level, research with regard to younger learners is relatively rare (Brunner 2015). The larger study by the Franco-German Youth Office (Krüger-Potratz et al. 2018; Melin and Wagner 2015) is worth mentioning. Using ethnographic videos and interviews with pupils and teachers, the study dealt with five-day encounters between school classes from five French and five German primary schools (generally two encounters of which one was in each partner's country). A specific feature of the exchange situations was that the learners were not taught each other's language as their first foreign language. The focus was not on foreign language learning, but on informal, social and intercultural learning. The study revealed a wide range of mostly non-verbal communication strategies, forms of intercultural learning as well as changes in self-perception and perception of others that were developed by the children. The authors point out that while exchange activities at primary school level are still scarce, they can pave the way for linguistic and intercultural learning. Based on their findings, they formulate recommendations for the design of exchange situations and teacher training, even in situations with no or very limited target language skills in the partner language.

Similarly, the study by Krok et al. (2010), also funded by the Franco-German Youth Office, documented the learning experiences of children who took part in an international children's encounter of at least 5 days. Although this study, too, focused on intercultural rather than language learning, it nevertheless provides interesting information on the language use of the pupils. Children from Germany aged between eight and twelve years were examined. The children participating in the group encounter stayed in a shared accommodation during the encounter. The survey of the children showed that they used a variety of communication methods when communication problems arose: they used a different language, communicated with gestures, sought help from counselors, spoke their first language or painted a picture. Another important finding of the study was that experiences and learning effects are variable. Those children, who due to their age, experience and attitudes, had more numerous resources, were more often represented in the group of learners who communicated in a versatile and flexible way – i.e. also used their resources. Older children, who had a wider range of experience and were more likely to have foreign language skills, used more versatile means of communicating with the exchange students. Furthermore, the observations indicated that the experiences and thus also the learning effects differed more among the children the more their own initiative was required when making contact with participants from abroad. This shows the importance of the didactic setting and support of such an encounter.

With her case study of a two-week exchange between two grade 6 classes (13-year-old and younger students) in Ontario and Quebec, MacFarlane (2001) examined the question if and how exchanges and classroom language learning were complementary. Focus groups, observation, and informal teacher interviews were used. Based on Wong Fillmore's multidimensional *Teachability Model*, social, linguistic, and cognitive processes were identified and compared between classroom language learning on the one hand and the exchange context on the other hand. The study concludes that these two language settings are complementary. The classroom and the exchange setting provide different L2 input contexts which are both needed for successful language acquisition. Language skills acquired in the classroom do not transfer automatically to authentic interactions with native speakers, as the language used in classroom differs from the exchange setting (e.g. easier vocabulary, slower tempo, non-native accent). Students therefore have difficulties when they are exposed to authentic language in an exchange setting and it takes time for them to adapt and to build social bonds. Face-to-face communication with native speakers is crucial to acquire a full repertoire of language skills, and, particularly, to be able to communicate successfully with native speakers in natural contexts. The author therefore advises to cyclically integrate exchange activities into the foreign language curriculum at school.

In Switzerland, there has only been one study on language exchange at primary level so far. The study by Botturi Kappler and Negrini (2018) examined the *AlpConnectar project*, an exchange project based on digital technology between nine primary schools (4th–6th grade) from the cantons of Grisons, Valais and Ticino. The exchange activities were spread out over a whole school year and included a virtual acquaintance phase on the one hand and the development of a more comprehensive didactic and interdisciplinary course unit on a topic in the field of biology, geography, and history by means of video sequences on the other hand. The aim was for learners to document certain aspects of their village or home in the foreign language and make them accessible to the other class. The project ended with an encounter of the respective classes involved. The interviews conducted with pupils and students indicated that a virtual language exchange in the Alpine arc between the German-, French-, and Italian-speaking parts of Switzerland had improved the language skills of the primary school pupils. In addition to improving their language skills, they expanded their communication strategies and reduced their inhibitions about communicating in the foreign language.

Apart from these studies at primary level, there are also a few studies at lower secondary school level. The study by Oakley et al. (2018) aimed at identifying challenges and successes associated with the design and implementation of a digital storytelling exchange between three Australian und four Chinese middle

schools. 151 students (aged between 13 and 15) participated, 34 in Australia and 117 in China. Pupils in both countries were asked to work in small groups and to describe their daily lives and a traditional tale from their own culture in the target language (Mandarin for Australian students, English for Chinese students) in a multimodal format. The digital stories were posted on an online platform in video format, so that the overseas pears could respond, comment, ask questions or offer corrections or suggestions. Teachers were interviewed at the end of the project to elucidate their perspectives on what had worked well and what challenges they and their students had experienced. According to the teachers the digital stories enabled student learning through authenticity and real-world relevance. Tokaryk (2019) makes a similar observation in her *eTwinning project*. Pupils aged 12–15 years from 7 European countries, all beginners of French, wrote a story in French together using digital tools. According to the teachers, in doing so, the young people developed reading strategies, vocabulary, and discovered new grammar rules.

It must be noted that none of these studies assessed the language skills of the students involved by means of tests. The statements on linguistic competence and progress are based on self-assessments of the participants, teacher assessment, and in some cases also participant observations. Furthermore, the studies presented so far are not intervention studies with a control group. While they provide valuable information about the development of age-appropriate exchange settings, they do not allow for well-founded statements about the effectiveness of the exchange activities studied. This is particularly true for longer projects such as *Alpconnectar*, which extend over a whole school year, as it can be assumed that the language skills of the pupils would also have improved in regular language classes.

While Massler's study (2008) is not based on a (quasi-)experimental design either, it is one of the few studies which examined the language competences using writing products of the participants. It showed that a project-oriented six-month e-mail contact between a class of German 9th grade pupils and a Canadian high school class had a positive effect on communicative and cooperative writing skills in English. Whether similar project-oriented writing lessons would have been just as effective without exchange with pupils from the target language area remains open, however, as no control group was examined. Evans and Fisher (2005) studied the effects of a six to eleven-day language stay of 68 English-speaking 13 to 14-year-old adolescents in France by analysing systematically pre- and post-departure tests on listening, reading, speaking, and writing. The greatest improvements were found in listening comprehension and writing (fluency and communicative content, but not accuracy). The study implied a comparison of the results taken two years later in the public GSCE examination with a matched con-

trol group. No long-term effects of the stay were found, but the analysis yielded a small correlation between the stay and listening comprehension. The study also showed a correlation between linguistic progress and the language support of the host family in terms of help and error correction, as reported by the pupils.

Noteworthy quasi-experimental studies on children that examined the language learning gains in exchanges settings using intervention and control groups include the study by Llanes (2012) and Llanes and Serrano (2017). Llanes (2012), which explored the short- and long-term effects of a two-month stay abroad (SA) on nine 11-year-old Catalan/Spanish bilingual's oral and written skills, found significant gains in their oral skills immediately after the stay as compared to seven peers who had stayed at home (AH). Contrary to Evans and Fisher (2005), no significant increase in their written skills was found. Llanes and Serrano (2017) compared the effect of learning context (SA vs AH) on three different age groups: children (age 10–11), adolescents (age 12–15) and adults in terms of their oral and written development (N=197). All SA participants spent 2–3 months abroad. While the SA context did not affect the participants' writing skills, it had an effect on their speaking skills with the SA context being more beneficial than the AH context. In terms of age, the adults showed greater increase in their written skills than the rest of the groups. In terms of oral skills, the differences were less pronounced. Here adults also outperformed the other groups but only with regard to lexical diversity. Regarding the interaction effects between learning context and age, younger participants in the SA context seemed to experience the greatest gains concerning the oral skills measured, but these effects were not so clear with respect to the written skills.

3 Methodology

3.1 Objectives and research questions

The aim of the present study was, on the one hand, to design and test an exchange setting at primary level in Switzerland that is appropriate to the level and can be implemented at reasonable cost, and, on the other hand, to investigate its effect on the productive language skills and language learning motivation of the pupils. With this dual objective, the aim was both to develop an exemplary and low-threshold didactic scenario with teaching materials that can be used for language exchanges between primary school classes in French- and German-speaking Switzerland, and to contribute to empirical research. Productive skills, as reported above, are at the heart of many educational interventions related to exchange activities. However, learning outcomes are usually not systematically

studied or, if they are, the results are controversial. Hence, the need for evidence-based knowledge in this specific area is high. The intervention had to take into account the political and educational framework mentioned in the introduction so that it could be implemented in school practice despite limited time and personnel resources. The present contribution presents the design of the exchange setting which was the basis of the intervention and focuses on the following research question:

> RQ: Do the implemented exchange activities between primary school classes in German- and French-speaking Switzerland have an impact on curriculum-relevant speaking and writing skills of pupils in French compared to no exchange activities?

3.2 Sample

The sample[2] consisted of seven intervention classes each from German-speaking Switzerland and French-speaking Switzerland who took part in a language exchange in the context of the present project (cf. Table 1). Interventions are time- and work intensive for all participants involved. This was also the case in the present study, which included exchange partners from both language regions. Fourteen classes for the intervention group and 7 classes for the control group was considered a feasible number to recruit from the field and at the same time a sufficient number to be able to demonstrate potential effects of the intervention as a number of studies in the European context, which have investigated the effect of rather short exchange activities of younger learners on linguistic skills or language learning motivation, were able to identify measurable effects with a similar or even much smaller sample size (Evans and Fisher 2005; Llanes 2012; Llanes and Muñoz 2009, Llanes and Muñoz 2013; Heinzmann et al. 2015). The original plan was to draw lots to assign classes to the intervention or control group for the purpose of randomization. However, not enough teachers registered, probably also due to the time-consuming nature of the project. Therefore, the classes of all the teachers who registered were recruited as intervention classes. In a second step, classes for the control group were recruited separately. It must therefore be assumed that a certain bias exists due to the selection of the intervention classes. However, the control group also includes three classes that had originally registered as intervention classes, which mitigates this bias.

2 The anonymized data set can be obtained from the documentation centre of the Research Centre for Multilingualism via the following link (project number 0058): https://centre-plurilinguisme.ch/en/centre-de-documentation

At the time of the study, the pupils from German-speaking Switzerland were in their second year of learning French (6th grade) with two to three hours of French per week. The pupils from French-speaking Switzerland were in the third or fourth year of learning German (5th or 6th grade), also with two to three hours of German per week. The minimum competence level aimed at by the end of sixth grade according to the curriculum is a CEFR level of A1.2. The control group consisted of four classes from German-speaking Switzerland and three classes from French-speaking Switzerland, who did not take part in a language exchange.

Table 1: Sample by study group and target language (N=392)

	German-speaking Switzerland (Target language French)	French-speaking part of Switzerland (Target language German)	Total
Intervention group	136 pupils 7 classes 6th school year	135 pupils 7 classes 5th and6th school year	271 pupils 14 classes
Control group	59 pupils 4 classes 6th school year	62 pupils 3 classes 5th and 6th school year	121 pupils 7 classes
Total	195 pupils 11 classes	197 pupils 10 classes	392 pupils 21 classes

Table 2 provides information on the socio-demographic data of the intervention and control groups.

Table 2: Sociodemographic data of the study groups (N=392)

	Intervention group	Control group	Total
Age	Minimum: 10.0 years Maximum: 13.6 years Average: 11.5 years	Minimum: 10.1 years Maximum: 13.6 years Average: 11.5 years	Minimum: 10.0 years Maximum: 13.6 years Average: 11.5 years
Gender	48.6 % girls 51.4 % boys	52.6 % girls 47.4 % boys	49.9 % girls 50.1 % boys
Nationality	59.9 % Swiss 17.5 % dual citizens 22.6 % non-Swiss citizens	46.6 % Swiss 16.4 % dual citizens 37.1 % non-Swiss citizens	55.8 % Swiss 17.2 % dual citizens 27.1 % non-Swiss citizens
Languages	48.8 % monolingual 51.2 % multilingual	36.0 % monolingual 64.0 % multilingual	44.9 % monolingual 55.1 % multilingual

At the outset of the study[3], the pupils in both groups were between 10.0 and 13.6 years old. On average, they were 11.5 years old. 49.9 percent of the sample were girls and 50.1 percent boys. The two study groups differed in terms of gender. In the control group, girls were slightly overrepresented, whereas boys were slightly overrepresented in the intervention group. Across both groups 55.8 percent were Swiss, 17.2 percent dual Swiss citizens, and 27.1 percent non-Swiss citizens. The proportion of non-Swiss citizens in the control group was significantly higher than in the intervention group. 44.9 percent of the children in the sample were monolingual and 55.1 percent multilingual. In line with the larger proportion of non-Swiss citizens, more pupils in the control group were multilingual.[4] While being aware of the necessity to match the control and intervention group as closely as possible with regard to possible influencing variables (see Marx and Steinhoff 2017), creating matched groups ultimately proved impossible as it was very difficult to recruit classes who would participate, either in the intervention or the control group.

3.3 Intervention

The didactic setting consisted of two physical exchange meetings at the partner classes' places of residence, each with one overnight stay, their preparation and follow-up, as well as activities to maintain contact between the two meetings. This didactic setting in its entirety constitutes the intervention in our study. The study, therefore, can only draw conclusions about the effectiveness of this setting as a whole as compared to no intervention. No conclusions can be drawn about the effectiveness of the physical reunions or the use of the accompanying materials in the classroom separately.

The teachers of the intervention classes carried out the intervention by themselves but received lesson plans and worksheets prepared by the project team with instructions to stick to these plans without major changes. The aim of this procedure was, on the one hand, to ensure the highest possible comparability of the intervention in the different classes. On the other hand, the teachers who al-

3 The 1.8.2017 was used as a reference date for calculating the age.

4 The multilingual speakers mentioned the following languages as family languages in addition to German and French: English (33), Italian (33), Albanian (29), Portuguese (22), Spanish (17), Serbian (8), Turkish (8), Croatian (7), Bosnian (6), Russian (5), Hungarian (4), Arabic (3), Tamil (3), Aramaic (2), Romanian (2), Kurdish (2), Moroccan (2), Somali (2), Polish (1), Brazilian (1), Chinese (1), Dioula (1), Slovak (1), Dutch (1), Greek (1), Macedonian (1), Japanese (1), Lingala (1), Oromo (1), Thai (1), Tigrinya (1), Amharic (1), Vietnamese (1), and sign language (1).

ready took on additional workload by carrying out the language exchange were to be relieved as far as possible.

When devising the didactic material, competencies from the Swiss-German *Lehrplan 21* and the Swiss francophone *Plan d'études romand* that are relevant in the context of an exchange activity were taken into account. In addition, references were made to the contents of the coursebook *envol* (Achermann et al. 2000, Achermann et al. 2001) or *Junior* (Endt et al. 2016; Endt et al. 2017) used in the cantons where the participating classes were from in order to show teachers synergies with regular teaching. The development also included materials from the *Plurimobil* project for primary school (Egli Cuenat et al. 2015) as well as materials already tested in a similar project (Paul and Aguirre 2014).

The teachers of the intervention classes were introduced to the planned programme and the use of the didactic material during a half-day training session. This preparatory meeting also served to inform the teachers about the exact course of the project with the different measurement dates and to give them the opportunity to get to know their tandem partner personally.

The intervention consisted of a preparation stage, two physical encounters half a year apart, and a follow-up phase which are delineated below. Overall, the students in the intervention group had about 13 extra hours of contact with the target language during the physical encounters compared to the control group with no language exchange but regular foreign language lessons. No such estimate can be made for the preparation and follow-up stages of the intervention, as it is not exactly known whether the teachers did more foreign language lessons than usual or whether they simply carried out the intervention during their regular foreign language lessons.

Preparation for the exchange: The pupils of the intervention classes were prepared for the first physical exchange meeting in their regular lessons by informing them about the upcoming exchange in form of an orientation run in the classroom. They designed and presented a self-portrait booklet with questions about themselves in the foreign language. A first introduction to the partner class took place through an exchange of letters. The preparation also included practising important chunks for the physical encounters by means of role-plays. The pupils also prepared a typical regional activity (e.g. a song, a dance, a story, a game), which they presented at the exchange meetings.

Programme of the two encounters: The two physical encounters took place in autumn 2017 and spring 2018 and followed the same pattern. In autumn 2017, the classes from French-speaking Switzerland visited their partner classes in German-speaking Switzerland, followed by a visit to the French-speaking part of Switzer-

land in spring 2018. On the first day, the two partner classes met in front of the group accommodation of the visiting class at some point in the afternoon, where they participated in activities to break the ice and get to know each other as well as cooperation games. Afterwards, they cooked dinner and ate together in mixed-language groups. The visiting class stayed in a group accommodation while the children of the local class went home.
The next day the two classes met in front of the local school. The visiting children taught their partners the previously prepared activity typical for their home region. Afterwards, the pupils completed an orientation run through the village or town of the host class in mixed-language groups. The assignments for the orientation run were formulated in the language of the visiting class, so that the pupils had to cooperate with their partners to find the answers. After lunch, which was an organised picnic by the host class, the visiting class went home again.

Between the two exchanges: Between the two physical encounters, the pupils talked about and reflected on the experiences of the first encounter in the context of their regular lessons. In addition, the pupils wrote short, simple picture descriptions in the foreign language to accompany photos from the exchange meeting, which they then sent to the partner class to supplement. This activity served as the basis for an e-mail exchange with the other class. Following this, the teachers produced an exchange newspaper consisting of all the pictures and picture descriptions as a reminder of the event.
To maintain contact between the two physical encounters, the two partner classes also exchanged a 'culture box'. The idea stems from the *Culture Box project* by Porczyk Fromowitz and Hainaut (2012). The learners first discussed and reflected on their own canton before putting together a culture box that would best represent their canton using seven objects. Possible objects in a culture box could be postcards, flags, specialities or books. The culture box was sent to the partner class, which in turn discussed its contents and thus learned about the other class' canton.

After the second exchange: After the second physical exchange meeting a short reflection within the class took place. This phase was deliberately kept short, as the time load of the project with two encounters and about 20 lessons in the regular classroom was already relatively high for the participating teachers.
The control group did not receive any special treatment and followed the usual programme, based on the curricula and course books currently used in the cantons of all participating classes.

3.4 Design and instruments

The productive language skills of the pupils were examined longitudinally using a quasi-experimental design with an intervention and control group and a pre-test post-test design. The data collection took place at two time points: in August 2017 before the start of the intervention (pre-test; t_0) and in May 2018 (post-test; t_1), approximately one month after the second encounter and a week after the last follow-up lesson. The same tasks were used at t_0 and t_1 and administered simultaneously to the intervention and the control group. For research economic reasons, the productive skills were only examined in the target language French (that is in the German-speaking intervention and control classes).

In addition to the language tests, a student questionnaire was administered before and after the intervention. It contained around 45 items and recorded socio-demographic data (gender, age, nationality, first languages), motivational aspects such as general language learning motivation, foreign language use anxiety and self-efficacy beliefs in relation to foreign language use. The development of these motivational aspects will not be discussed further below as this is done in another article (Heinzmann et al. 2022). Furthermore, a teacher questionnaire was administered to the teachers of both the intervention and control classes in order to gauge the teacher's motivational dispositions and self-efficacy beliefs as well as the intervention class teachers' perception of the effects of the intervention. Due to the focus on the competences of the students, this data source will not be discussed in detail in this article, but will be used to support the interpretation of the results.

3.4.1 Speaking tasks

The French speaking competences of the pupils in the intervention and control classes in German-speaking Switzerland were assessed by means of an interview and a picture comparison. The dialogic interview consisted of six simple questions for the pupils, such as *Comment ça va?*' How are you?'. The questions were based on the DELF Prim A 1.1 exam, whereby the vocabulary being used was adapted to the coursebook used. In the monologic picture comparison, the pupils had to name three differences in two similar but not quite identical pictures. The pictures used were taken from the *Cambridge Young Learners Test* (OCR 2006, OCR 2007; new Cambridge Assessment 2019). It was ensured that they could well be described by means of vocabulary that the pupils knew from their coursebook. The assessment of pupils' speaking competences took between five and ten min-

utes per pupil. The interlocutors were researchers from the St. Gallen University of Teacher Education. Group assignment (intervention vs control group) was known to the testers. The pupils were given no preparation time.

3.4.2 Communicative writing task and C-test

The French writing competences of the pupils from the intervention and control classes were assessed with a communicative writing task from the *lingualevel*[5] repository (unpublished *lingualevel* task). As part of this task, the pupils had to write a short e-mail with information about themselves. The contents that had to be addressed in the e-mail were specified in German. A C-test was used to assess the pupils' general language competence. The C-test also originated from *lingualevel* and contained three short texts with 16 gaps each. The pupils had to fill the gaps with common French words (cf. Figure 1).

Marie-Claire se présente

Salut, je m'appelle Marie-Claire. J'ai 15 a_ _. J'habite à l_ campagne, pr_ _ de Bordeaux. J'a_ _ _ les ani_ _ _ _. Mon ani_ _ _ préféré, c'_ _ _ le che_ _ _.

Figure 1: Excerpt from C-Test

As a further source of data, teacher questionnaires were used to collect information on the organisation of the exchange, the teachers' qualifications and experiences as well as their attitudes and motivation with respect to carrying out the exchange. Furthermore, the researchers were present as participant observers at each physical encounter and took field notes.

3.5 Data analysis

The oral interactions and productions as well as the written productions of the students were rated according to specific rating criteria. Group assignment of students (intervention vs control group) was known to the raters. The rating criteria and the guidelines for their use were developed by the project team, partly based

5 *lingualevel* offers foreign language assessment tools for use in Swiss schools (https://www.zebis.ch/lingualevel).

on previous projects (see Haenni Hoti et al. 2009). In the developed rating scheme, the focus was set on task fulfilment (cf. Table 3). In the oral tests, task fulfilment was the only rating criterion. For the written production it was decided that, in addition to task fulfilment, the lexical diversity of the texts would be evaluated. After considering various suitable measures, the project team decided on *Guiraud's Index for Lexical Complexity* (RTTR; Guiraud 1960)[6]. This index is one of the best predictors for human ratings of lexical diversity (Vanhove et al. 2019) and provides a measure of text length (number of tokens i.e. single words) as well as another measure of lexical diversity, which is often used for studies with children and studies dealing with short texts. For the oral skills lexical diversity was not evaluated because of the great effort that would have been involved in a literal transcription of the sometimes difficult to understand utterances of the students.

In the beginning of the rating phase, all project staff were trained by the project directors. During this training, the rating criteria and guidelines developed by the project team were explained and then tested using a series of sample essays and oral tests to ensure the highest possible interrater reliability. Then, the data from three classes was rated separately by two raters, which represents 27 percent of the total data. After two weeks, a meeting was held to discuss and adjust any non-matching ratings. On the basis of the non-conformities and uncertainties that arose during the double rating, the rating guidelines were further specified and revised. Afterwards, the project staff independently rated the remaining essays and oral tests, with the project directors being consulted in unclear cases. The interrater reliability was 95 percent for oral competence and 83 percent for writing competence.

Tables 3 and 4 provide an overview of the gradations used in the rating of the oral productions. As can be seen from the table, the number of points achieved is based on the amount of assistance required by the students to complete the task. This approach reflects the conceptualisation of progress as an increasing independence from assistance from interlocutors and from contextual and situational references (Werlen 2006). In addition, points were awarded in the interview if the pupils themselves initiated turns (e.g. by asking the interviewer questions).

6 Formula for calculating Guiraud's Index for Lexical Complexity (Guiraud 1960) : $RTTR = \frac{Types}{\sqrt{Tokens}}$

Table 3: Rating categories for the interview tasks (oral)

Task	*Fulfilled at the first go (4P)*	*Fulfilled after repetition of the question (3P)*	*Fulfilled after verbal or non-verbal assistance (2P)*	*Fulfilled after translation (1P)*	*Not fulfilled (0P)*
Answering interview questions	Spontaneous, content-wise, appropriate and comprehensible statement/ answer	Spontaneous, content-wise, appropriate and comprehensible statement/ answer after the question has been repeated in French	Spontaneous, content-wise, appropriate and comprehensible statement/ answer after assistance (e.g. non-verbal hints, modelling, paraphrasing)	Spontaneous, content-wise, appropriate and comprehensible statement/ answer in French after German translation	No reply, German reply

Table 4: Rating categories for the picture comparison tasks (oral)

Task	*Fulfilled (2P)*	*Partly fulfilled (1P)*	*Not fulfilled (0P)*
Describing differences between two pictures	The test person describes the differences in picture A and B. The test person points to differences and can explain the difference when asked to do so. He/she can name both elements. The test person does not recognise any difference but can describe the difference (both elements) when shown to him/her.	The test person can only name the differing element in one picture but not in the other. The test person can only name the differing element in one picture but not in the other even after having been asked to specify or after having been shown the element that has to be described on the picture.	The test person cannot name the differences even after having been asked to specify or after having been shown the elements that have to be described on the picture. The test person only points to the picture. No answer.

Table 5 provides an overview of the gradation used for the criterion *task fulfilment* in written production.

To keep analysis pragmatic and ensure a high interrater reliability, a three-level gradation was chosen. Two additional points were awarded for following genre conventions. One point was awarded for including the name of the addressee in the E-Mail and the other one for including the name of the sender.

Table 5: Rating categories for the writing tasks (written)

Fulfilled (2P)	*Partly fulfilled (1P)*	*Not fulfilled (0P)*
The required information (e.g. greeting, birthday) is available as a complete sentence (except for the greeting and farewell) and completely in French. All sentences containing a subject, verb and object are considered complete sentences. The sentence does not have to be correct, however (i.e. grammatical errors, for example, are allowed).	The required information (e.g. greeting, birthday) is available but not as a complete sentence (e.g. *Je 12 ans [I 12 years], Je mon anniversaire 5.8.2012 [I my birthday 5.8.2012*] and/or not completely in French. However, the meaningful element (context relevant words) is understandable/ accessible to a benevolent French-speaking person (eg. *My anniversaire 12) [my birthday 12]*.	The required information (e.g. greeting, birthday) is missing or the required information is present, but not as a complete sentence or completely in another language, or all context relevant words are in another language, so that the message is not understandable/ accessible for a benevolent French-speaking person (e.g. *ce Zürich et suiss*) [it Zürich and Switzerland]) where one cannot deduce that he/she lives there.

The data from the questionnaires was analysed descriptively and aggregated by means of factor analyses. Scales were constructed on the basis of the identified factors by calculating the mean of the item values. These scale values were used together with the socio-demographic variables as covariates in the regression analyses (see following chapter) to investigate the influence of individual characteristics of the students on the development of language skills. For this purpose the difference of the test score from t_0 to t_1 was used as a dependent variable. In addition to group membership, the measured test score at t_0 was also included as an independent variable, since it can be assumed that the development of language skills depends on the baseline level before the intervention. In addition, given the clustered nature of the sample the class mean of the measured skills before the intervention was included as a predictor in the regression models to account for the multilevel structure of the data and to control for the influence of intraclass correlation. Due to the differences in demographic characteristics in the intervention and control groups, gender and language background were included as control variables in the exploratory stepwise procedure when calculating all regression models. Since no significant influence of these variables could be demonstrated, they were not included in the regression models presented.

4 Results

4.1 Speaking skills

The speaking test score was calculated by adding up the points achieved for the individual subtasks of the interview, the picture description and the points awarded for self-initiated tasks[7] (cf. Table 3 and Table 4). The maximum achievable score was 36. The empirical test scores of the pupils ranged from 6 to 31 for t_0 and from 7 to 32 for t_1, which indicates that there is neither a ceiling nor a floor effect.

In both groups, there was a significant increase in the test score from the pre-test to the post-test. The mean values of the two groups at t_1 are practically identical, but the increase across time is significantly greater in the intervention group than in the control group which was confirmed by linear regression (cf. Table 7). It needs to be pointed out, however, that the control group started out with a significantly higher pre-test score than the intervention group.

Table 6: Speaking skills – Intervention and control groups t_0 and t_1 in comparison

Study group *Mean value difference t_1–t_0*		***N***	***Minimum***	***Maximum***	***Mean value***	**Standard deviation**
Control group +4.20	Speaking skills t_0	58	6	31	21.19	5.20
	Speaking skills t_1	56	9	31	25.32	4.24
	Valid values	56				
Intervention group +7.35	Speaking skills t_0	107	6	28	17.79	5.93
	Speaking skills t_1	103	7	32	25.14	4.51
	Valid values	103				

7 We also ran the regression analyses without these additional points as they may reflect competencies more guided by real-world interactions and the results were the same.

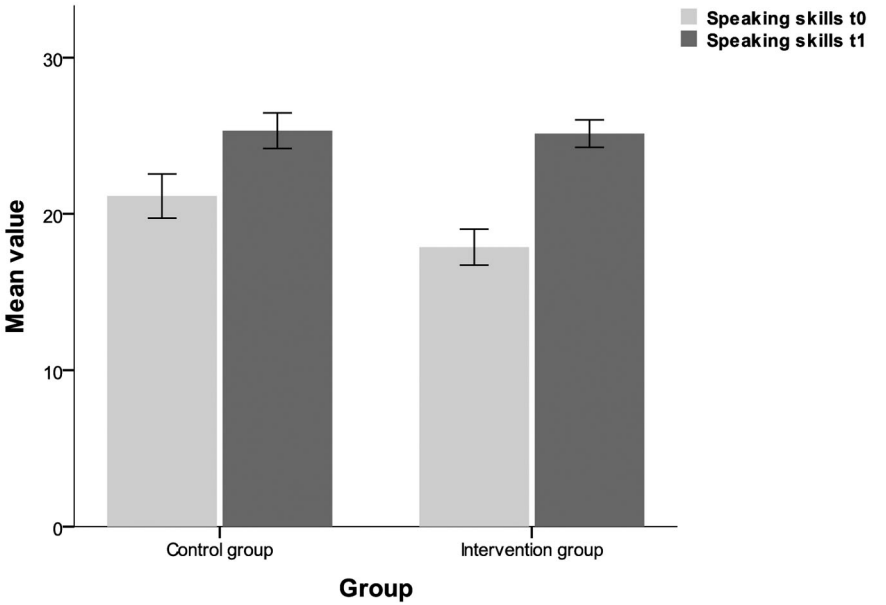

Figure 2: Speaking skills – Mean values and confidence intervals t_0 and t_1

Table 7 presents a regression model with the difference in test scores between t_0 and t_1 as the dependent variable and the intervention and other significant predictors as independent variables. The *intervention* turns out to be a significant predictor for the development of pupils' speaking skills. Additional explanatory variables are the *speaking test score* as well as learners' *language use anxiety* at t_0, both of which have a negative impact on the development. The *class mean of the speaking test score* t_0 is also included in the regression to account for the clustered nature of the sample. However, this control variable is not significant with respect to the development of speaking skills. The model explains 59.7 per cent of the variance and thus indicates a large effect.[8]

8 The adjusted coefficient of determination R^2 is always given and interpreted according to the guidelines of Cohen (1992): small effect: R^2 = .0196; medium effect: R^2 = .1304; large effect: R^2 = .2592.

Table 7: Speaking skills – Regression models to explain the difference t_1–t_0

Model		B	Standard error	Standard Beta	T	Significance
R^2 = .597 F(4,148) = 55.813 p = .000	(Intercept)	17.72	1.92			
	Intervention	1.75	.57	.19	3.08	.00
	Speaking skills t_0	-.57	.05	-.71	-11.44	.00
	Speaking skills t_0 (class mean)	.04	.09	.03	.47	.64
	Language use anxiety t_0	-.99	.36	-.14	-2.72	.01

4.2 Writing skills

In the writing task, there were 7 subtasks that were rated according to the rating criteria in Table 5 and two additional points for following genre conventions[9]. Therefore, the maximum score is 16 points. The students' test scores for t_0 ranged from 0 to 14 and for t_1 from 3 to 16.

With respect to the writing skills, too, a significant increase in test scores could be observed between the pre-test and the post-test. The increase is comparable in both groups. However, once again the control group already had higher scores at the outset.

Table 8: Writing skills – Intervention and control groups t_0 and t_1 in comparison

Study group *Mean value difference t_1–t_0*		*N*	*Minimum*	*Maximum*	*Mean value*	Standard deviation
Control group +3.48	Writing skills t_0	58	1	14	9.31	3.11
	Writing skills t_1	56	7	16	12.79	2.12
	valid values	55				
Intervention group +3.87	Writing skills t_0	107	0	14	7.82	3.22
	Writing skills t_1	105	3	16	11.69	2.70
	valid values	103				

9 We also ran the regression analyses without these additional points as they may reflect competencies more guided by real-world interactions and the results were the same.

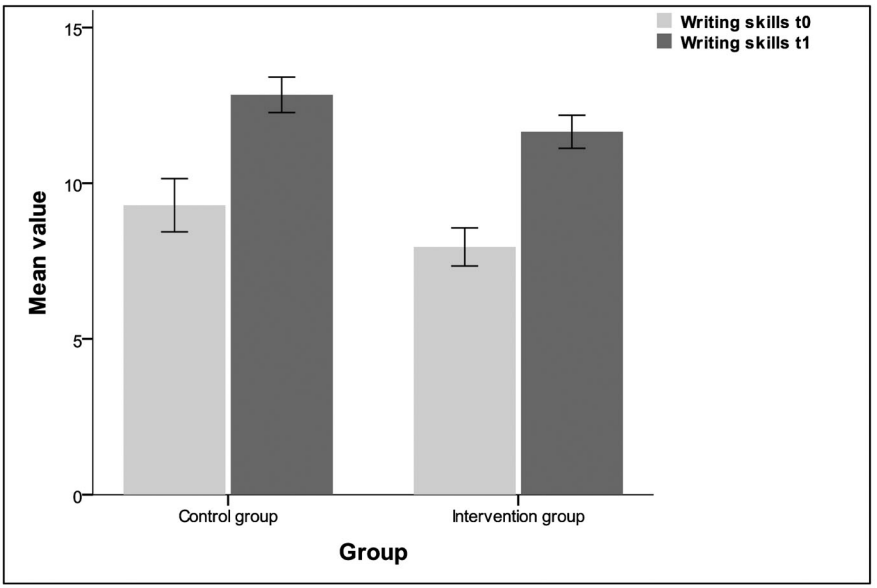

Figure 3: Writing skills – Mean values and confidence intervals t_0 and t_1

Table 9 shows a regression model with covariates to predict the change in the writing skills test score.

Table 9: Writing skills – Regression models to explain the difference t_1–t_0

Model		B	Standard error	Standard Beta	T	Significance
R^2 = .669 F(4,142) = 72.898 p = .000	(Intercept)	3.60	1.17			
	Intervention	.07	.31	.01	.23	.82
	Writing skills t_0	-.75	.05	-.88	-16.57	.00
	Writing skills t_0 (class mean)	.36	.11	.19	3.17	.00
	Language learning motivation t_0	1.23	.21	.29	5.78	.00

In the model, which includes control variables, the *intervention* is not significant. Significant predictors that contribute to the explanation of the development of pupils' writing skills are their *writing skills* at t_0, the *class mean of writing skills* at t_0 as well as *their motivation to learn the target language* at t_0. Once more, the test scores of the pre-test have a negative effect on development, i.e. learners who already have a

high test score before the intervention show a lower increase on average. Additionally, class context plays a role in explaining the development of writing skills, as the class mean is a significant predictor. Furthermore, pupils with high *motivation to learn the target language* before the language exchange exhibit a more favourable development regarding writing skills. The corrected model explains 66.9 percent of the variance in the difference in test scores and is significant.

Lexical diversity: To determine lexical diversity, *Guiraud's Index for Lexical Complexity* (RTTR) was calculated for the written texts of each pupil at both time points. In both groups, a significant increase in lexical diversity can be observed across time. The increase is somewhat greater in the intervention group than in the control group. However, here too, the control group already had a higher starting level at the outset.

Table 10: Lexical diversity – Intervention and control group t_0 and t_1 in comparison

Study group *Mean value difference t_1–t_0*		***N***	***Minimum***	***Maximum***	***Mean value***	**Standard deviation**
Control group +0.49	RTTR t_0	58	1.00	4.80	3.48	.64
	RTTR t_1	56	2.12	5.98	3.97	.81
	Valid values	55				
Intervention group +0.78	RTTR t_0	107	1.41	5.37	3.15	.76
	RTTR t_1	105	2.11	5.61	3.93	.76
	Valid values	103				

In the regression model (cf. Table 11) the intervention turns out to be a significant predictor for the development of lexical diversity in writing. Besides that *lexical diversity at t_0*, the class mean of *lexical diversity at t_0* and the *motivation to learn the target language* turn out to be significant predictors of development. The higher the language learning motivation at the outset, the more lexical diversity increases. The lexical diversity of the pre-test has a negative effect on development. Learners with high scores at the beginning show a lower increase on average. As was found for writing skills in general, the class context also plays a role in explaining the differences in development of lexical diversity. Students from classes with a higher class mean before the intervention increase their lexical diversity more. The model explains a total of 30.8 percent of the variance in the change in lexical diversity.

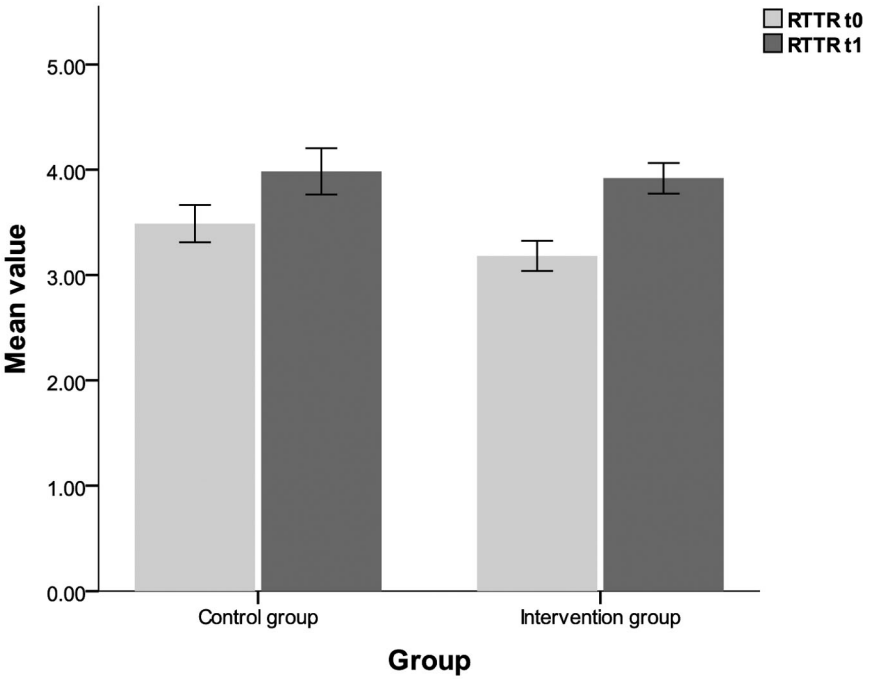

Figure 4: Lexical diversity – Mean values and confidence intervals t_0 and t_1

Table 11: Lexical diversity – Regression models to explain the difference t_1–t_0

Model		B	Standard error	Standard Beta	T	Significance
R^2 = .308 F(4,148) = 17.456 p = .000	(Intercept)	-24.40	5.82			
	Intervention	3.74	1.17	.24	3.20	.00
	RTTR t_0	-4.94	.90	-.45	-5.50	.00
	RTTR t_0 (class mean)	10.09	1.78	.49	5.67	.00
	Language learning motivation t_0	4.58	.81	.40	5.63	.00

C-Test: The test value was calculated as the sum of all correctly filled in gaps, so that a maximum score of 48 points could be achieved. Only gaps where each individual letter of the word filled in was correct and in the correct position were considered correct. The C-test was very difficult for the pupils, which is evidenced by the moderate test scores at both time points. These ranged between 0 and 29 for t_0 and between 2 and 34 for t_1.

Both in the intervention group and in the control group, a significant increase in the test scores across time can be observed. The increase is significantly higher in the intervention group than in the control group. It should be noted, however, that the test scores already differed between the two groups at the outset (t_0). The control group started with a significantly higher mean value before the intervention, which indicates a higher initial level in the control group.

Table 12: C-Test – Intervention and control group t_0 and t_1 in comparison

Study group Mean value difference t_1–t_0		N	Minimum	Maximum	Mean value	Standard deviation
Control group +5.61	C-Test t_0	58	0	29	16.53	5.26
	C-Test t_1	56	9	33	22.14	5.58
	Valid values	55				
Intervention group +8.62	C-Test t_0	106	0	29	13.12	5.57
	C-Test t_1	105	2	34	21.74	5.99
	Valid values	102				

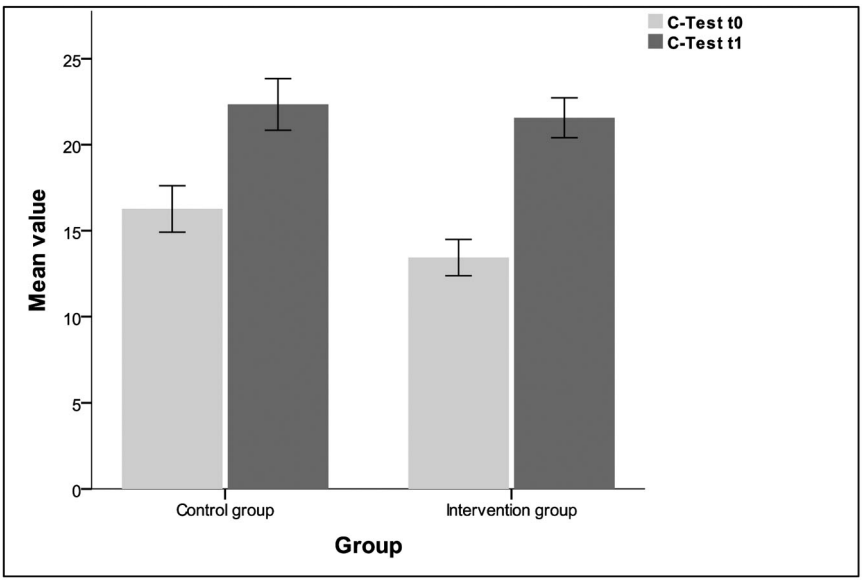

Figure 5: C-Test – Mean values and confidence intervals t_0 and t_1

In the regression model to predict the development in the C-test test scores (cf. Table 13), the *intervention* turns out to be a significant predictor despite the inclusion of control variables. In addition to the intervention, the pupils' *C-test scores* at t_0, *the class mean of C-test scores* at t_0 as well as pupils' *self-efficacy beliefs* and the *foreign language use anxiety* at t_0 proved to be significant predictors of the linguistic progress as measured by the C-test. Besides the C-test scores at the outset, the clearest influence on the development of overall proficiency are pupils' *self-efficacy beliefs* at the outset: learners who believe in their foreign-language skills show a greater increase in their overall proficiency. The *test result* t_0 on the other hand has a negative effect on development, i.e. the higher the test score of a person at the outset, the more likely it is that their increase is lower. The class context determines part of the variance as the class mean at t_0 is significant. The *language use anxiety*, surprisingly, has a positive effect on the development of learners' C-test scores: students with higher anxiety show a greater increase in test scores. Overall, the corrected model explains 29.1 percent of the variance and is significant.

Table 13: C-Test – Regression models to explain the difference t_1–t_0

Model		B	Standard error	Standard Beta	T	Significance
R^2 = .329 F(5,146) = 15.313 p = .000	(Intercept)	-7.92	3.36			
	Intervention	2.96	.80	.33	3.68	.00
	C-Test t_0	-.45	.07	-.52	-6.48	.00
	C-Test t_0 (class mean)	.43	.16	.25	2.70	.01
	Self-efficacy beliefs t_0	3.15	.57	.47	5.50	.00
	Language use anxiety t_0	1.67	.51	.26	3.24	.00

5 Discussion

Action-oriented tasks were used to assess both speaking and writing skills. In addition, a C-test was used to measure general language competence. With regard to writing skills, results are mixed. While no effect of the intervention could be demonstrated for task fulfilment, a positive effect could be demonstrated for lexical diversity. As far as writing task fulfilment is concerned, significant progress was observed across time in both groups. This progress was comparable across groups, however, and did not differ significantly between the groups. This finding contrasts

with the results of the Evans and Fisher (2005) study, where, however, the learners were slightly older and more advanced in their learning of French, and the stay was longer and more intensive. On the other hand, it is in line with the study by Llanes (2012) as well as Llanes and Serrano (2017), who also found no effect for the development of writing skills. However, the positive effects found for lexical diversity are in contrast to the study by Llanes (2012) as well as Llanes and Serrano (2017), where the same measure for lexical diversity was used and no effect of a stay abroad could be found. Given the mixed results of studies dealing with the effects of mobility programs on writing skills both with older and younger learners in a study abroad context (Pérez-Vidal and Barquin 2014; Pérez-Vidal and Juan-Garau 2009; Sasaki 2004, Sasaki 2007, Sasaki 2009, and Sasaki 2011), the mixed results in this study are not overly surprising. Llanes and Serrano (2017) point out that apparent contradictions could be related to the instruments used for measuring writing skills. As Evans and Fisher's (2005) analyses suggest, qualitative analyses of written texts may capture more subtle morphosyntactic, lexical and idiomatic changes related to authentic communicative experiences in the target language. However, Llanes and Serrano (2017) also suggest that the inconclusive results might be attributable to the different length of programs investigated or the amount and type of practice that learners receive during the intervention. It may be that a longer stay or an intervention that is more clearly targeted at the promotion of writing skills would be more effective for their development. While the subjects in our intervention group also had written contact with the partner class throughout the school year, communication during the two physical exchange meetings was entirely oral. It can be assumed that pupils in regular French lessons (and hence the students in our control group) also repeatedly encountered writing occasions in their French classes. On the other hand, the pupils in the control group did not have the opportunity for repeated oral interaction with French-speaking pupils. This could explain why the intervention primarily had a positive effect on speaking skills.

Although with regard to speaking skills, too, significant progress can be observed in both groups across time, the progress is significantly greater in the intervention group (cf. Table 7). Consequently, the intervention is a significant predictor of the development of speaking skills. It needs to be pointed out, however, that the intervention and control group already differed significantly with regard to their speaking skills before the intervention. The pupils in the control group had a significantly higher starting level. However, the test values do not suggest a ceiling effect, i.e. the pupils were not able to solve all or almost all the test items. Accordingly, the test was not too easy for either group.

Apart from the positive effect on speaking skills and lexical richness in writing, the intervention also had a positive effect on pupils' general language competence, which was assessed by means of the C-test. The picture here is analogous to that

identified with respect to speaking. The test scores of both the control and the intervention group increased significantly from the first to the second test. However, the increase was significantly greater in the intervention group than in the control group (cf. Table 13). Once again, however, the starting level of the two groups differed significantly at the outset of the study (t_0) with the control group outperforming the intervention group, which indicates a higher initial level of language competence in the control group. Due to the relatively low test scores of both groups in the C-test even after the intervention, a ceiling effect can be ruled out.

The fact that learners with initially lower levels of competence made greater progress than learners with initially higher levels of competence reflects the results of two large-scale studies that looked at the predictors of learning gains in study abroad settings. In the study by Brecht et al. (1995) involving American students, those students who had lower test scores in speaking, listening and reading at the beginning of their stay abroad showed greater gains. The same pattern was found in a study by Lapkin et al. (1995), which focused on secondary school students. Again, those students with lower initial test scores in listening, speaking, reading and general language proficiency benefited more. However, Dewey (2007) notes that these findings could also be attributable to a ceiling effect and that the tests used were not demanding enough to detect progress by more advanced learners. Such a ceiling effect can be largely ruled out for the present study, however, based on the test scores achieved. On the other hand, the results of the present study on the influence of prior competences on learning progression also differ from the findings of two previous studies in which children and young people with a better starting position benefited more (Krok et al. 2010; Heinzmann et al. 2014). The question as to which learners benefit more or less from which exchange formats and why needs further clarification.

All in all, the present study shows a small positive effect of the exchange setting on the general language competence of the pupils, their speaking skills and their writing skill in terms of lexical richness. When carrying out a language exchange at primary level, it is certainly important not to overlook the fact that pupils are at an elementary level of language use only. Accordingly, an implementation at this level needs appropriate scaffolding measures on the part of the teachers and didactic materials as well as realistic expectations of progress on the part of pupils and teachers.

6 Limitations

Obviously, the results of the study have to be interpreted in light of certain limitations. Firstly, we were unable to create a randomized sample or to closely match

control and intervention groups with regard to target and control variables due to the difficulty of recruiting classes. Therefore, a selection bias must be assumed to exist with teachers in the intervention group being those who were eager to participate with their class in an exchange activity. Also, the two groups already differed with regard to their language skills at the beginning (t_0), which might have had an influence on the results. A further methodological limitation is that both testers and raters were aware of the group assignment of the students that they tested/rated, which might have had an influence on their testing or rating behaviour. Thirdly, we used holistic rating criteria rather than fine-grained linguistic analyses for the evaluation of the pupils' written and oral productions. Using the latter would have allowed for more detailed information on pupils' progress in the foreign language. Furthermore, no literal transcriptions of pupils' oral productions were made for research economic reasons for which reason it was not possible to investigate lexical diversity in students' speaking skills in analogy with the analyses conducted for writing.

7 Conclusion

Despite its brevity and low-threshold nature, the intervention had a positive impact on the pupils' speaking skills and general language competence as well as partly on their writing skills. This finding suggests that even short encounters at primary and beginner level can improve language competencies. As mentioned above, the design of the study does not allow any conclusions to be drawn about the extent to which the use of the accompanying material contributed to this result. Still, it can be stated that he majority of the teachers appreciated the authentic use of the foreign language and the fact that they worked towards the mastery of a specific encounter during the regular French lessons in which the acquired skills could be applied. They also appreciated the didactic material provided by the project team, which they found very useful, practical and helpful. Due to the high effort involved in preparing, implementing and following up a language exchange, it is important to familiarise student teachers in initial and in-service training with existing didactic materials and to make them available to other interested parties (see Krüger-Potratz et al. 2018)[10]. In the interest of sustainability, it

10 The didactic material developed in the present project is available for interested parties under the following link: https://www.phsg.ch/sites/default/files/cms/Forschung/Institute/Institut-Fachdidaktik-Sprachen/Downloads/Didaktisches%20Material%20Deutsch_Website.pdf

is also advisable to continue working with the class, teacher or school tandems after a successfully completed exchange activity, as this can also considerably reduce the effort required for a renewed implementation.

Another aspect, which needs to be considered in view of obtaining greater benefit from exchange projects, is the timing of such projects. The intervention took place in 6th grade. The end of 6th grade in Switzerland marks the transition from primary to lower secondary level and, thus, the dissolution of the existing school classes and rearrangement of classes into three different performance levels. As a result of the transfer, the tandem classes can no longer be brought together in the following year, so that the contacts once established between pupils are at best continued at a private level. Where this is possible, e.g. in smaller municipalities, it would make sense for the schools as institutions to continue the exchange activities initiated at primary level at lower secondary level. Given the superiority of children in benefitting from exchange and study abroad experiences (Llanes and Serrano 2017) and in the light of the encouraging results of Krüger-Potratz et al. (2018) in contexts with little previously developed target language competence, an earlier start of such activities could also be envisaged, provided sufficient scaffolding by accompanying teachers.

While our study points to the potential of short exchange activities to foster the development of productive skills of primary school children, it is clear that not every student profited to the same degree – as was also shown in other studies (e.g. Evans and Fisher 2005). This warrants future studies on the processes influencing why some students develop more than others in exchange settings. Finding out more about the processes behind learning and development also opens up paths for organizing and implementing exchange programs in such a way as to facilitate optimal learning conditions.

References

Achermann, Brigitte, Michel Bawidamann, Martine C. Tchang-George & Hanna Weimann. 2000. *Envol 5. Französischlehrmittel für das 5. Schuljahr.* Zürich: Lehrmittelverlag des Kantons Zürich.

Achermann, Brigitte, Michel Bawidamann, Martine C. Tchang-George & Hanna Weimann. 2001. *Envol 6. Französischlehrmittel für das 6. Schuljahr.* Zürich: Lehrmittelverlag des Kantons Zürich.

Borel, Virginie & Jessica Gygax. 2011. *Sprachliche Austauschmöglichkeiten für Schüler und Lehrpersonen in der Hauptstadtregion Schweiz: Überblick über die angebotenen und nachgefragten Möglichkeiten auf allen Stufen von der Primarschule, Sekundarstufe 1 und 2, Berufsbildung und Fachhochschulen*. Interne Publikation. Biel: Forum für die Zweisprachigkeit/Hauptstadtregion Schweiz.

Botturi, Luca, Daniela Kappler & Lucio Negrini. 2018. Digitally-supported language exchanges in primary school: The Alp Connectar project. *Studies in Second Language Learning and Teaching* 8(4). 795–843.

Brecht, Richard D., Dan E. Davidson & Ralph B. Ginsberg. 1995. Predictors of foreign language gain during study abroad. In Barbara Freed (ed.), *Second language acquisition in a study abroad context*, 37–66. Amsterdam: John Benjamins.

Brunner, Stefan. 2015. *Wirkungskompendium Jugendaustausch. Eine Übersicht wissenschaftlicher Forschungsergebnisse zur Wirkung von interkulturellem Austausch.* Bern: Intermundo.

Bundesamt für Statistik. 2017. *Statistik der Lernenden.* https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/personen-ausbildung/obligatorische-schule.html (accessed 8 January 2021).

Cambridge Assessment. 2019. *A1 Movers*. https://www.cambridgeenglish.org/exams-and-tests/movers/ (accessed 8 January 2021).

ch Stiftung für eidgenössische Zusammenarbeit. 2006. *Wege aufeinander zu. Persönliche Erfahrungen mit Austauschbegegnungen und Konsequenzen für die Förderung von Austausch.* Solothurn: ch Stiftung für eidgenössische Zusammenarbeit.

ch Stiftung für eidgenössische Zusammenarbeit. 2017. *Austauschstatistik 2015/16.* Solothurn: ch Stiftung für eidgenössische Zusammenarbeit.

Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112(1). 155–159.

Dewey, Dan P. 2007. *Language learning during study abroad: What we know and what we have yet to learn.* Japanese Language and Literature 41(2). 45–269.

Dewey, Dan P. 2008. Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad* 15(1). 127–148.

Egli Cuenat, Mirjam, Nicole Schallhart & Lukas Bleichenbacher. 2015. *Austausch wirkt – Wie profitieren Schüler/-innen der obligatorischen Volksschule von einem Austausch?* Solothurn: chStiftung.

Egli Cuenat, Mirjam, Kristin Brogan, Josephine Cole, Anna Czura, Chantal Muller, Anna Szczepańska, Lukas Bleichenbacher, Katharina Höchle Meier & Barbara Wolfer. 2015. *PluriMobil lesson plans for primary school.* Graz: ECML. https://plurimobil.ecml.at/Lessonplans/Mobilityinprimaryschool/tabid/3765/language/en-GB/Default.aspx (accessed 8 January 2021).

Endt, Ernst, Michael Koenig, Nadine Ritz-Udry & Hannelore Pistorius. 2016. *Junior. Deutsch für die Romandie. Kursbuch 7. Klasse.* München: Klett-Langenscheidt.

Endt, Ernst, Michael Koenig, Petra Pfeifhofer, Nadine Ritz-Udry & Hannelore Pistorius. 2017. *Junior. Deutsch für die Romandie. Kursbuch 8. Klasse.* Stuttgart: Ernst Klett Sprachen.

Engle, Lilli, & John Engle 2004. Assessing language acquisition and intercultural sensitivity development in relation to study abroad program design. *Frontiers: The Interdisciplinary Journal of Study Abroad* 10(1). 219–236.

Evans, Michael & Linda Fisher. 2005. Measuring gains in pupils– foreign language competence as a result of participation in a school exchange Visit. The case of Y9 pupils at three comprehensive schools in the UK. *Language Teaching Research* 9(2). 173–192.

Freed, Barbara F. (ed.). 1995. *Second language acquisition in a study abroad context.* Amsterdam & Philadelphia: John Benjamins.

Guiraud, Pierre. 1960. *Problèmes et méthodes de la statistique linguistique.* Dordrecht: Reidel.

Haenni Hoti, Andrea, Marianne Müller, Sybille Heinzmann, Werner Wicki & Erika Werlen. 2009. *Schlussbericht zum Forschungsprojekt Frühenglisch – Überforderung oder Chance? Eine*

Längsschnittstudie zur Wirksamkeit des Fremdsprachenunterrichts auf der Primarstufe. Luzern: Pädagogische Hochschule Zentralschweiz, Hochschule Luzern.

Heinzmann, Sybille, Nicole Schallhart, Marianne Müller, Roland Künzle & Werner Wicki. 2014. *Sprachliche Austauschaktivitäten und deren Auswirkungen auf interkulturelle Kompetenzen und Sprachlernmotivation. Forschungsbericht Nr. 44*. Luzern: Pädagogische Hochschule Luzern.

Heinzmann, Sybille, Roland Künzle, Schallhart Nicole & Marianne Müller (2015). The effect of study abroad on intercultural competence: Results from a longitudinal quasi-experimental study. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 26(1). 187–208. https://doi.org/10.36366/frontiers.v26i1.366

Heinzmann, Sybille, Robert Hilbe, Seraina Paula & Nicole Schallhart (2022). Sprachaustausch in der Grundschule als Motivator für das Erlernen der Landessprachen Französisch und Deutsch? Befunde einer Interventionsstudie in der Schweiz. In Anka Bergmann, Christoph Oliver Mayer, Jochen Pilkat (eds.), *Welche Zielsetzungen sind für Französisch, Spanisch, Russisch & Co. (noch) zeitgemäss? Zu Perspektiven der Schulfremdsprachen im Zeitalter von Global English und Digitalisierung*, 169–190. Frankfurt am Main: Lang.

Krok, Isabelle, Barbara Rink & Kirsten Bruhns. 2010. *Interkulturelle Kompetenz durch internationale Kinderbegegnung. Hauptstudie: Die Kinder im Mittelpunkt*. München: Deutsches Jugendinstitut e.V. https://www.dji.de/ueber-uns/projekte/projekte/interkulturelle-kompetenz-durch-internationale-kinderbegegnung.html (accessed 8 January 2021).

Krüger-Potratz, Marianne, Bernd Wagner, Christiane Montandon & Philippe Sarremejane (eds.). 2018. *Deutsch-französischer Grundschulaustausch – informelles und interkulturelles Lernen: Eine videoethnographische Studie*. Münster: Waxmann.

Lapkin, Sharon, Doug Hart & Merrill Swain. 1995. A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In Barbara F. Freed (ed.), *Second language acquisition in a study abroad context,* 67–94. Amsterdam: John Benjamins.

Llanes, Àngels. 2012. The short- and long-term effects of a short study abroad experience: The case of children. *System* 40(2). 179–190.

Llanes, Àngels & Carmen Muñoz. 2009. A short stay abroad: Does it make a difference? *System* 37(3). 353–365.

Llanes, Àngels & Carmen Muñoz. 2013. Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning Research* 63. 63–90.

Llanes, Àngels & Raquel Serrano. 2017. The effectiveness of classroom instruction 'at home' versus study abroad for learners of English as a foreign language attending primary school, secondary school and university. *The Language Learning Journal* 45. 434–446.

MacFarlane, Alina. 2001. Are brief contact experiences and classroom language learning complementary? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 58(1). 64–83.

Magnan, Sally S. & Michele Back. 2007. Social interaction and linguistic gain during study abroad. *Foreign Language Annals* 40. 43–61.

Marx, Nicole & Torsten Steinhoff. 2017. Unterrichtsbezogene Interventionen. In Michael Becker-Mrotzek, Joachim Grabowski & Torsten Steinhoff (eds.), *Forschungshandbuch empirische Schreibdidaktik*, 253–266. Münster & New York: Waxmann.

Massler, Ute. 2008. Ausdruck, Analyse und Förderung der schriftlichen Kommunikations- und Kooperationsfähigkeit in E-Mail-Projekten der Sekundarstufe I. *Zeitschrift für interkulturellen Fremdsprachenunterricht* 13(1).

Melin, Valérie & Bernd Wagner. 2015. The intercultural and non-formal learning processes of children in primary school exchange programmes in France and Germany. *Comparative and International Education 10*(3). 407–422.

Ogay, Tania, Myriam Gremion, Yulia Zharkova Fattore, Valérie Hutter,Théogène Gakuba & Stéphanie Gendre-Borruat. 2007. *La recherche suisse en éducation interculturelle. Panorama des recherches empiriques réalisées entre 1993 et 2006.* Genève: Société Suisse de la Recherche en Éducation.

Oakley, Grace, Mark Pegrum, Xi B. Xiong, Cher P. Lim & Hanbing Yan. 2018. An online Chinese-Australian language and cultural exchange through digital storytelling. *Language, Culture and Curriculum* 31(2). 128–149.

Oxford, Cambridge and RSA Examinations (OCR). 2006. *Asset Languages. External Assessment Sample Tasks.* Cambridge: OCR.

Oxford, Cambridge and RSA Examinations (OCR). 2007. *Asset Languages. External Assessment Sample Tasks.* Cambridge: OCR.

Paul, Seraina & Mai Ling Aguirre. 2014. *Veränderung der Sprachlernmotivation durch Sprachaustausch in der Primarschule*. Freiburg: Universität Fribourg/Freiburg (CH) Unpublished MA thesis.

Pérez-Vidal, Carmen & Elisa Barquin. 2014. Comparing progress in academic writing after formal instruction and study abroad. In Carmen Pérez-Vidal (ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts*, 217–234. Amsterdam: John Benjamins.

Pérez-Vidal, Carmen & Maria Juan-Garau. 2009. The effect of study abroad (SA) on written performance. *EUROSLA Yearbook 9*(1). 269–295.

Porczyk Fromowitz, Aneta & Christine Hainaut. 2012. *Culture Box.* Graz: ECML. https://plurimobil.ecml.at/LinkClick.aspx?fileticket=E2antKqcXYQ%3D&tabid=3581&language=en-GB (accessed 8 January 2021).

Regan, Vera. 1995. The acquisition of sociolinguistic native speech norms: Effects of a year abroad on second language learners of French. In Barbara F. Freed (ed.), *Second Language Acquisition in a Study Abroad Context*, 245–267. Amsterdam & Philadelphia: John Benjamins.

Sasaki, Miyuki. 2004. A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning* 54(3). 525–582.

Sasaki, Miyuki. 2007. Effects of study-abroad experiences on EFL writers: a multiple-data analysis. *The Modern Language Journa*l 91(4). 602–620.

Sasaki, Miyuki. 2009. Changes in English as a foreign language students' writing over 3.5 years: a sociocognitive account. In Rosa M. Manchón (ed.), *Writing in Foreign Language Contexts: Learning, Teaching, and Research*, 49–76. Bristol: Multilingual Matters.

Sasaki, Miyuki. 2011. Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: a longitudinal study. *TESOL Quarterly* 45(1). 81–105.

Schweizerische Eidgenossenschaft & EDK. 2017. *Schweizerische Strategie Austausch und Mobilität von Bund und Kantonen.* https://edudoc.ch/record/212831?ln=de (accessed 21 September 2021).

Segalowitz, Norman & Barbara F. Freed. 2004. Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition* 26. 173–199.

Serrano, Raquel, Àngels Llanes & Elsa Tragant. 2011. Analyzing the effect of context of second language learning: domestic intensive and semi-intensive courses vs. study abroad in Europe. *System* 39. 133–143.

Tokaryk, Jana. 2019. Deutsch-französischen Projektaustausch und multilaterale europäische Projekte mit eTwinning in den Schulalltag integrieren. *Französisch heute* 50(3). 14–17.
Vanhove, Jan, Audrey Bonvin, Amelia Lambelet & Rafael Berthele. 2019. Predicting perceptions of the lexical richness of short French, German, and Portuguese texts using text-based indices. *Journal of Writing Research* 10(3). 499–525.
Werlen, Erika. 2006. Neuland Grundschul-Fremdsprachenunterricht. Empirische Ergebnisse der Wissenschaftlichen Begleitung „Fremdsprachen in der Grundschule“ in Baden-Württemberg. *Babylonia* 1. 10–16.